



HAL
open science

Caractérisation structurale de la glycosylation des protéines : de la glycomique à la glycoprotéomique, une stratégie utile pour l'étude des gonadotrophines

Nicolas Eskenazi

► **To cite this version:**

Nicolas Eskenazi. Caractérisation structurale de la glycosylation des protéines : de la glycomique à la glycoprotéomique, une stratégie utile pour l'étude des gonadotrophines. Chimie analytique. Université Paris sciences et lettres, 2019. Français. NNT : 2019PSLET032 . tel-02896596v2

HAL Id: tel-02896596

<https://pastel.hal.science/tel-02896596v2>

Submitted on 25 Aug 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.



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'École Supérieure de Physique et de Chimie
Industrielles de la ville de Paris

**Caractérisation structurale de la glycosylation des
protéines : de la glycomique à la glycoprotéomique,
une stratégie utile pour l'étude des gonadotrophines**

Soutenue par

Nicolas ESKENAZI

Le 6 novembre 2019

Ecole doctorale n° 388

**Chimie physique et chimie
analytique de Paris Centre**

Spécialité

Chimie analytique

Composition du jury :

Clotilde, POLICAR Pr ENS-SU	<i>Présidente</i>
Thierry, RABILLOUD DR CEA-CNRS	<i>Rapporteur</i>
François, FENAILLE Ingénieur-chercheur CEA	<i>Rapporteur</i>
Charles, PINEAU DR INSERM - Université de Rennes	<i>Examineur</i>
Joëlle, VINH DR ESPCI - CNRS	<i>Directrice de thèse</i>



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'École Supérieure de Physique et de Chimie
Industrielles de la ville de Paris

**Caractérisation structurale de la glycosylation des
protéines : de la glycomique à la glycoprotéomique,
une stratégie utile pour l'étude des gonadotrophines**

Soutenue par

Nicolas ESKENAZI

Le 6 novembre 2019

Ecole doctorale n° 388

**Chimie physique et chimie
analytique de Paris Centre**

Spécialité

Chimie analytique

Composition du jury :

Clotilde, POLICAR Pr ENS-SU	<i>Présidente</i>
Thierry, RABILLOUD DR CEA-CNRS	<i>Rapporteur</i>
François, FENAILLE Ingénieur-chercheur CEA	<i>Rapporteur</i>
Charles, PINEAU DR INSERM - Université de Rennes	<i>Examineur</i>
Joëlle, VINH DR ESPCI - CNRS	<i>Directrice de thèse</i>

REMERCIEMENTS

Je tiens à adresser les premiers mots de ce manuscrit à tous ceux qui m'ont permis de mener à bien ce doctorat. Ma directrice de thèse, Joëlle Vinh, que je remercie évidemment pour ma formation scientifique pendant ces trois années (parfois tumultueuses), mais également pour ses discussions éclairées, ses critiques enrichissantes, ses idées justes et son soutien sans faille.

Je remercie les membres de mon jury de thèse, qui m'ont fait l'honneur d'évaluer ce travail et avec qui j'espère avoir l'occasion de collaborer dans les années à venir : Madame Clotilde Policar, Messieurs François Fenaille, Thierry Rabilloud et Charles Pineau.

Mes remerciements vont également à Monsieur Yves Combarous, à qui je dois beaucoup pour son aide précieuse.

Pour leur rôle dans ma formation, je dois sans conteste associer Chiara Giangrande et Ha Phuong Ta à ce manuscrit. Un grand merci à toutes les deux pour l'aide que vous m'avez apportée dans le développement de ce projet.

Merci à Manue, ma « coloc » et sponsor hydratation du labo, grande prêtresse du MALDI, à Iman, (ex-)coloc' incollable au blind-test et en informatique, à Shakir et son amour pour la Suisse (ses prairies, son air pur, Roger Federer...), à Yann qui lit probablement ces mots en pensant à sa prochaine course folle (200 km ?), à Gio le passionné (de science, de philosophie, mais surtout de Napoli). Merci également à celles et ceux qui ont été là tous les jours de cette aventure, qui ont été ma deuxième famille au SMBP. Massamba (bientôt ton tour, courage !), Alex, Martha, Ophélie, Huu Hien, Alexia, Stéphanie, et tous ceux qui ont croisé la route du SMBP, merci pour ces moments.

À toutes celles et ceux avec qui j'ai eu la chance de travailler un jour ou un an, ces interactions ont été autant d'expériences enrichissantes pour moi.

Un mot aussi pour les membres du Club Jeune de la SFSM et de son Bureau, c'est un plaisir de côtoyer une si belle association. Je rendrai mon étiquette « jeune » dans 5 ans mais j'espère conserver longtemps ces liens.

À mes amis, Max, merci d'être là depuis tout ce temps, Isa, Jules, Elsa, Pauline, loin des yeux, près du cœur, merci pour tout ce chemin parcouru ensemble, Marianne, team DID, Ben, Marine, Mégane, Nico, Kévin, quelle chance j'ai de vous avoir, Claude, Keke, pas changer assiette pour fromage ! Vous m'êtes tous indispensables.

Une mention spéciale à mes parents, ma sœur et ma famille, merci pour votre soutien.

*DILEGUA, O NOTTE! TRAMONTATE, STELLE! TRAMONTATE, STELLE!
ALL'ALBA VINCERÒ!*

TURANDOT – GIACOMO PUCCINI

Caractérisation structurale de la glycosylation des protéines : de la glycomique à la glycoprotéomique, une stratégie utile pour l'étude des gonadotrophines

RÉSUMÉ

La composante sucrée des glycoprotéines joue un rôle majeur dans de nombreux processus physiologiques et pathologiques. Leur analyse structurale nécessite des outils performants capables d'évaluer la forte hétérogénéité de ces glycanes. Des approches complémentaires ont été développées : l'analyse des N-glycanes par MALDI-TOF et l'approche protéomique par LC-ESI-FTMS.

La glycomique consiste en l'analyse des sucres indépendamment de la protéine, pouvant livrer une empreinte globale de la glycosylation au niveau d'une protéine, d'un fluide *etc.* Le développement de nouvelles conditions de préparation d'échantillon et de matrice incluant la modification chimique des sucres nous a permis d'améliorer leur détection et leur fragmentation par spectrométrie de masse.

La considération des glycanes comme partie des glycoprotéines et glycopeptides permet une sélectivité accrue et une caractérisation de l'hétérogénéité intramoléculaire de ces glycosylations. L'intégration de protéases aspécifiques et de fragmentation à énergie variable (SCE) nous a conduit à plus de diversité et de fiabilité dans les glycoformes identifiées.

De par leur implication dans la grossesse et le développement de l'embryon, les glycohormones gonadotropes ont un fort intérêt médical et vétérinaire. La combinaison des méthodes analytiques a été entreprise dans le but de caractériser les motifs de glycosylation spécifiques à différentes populations.

MOTS CLÉS

Glycosylation ; Spectrométrie de masse ; Protéomique

ABSTRACT

Polysaccharides from glycoproteins play an important part in numerous biological processes and pathologies. Their structural elucidation requires performant and sensitive tools capable of assessing the heterogeneity of such glycans. Different approaches were developed: N-glycan analysis by MALDI-TOF and LC-ESI-FTMS proteomics.

Glycomics is the analysis of glycans independently of their linked-protein and can be using to obtain a glycosylation fingerprint of a protein, fluid *etc.* Developments of innovative sample and matrix preparation conditions, including glycan chemical modification, improved their mass spectrometry detection and fragmentation.

Glycopeptide and glycoprotein analysis permits selective assessment of the glycan population across the protein hence allowing the characterization of molecular microheterogeneity of the different glycoforms. The development and integration of non-specific protease digestion and stepped-energy fragmentation produced more diverse and reliable identifications.

Because of their critical role in gestation and embryo development, glycosylated gonadotropins are of major medical and veterinary interest. Combining analytical methodologies granted access to the characterization of population-specific glycosylation motifs.

KEYWORDS

Glycosylation ; Mass spectrometry ; Proteomics

Sommaire

Valorisation.....	15
Liste des figures.....	19
Liste des tables.....	22
Liste des abréviations.....	23
A. Introduction.....	25
B. Première Partie : Etat de l'art	31
1. Glycosylation des protéines et glycohormones.....	33
1.1 Introduction.....	33
1.2 Nomenclature.....	37
1.3 Fonctions biologiques	40
1.4 Les hormones glycoprotéiques gonadotropes	42
1.4.1 La gonadotrophine chorionique humaine (hCG)	43
1.4.2 La gonadotrophine chorionique équine (eCG)	46
1.5 Synthèse in vivo.....	47
1.5.1 N-glycanes.....	47
1.5.2 O-glycanes (O-GalNAc).....	51
1.5.3 Autres glycoconjugués.....	53
2. Chimie analytique et glycoprotéomique	55
2.1 Introduction à la spectrométrie de masse	57
2.2 N-glycomique.....	60
2.2.1 Production des N-glycanes	61
2.2.2 Etude structurale des N-glycanes par MS.....	72
2.3 Glycoprotéomique bottom-up	79
2.3.1 Généralités.....	80
2.3.2 Glycoprotéomique <i>bottom-up</i>	85

2.4	Approches intacte et <i>top-down</i>	100
C.	Deuxième Partie : N-glycomique pour l'étude des choriogonadotropines 105	
1.	Développement d'une méthode de N-glycomique sensible pour l'étude de la glycosylation des glycohormones	107
1.1	Introduction.....	107
1.2	Développement de la méthode.....	109
1.2.1	Reproduction du protocole publié	109
1.2.2	Optimisation de la méthode	111
1.3	N-glycomique pour l'analyse de glycohormones urinaires	117
1.4	Conclusions	119
D.	Troisième Partie : Caractérisation structurale de la glycosylation des gonadotrophines chorioniques par approche bottom-up	123
1.	Protéolyse aspécifique et fragmentations <i>stepped-HCD</i> pour la comparaison de deux hCG d'origines géographiques distinctes.....	125
1.1	Introduction et problématique.....	125
1.2	Protocole de glycoprotéomique <i>bottom-up</i> et identification	128
1.2.1	Préparation de l'échantillon.....	128
1.2.2	Analyse LC-MS/MS.....	128
1.2.3	Identification des glycoprotéines.....	130
1.3	Résultats	131
1.3.1	Traitement des résultats d'interrogation.....	131
1.3.2	Comparaison des méthodes de protéolyse	134
1.3.3	Fragmentation <i>stepped-HCD</i>	136
1.4	Glycosylation de la hCG et origine géographique	142
1.5	Conclusions	145

2. Analyse approfondie des données spectrales pour l'identification de glycoformes minoritaires.....	147
2.1 Introduction et problématique.....	147
2.2 Protocole de glycoprotéomique <i>bottom-up</i> et identification	148
2.2.1 Préparation de l'échantillon.....	149
2.2.2 Analyse LC-MS/MS.....	151
2.2.3 Identification glycoprotéique.....	152
2.2.4 Extraction des signaux d'intérêt.....	153
2.3 Résultats : Analyse des données spectrales	154
2.4 Conclusions	164
E. Synthèse et Discussion	165
F. Annexes.....	173
1. Monosaccharides et glycanes.....	175
2. Spectrométrie de masse.....	177
2.1 Notions en MS.....	177
2.2 Source d'ionisation.....	180
2.3 Analyseur.....	182
2.4 Modes de fragmentation (MS/MS).....	184
Article 1.....	187
Article 2.....	216
Bibliographie.....	246

Valorisation

PUBLICATIONS

S. Perchepped*, **N. Eskenazi***, C. Giangrande, J. Camperi, T. Fournier, J. Vinh, N. Delaunay, V. Pichon. "Development of Immobilized Enzyme Reactors for the Characterization of the Glycosylation Heterogeneity of a Protein." *Talanta* 206 (2020): 120171. <https://doi.org/10.1016/j.talanta.2019.120171>.

Article 1 (Annexe)

R. Thibeaux, M. E. Soupé-Gilbert, M. Kainiu, D. Girault, E. Bierque, J. Fernandes, H. Bähre, A. Douyère, **N. Eskenazi**, J. Vinh, M. Picardeau, C. Goarant. "The zoonotic pathogen *Leptospira interrogans* mitigates environmental stress through cyclic-di-GMP-controlled biofilm production" *Nature Microbiology*. *Article sous presse*.

Article 2 (Annexe)

N. Eskenazi, E. Demey, C. Giangrande, Y. Combarous, J. Vinh. "Sensitive N-Glycomics using One-pot Carbonyl-derivation MALDI matrix for Glyc hormone Characterization" *Article en cours de rédaction*.

Article 3

N. Eskenazi, E. Demey, C. Giangrande, Y. Combarous, J. Vinh. "Unspecific Proteolysis and Stepped-Energy MS² Fragmentation for the Mapping of Chorionic Gonadotropins Glycosylation between Geographically-distinct Populations" *Article en cours de rédaction*.

Article 4

N. Eskenazi, E. Demey, C. Giangrande, Y. Combarous, J. Vinh. "The Controlled Art of Destruction : Low-Abundance N-Glycopeptide Identification using MS² Oxonium Ions" *Article soumis*.

Article 5

CONFERENCES INVITEES

1. **N. Eskenazi**, J. Camperi, E. Demey, A. Combès, T. Fournier, J. Guibourdenche, V. Pichon, N. Delaunay, J. Vinh. *Journée Thématique Club IdF de l'AFSEP, Analytics & Biologics : de la cellule à la petite molécule*. Paris (France) 15 Mars 2018. Conférencier invité. "Comment analyser l'hétérogénéité structurale des glycoprotéines au niveau des peptides et des glycanes ?".
2. **N. Eskenazi**, C. Giangrande, O. Djimatey, J. Vinh. *Congrès Spectrométrie de Masse et Analyse Protéomique (SMAP 2019)*. Strasbourg (France) 16-19 Septembre 2019. Prix de la meilleure conférence des Clubs Jeunes de la SFSM et de la SFEAP. "Human Chorionic Gonadotropin glycoforms heterogeneity from two geographically distinct populations".

COMMUNICATIONS ORALES

3. **N. Eskenazi**, C. Giangrande, J. Camperi, S. Perchepied, A. Combès, J. Guibourdenche, V. Pichon, T. Fournier, N. Delaunay, J. Vinh. *Recent Developments in Pharmaceutical Analysis (RDPA 2017)* Rimini (Italy) 20-23 Septembre 2017. "Human chorionic gonadotropin glycoproteoforms characterization using hydrophilic affinity enrichment and mass spectrometry approaches".
4. **N. Eskenazi**, J. Camperi, T. Fournier, V. Pichon, N. Delaunay, J. Vinh. *Ecole thématique du CNRS sur la Spectrométrie de Masse à Transformée de Fourier (FTMS 2018)* Cabourg (France) 3-6 Avril 2018. "The combination of multiple proteases digestion for the structural characterization of glycosylated proteoforms detected in FTMS".
5. **N. Eskenazi**, J. Camperi, O. Djimatey, T. Fournier, V. Pichon, N. Delaunay, J. Vinh. *Congrès de la Société Française d'Electrophorèse et d'Analyse Protéomique et de la Société Espagnole de Protéomique (SFEAP 2018)* Albi (France) 9-12 Octobre 2018. "Comprehensive glycosylation mapping of a highly glycosylated hormone using high resolution mass spectrometry of glycoproteoforms and glycopeptides combined to matrix-assisted glycan population analysis".
6. **N. Eskenazi**, J. Camperi, Y. Combarous, V. Pichon, J. Vinh. *24èmes Rencontres du Club Jeunes de la Société Française de Spectrométrie de Masse (RCJSM 2019)* Quiberon (France) 18-22 Mars 2019. "nanoLC-MS/MS of glycoproteoforms and

proteolytic glycopeptides combined with glycan fingerprinting: a powerful strategy for glycoprotein characterization”.

POSTERS

- I. **N. Eskenazi**, C. Giangrande, J. Camperi, V. Pichon, N. Delaunay, J. Vinh. *Congrès Spectrométrie de Masse, Métabolomique et Analyse Protéomique (SMMAP 2017)* Marne-la-Vallée (France) 2-5 Octobre 2017. “Mass spectrometry and hydrophilic proteoform enrichment method-development for the characterization of Chorionic Gonadotropin glycosylation”.
- II. **N. Eskenazi**, G. Chiappetta, J. Camperi, T. Fournier, V. Pichon, N. Delaunay, Y. Combarous, J. Vinh. *1st European Top-Down Proteomics Symposium (TDP 2019)* Paris (France) 12-14 Février 2019. “nanoLC-MS/MS of glycoproteoforms and proteolytic glycopeptides combined with glycan fingerprinting: a powerful strategy for glycoprotein characterization”.
- III. **N. Eskenazi**, C. Giangrande, J. Vinh. *67th American Society for Mass Spectrometry Congress (ASMS 2019)* Atlanta (Géorgie, USA) 2-6 Juin 2019. “Identification of human Chorionic Gonadotropin glycoforms in two populations using improved bottom-up analysis”.
- IV. **N. Eskenazi**, O. Djimatey, C. Giangrande, J. Vinh. *67th American Society for Mass Spectrometry Congress (ASMS 2019)* Atlanta (Géorgie, USA) 2-6 Juin 2019. “N-glycan MALDI fingerprinting and all-in-one reducing-end derivation matrix optimization”.
- V. **N. Eskenazi**, O. Djimatey, J. Vinh. *47th IUPAC World Chemistry Congress (IUPAC 2019)* Paris (France) 5-12 Juillet 2019. “All-in-one N-glycan derivation matrix optimization for N-glycofingerprinting”.

Liste des figures

Figure 1 : Evolution du nombre de publications scientifiques référencées sur PubMed contenant les termes « protein glycosylation » depuis 1969.	28
Figure 2 : Formes linéaire et cyclique (hémiacétal) du glucose.	34
Figure 3 : Formes α -pyranosiques (hétérocycle hémiacétal à 6 atomes) des monosaccharides usuellement rencontrés dans les glycanes des mammifères.	34
Figure 4 : N-glycosylation, représentations SNFG du noyau (ou « core ») commun aux N-glycanes et des trois familles de N-glycanes. La représentation SNFG est explicitée au paragraphe 1.2 Nomenclature.	35
Figure 5 : Nomenclature graphique pour les glycanes, SNFG.	37
Figure 6 : Nomenclatures graphiques et textuelles, exemples représentatifs de N- et O-glycanes.	39
Figure 7 : Nomenclature des fragments des polysaccharides.	39
Figure 8 : Les glycoprotéines TSH, FSH, LH et CG.	42
Figure 9 : Gonadotrophine chorionique humaine.	43
Figure 10 : Synthèse du précurseur Dol-P-P-GlcNAc ₂ Man ₉ Glc ₃ à la surface externe puis dans le réticulum endoplasmique.	48
Figure 11 : Maturation d'un N-glycane dans le RE et l'appareil de Golgi.	50
Figure 12 : Noyaux des O-glycanes de type mucine.	52
Figure 13 : Voies de synthèse des cores O-GalNAc 1 à 4 et formes de Lewis.	53
Figure 14 : Trois approches complémentaires pour l'étude des glycoprotéines,	55
Figure 15 : Représentation schématique des éléments constituant un spectromètre de masse et leurs fonctions principales.	58
Figure 16 : Déglycosylation par la PNGase F, spécificité des sites de clivage.	62
Figure 17 : Structure de la 3-aminoquinoline.	65
Figure 18 : Représentation de deux fragments isobares issus de la fragmentation en MS/MS du sucre A2G2.	67
Figure 19 : Illustration de la séparation chromatographique des glycanes.	71
Figure 20 : Composition de la matrice liquide 3-AQ/CA,	74

Figure 21 : Comparaison visuelle de la forme des cristaux formés par différentes matrices MALDI.	74
Figure 22 : Diagramme d'une cellule de spectrométrie de mobilité ionique...	77
Figure 23 : Schématisation des approches bottom-up et top-down en protéomique.....	79
Figure 24 : La digestion d'une glycoprotéine par une protéase (par exemple, la trypsine) produit un mélange de peptides modifiés ou non.	82
Figure 25 : Nomenclature simplifiée des fragments peptidiques en spectrométrie de masse.....	84
Figure 26 : Etapes classiques d'un protocole de protéomique bottom-up (shotgun). (139).....	85
Figure 27 : Séquences protéiques des formes annotées des sous-unités α (GLHA_HUMAN)(a) et β (CGB3_HUMAN)(b) de la hCG.....	88
Figure 28 : Spectre de fragmentation HCD type (nanoLC-MS/MS) du glycopeptide CamCamCRPIN(+A2G2S1)ATLAVEK.....	93
Figure 29 : Différentes approches de protéomique quantitative, avec marquage métabolomique ou chimique, ou sans marquage.	99
Figure 30 : Séparation multi-dimensionnelle en protéomique top-down.	102
Figure 31 : Spectre MALDI-TOF MS, N-glycanes issus de la déglycosylation enzymatique de la RNase B, protocole de référence.	110
Figure 32 : Paramètres testés pour l'optimisation du protocole d'analyse des N-glycanes par MALDI-TOF avec une matrice liquide 3-AQ/p-CA.	112
Figure 33 : Optimisation du protocole, efficacité de dérivation.....	113
Figure 34 : Optimisation du protocole, sensibilité.....	114
Figure 35 : Spectre MALDI-TOF MS, N-glycanes issus de la déglycosylation enzymatique de la RNase B, protocole optimisé.	115
Figure 36 : Comparaison des spectres MALDI-TOF des N-glycanes de la hCG-CY obtenus avec la matrice 3-AQ/p-CA (haut) ou matrice DHB (bas).....	116
Figure 37 : Spectres MALDI-TOF MS des échantillons hCG-CY (haut) et hCG-BBT (bas), matrice 3-AQ/p-CA.	118
Figure 38 : Spectres MALDI-TOF MS des N-glycanes A1G1S1 (gauche), A2G2 (milieu) et A3G2NAcG1S2F (droite).	118

Figure 39 : Séquence en acides aminés de la hCG β (UniProt P0DN86 - CGB3_HUMAN).....	126
Figure 40 : Spectres illustrant la mauvaise identification de peptides multi-fucosylés.....	133
Figure 41 : Comparaison des glycopeptides identifiés pour chaque site de N-glycosylation avec différentes protéases, pour l'échantillon hCG-CY avec une NCE fixe.....	134
Figure 42 : Comparaison des glycopeptides identifiés pour chaque site de O-glycosylation avec différentes protéases, pour l'échantillon hCG-CY avec une NCE fixe.....	135
Figure 43 : Principe de la fragmentation stepped-HCD.	136
Figure 44 : Comparaison des fragmentations à énergie unique ou stepped-HCD pour l'identification des O-glycopeptides produits par une digestion par la pronase.....	136
Figure 45 : Spectre MS/MS du glycopeptide LPGPSDTPILPQ porteur du O-glycane HexNAc(2)Hex(3).	140
Figure 46 : Comparaison des glycopeptides identifiés pour chaque site de N- et de O-glycosylation entre deux échantillons d'origines géographiques distinctes.....	142
Figure 47 : Spectres de fragmentation HCD des glycopeptides NVTSESTCCVAK portant un N-glycane complexe ou high-mannose.....	149
Figure 48 : Histogramme des contributions à la variance totale de chaque composante principale.....	159
Figure 49 : Analyse de composante principale, ions oxoniums, clusterisation par famille.	161
Figure 50 : Analyse de composante principale, ions oxoniums, clusterisation par structure.....	162
Figure 51 : Analyse de composante principale, ions oxoniums, clusterisation par séquence peptidique.	162
Figure 52 : Massif d'intérêt détecté par MALDI-TOF.	216

Liste des tables

Tableau 1 : Fragmentation HCD type (nanoLC-MS/MS) du glycopeptide CamCamCRPIN(+A2G2S1)ATLAVEK.	95
Tableau 2 : Conditions de préparation de la matrice 3-AQ/p-CA testées.	113
Tableau 3 : N-glycanes représentatifs de la diversité des structures identifiées par MALDI-TOF sur le spectre précédent (Figure 36).	117
Tableau 4 : Structures O-glycanes identifiées pour le site O4 de la hCG-CY lors d'une digestion par la pronase, comparaison des méthodes de fragmentation à énergie constante ou stepped-HCD.	138
Tableau 5 : Comparaison des caractéristiques des glycanes identifiés uniquement dans un des deux échantillons.	144
Tableau 6 : Identification de glycopeptides par analyse des ions oxoniums.	155
Tableau 7 : Valeurs propres des composantes principales.	159
Tableau 8 : Vecteurs propres des composantes principales.	160
Tableau 9 : Table récapitulative des structures monosaccharidiques des glycoprotéines humaines.	175

Liste des abréviations

- 2-AB : 2-aminobenzamide
2-AP : 2-aminopyridine
3-AQ : 3-aminoquinoline
- ACN : Acétonitrile
ADC : Anticorps-molécule active (*Antibody-Drug Conjugate*)
ADCC : Cytotoxicité à médiation cellulaire dépendante des anticorps (*Antibody-Dependent Cellular Cytotoxicity*)
ADN : Acide désoxyribonucléique
ADP : Phosphate de monoammonium (*Ammonium Dihydrogen Phosphate*)
- Bottom-up : Approche ascendante
- CG(s) : Hormone(s) chorionique(s) gonadotrope(s)
CHCA : Acide α -cyano-4-hydroxycinnamique (*α -Cyano-4-HydroxyCinnamic Acid*)
CID : Dissociation induite par collision (*Collision-Induced Dissociation*)
CMP : Cytidine monophosphate
CTP : Peptide C-terminal des CGs (*C-Terminal Peptide*)
- DDA : *Data-dependent acquisition*
DHB : Acide 2,3-dihydroxybenzoïque (*2,3-DiHydroxyBenzoic acid*)
DIA : *Data-independent acquisition*
Dol-P : Dolichol-phosphate
Dol-P-P : Dolichol-pyrophosphate
DTT : Dithiothréitol
- ECD : Dissociation par capture d'électron (*Electron Capture Dissociation*)
eCG : Hormone chorionique gonadotrope équine
ERLIC : *Electrostatic Repulsion Hydrophilic Interaction Chromatography*
ESI : Ionisation électrospray (*ElectroSpray Ionization*)
ETD : Dissociation par transfert d'électron (*Electron Transfer Dissociation*)
- FA : Acide formique
FDR : Taux de faux-positifs (*False Discovery Rate*)
FSH : Hormone folliculo-stimulante (*Follicle-Stimulating Hormone*)
Fuc : Fucose
FWHM : Largeur à mi-hauteur (*Full Width at Half Maximum*)
- Gal : Galactose
GalNAc : N-acétylgalactosamine
GDP : Guanosine diphosphate
Glc : Glucose
GlcNAc : N-acétylglucosamine
GPI : Glycosylphosphatidylinositol
- HCD : Dissociation induite par collision à haute énergie (*Higher-energy Collisional Dissociation*)

hCG : Hormone chorionique gonadotrope humaine
Hex : Hexose
HexNAc : N-acétylhexosamine
HILIC : *Hydrophilic Interaction Liquid Chromatography*
IRMPD : Dissociation multiphotonique infrarouge (*InfraRed MultiPhoton Dissociation*)
IAM : Iodoacétamide
ILM : Matrice liquide ionique (*Ionic Liquid Matrix*)
IMS : Spectrométrie de mobilité ionique (*Ion-Mobility Spectrometry*)
IUPAC : Union Internationale de Chimie Pure et Appliquée (*International Union of Pure and Applied Chemistry*)
(HP)LC : Chromatographie liquide à haute performance (*High Performance Liquid Chromatography*)
LH : Hormone lutéinisante (*Luteinizing Hormone*)
MALDI : Ionisation/désorption Laser assistée par matrice (*Matrix-Assisted Laser Desorption/Ionization*)
Man : Mannose
MS : Spectrométrie de masse (*Mass Spectrometry*)
m/z : masse / charge
NeuAc : Acide N-acétylneuraminique
NeuGc : Acide N-glycolylneuraminique
NPLC : Chromatographie liquide en phase normale, *Normal-Phase Liquid Chromatography*
pCA : Acide para-coumarique (*para-Coumaric Acid*)
PEP : *Posterior Error Probability*
PGC : *Porous Graphitized Carbon*
PSD : Fragmentation post-source (*Post-Source Decay*)
PSM : Spectre attribué à un peptide (*Peptide-Spectrum Match*)
PTM(s) : Modification(s) post-traductionnelle(s) (*Post-Translational Modification*)
RPLC : Chromatographie liquide en phase inverse, *Reversed-Phase Liquid Chromatography*
SNFG : *Symbol Nomenclature for Glycans*
TFA : Acide trifluoroacétique
TOF : (spectromètre de masse à) Temps de vol (*Time-of-Flight*)
Top-down : Approche descendante
TSH : Thyroïdostimuline (*Thyroid-stimulating hormone*)
UDP : Uridine diphosphate
UVPD : Photodissociation ultraviolette (*UltraViolet PhotoDissociation*)

A. Introduction

La glycosylation est un processus enzymatique consistant à former une liaison covalente entre un glycane, ou polysaccharide, et une protéine ou une autre biomolécule. Il s'agit d'une des modifications protéiques les plus communes puisqu'on estime qu'au moins la moitié des protéines circulantes sont glycosylées, et l'une des plus importantes de par son implication dans de nombreuses réactions biochimiques. Toutefois, il existe une très grande hétérogénéité au sein de ces modifications, en termes de composition, de structure, ou de site de liaison à la protéine. De plus, si l'ADN semble contenir une information majoritairement immuable, les protéines pour lesquelles il code sont en constante dynamique. Les interactions entre ces biomolécules ou les modifications qu'elles peuvent subir au niveau de leur conformation, de leur composition ou de leur état d'oligomérisation sont en effet responsables d'un grand nombre de fonctions biologiques au sein de tout organisme vivant.

Dans le cadre de cette étude, nous nous sommes plus particulièrement intéressés à l'hétérogénéité de certaines glycohormones comme l'hormone chorionique gonadotrope, qui est une glycoprotéine produite par un nombre restreint d'organismes, en particulier chez les primates dont l'espèce humaine, et les équidés. Il s'agit d'une hormone sécrétée à de fortes concentrations dès les premiers mois après la fécondation et fortement impliquée dans le bon déroulement de la grossesse chez les femelles de ces espèces. Son activité biologique est liée à son interaction avec des récepteurs transmembranaires présents dans la région gonado-utérine, déclenchant une cascade signalétique aboutissant à la synthèse d'hormones stéroïdes nécessaires au maintien précoce de l'embryon jusqu'à la maturité du placenta. Elle est donc tout à fait centrale dans les problématiques contemporaines de fécondation *in vitro* ou pour la médecine vétérinaire des animaux d'élevage (équins, bovins, ovins, caprins). De plus, la glycosylation de la hCG joue un rôle dans la capacité de transduction du signal de son récepteur et dans les fonctions des cellules cibles. Cependant, et bien que la concentration de ces protéines soit précisément mesurée durant la grossesse ou les protocoles cités précédemment, leur état de glycosylation n'est pas couramment recherché, probablement du fait du manque d'outils

pour explorer la complexité de leurs différentes glycoformes. Il est nécessaire de faire progresser l'état des connaissances concernant ces structures afin d'élucider leur importance dans la fonction biologique des gonadotrophines.

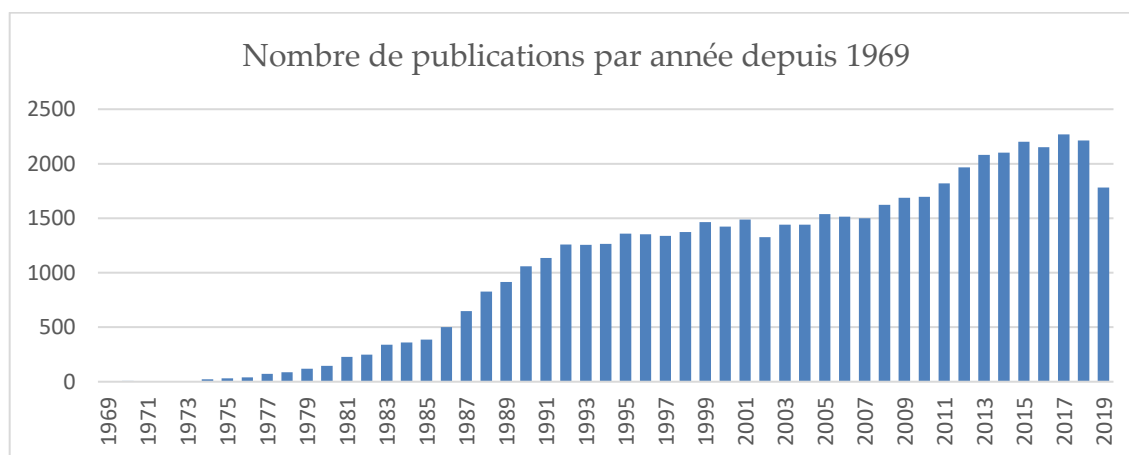


Figure 1 : Evolution du nombre de publications scientifiques référencées sur PubMed contenant les termes « protein glycosylation » depuis 1969. (<https://www.ncbi.nlm.nih.gov/pubmed/>) Données consultées le 2 octobre 2019.

Les stratégies d'études mises en œuvre dans cette étude reposent sur la spectrométrie de masse et dérivent des approches développées en protéomique. La spectrométrie de masse est une des techniques permettant l'identification de ces molécules. Le principe de la méthode repose sur la séparation dans un champ électrique ou magnétique de molécules ionisées en phase gazeuse selon leur rapport masse / charge. Elle permet de caractériser la structure chimique de molécules en mélange sur la base de la mesure de leurs masses et de celles de leurs fragments. L'étude de l'hétérogénéité structurale des glycoformes de ces hormones est un défi technologique considérable et néanmoins une préoccupation majeure de santé publique. L'analyse des glycoprotéines est un domaine de recherche s'étant développé dans les dernières décennies du XX^{ème} siècle, et qui a connu un regain d'intérêt récent concomitant avec les avancées technologiques et l'arrivée de solutions commerciales en spectrométrie de masse à haut débit et haute résolution (Figure 1).

Ce projet a pour but de caractériser les glycoformes des gonadotrophines humaine et équine, ce qui a nécessité le développement de méthodologies

performantes pour leur détection et leur identification. Les travaux présentés dans ce manuscrit portent sur les développements analytiques en glycomique et glycoprotéomique mis en place pour l'étude des gonadotrophines chorioniques. Ils seront précédés d'une revue de l'état de la recherche sur la glycosylation des choriogonadotropines et les méthodes de spectrométrie de masse pour la glycobiologie (Première Partie), et les résultats du projet seront séparés entre l'approche glycomique (Deuxième Partie) et l'approche glycoprotéomique *bottom-up* (Troisième Partie). Des méthodes robustes et sensibles de détection et d'identification de ces espèces sont nécessaires, notamment pour leur transposition en milieu médical pour le diagnostic et le développement de molécules biothérapeutiques.

B. Première Partie :

Etat de l'art

1. Glycosylation des protéines et glycohormones

1.1 Introduction

Les protéines subissent, en aval de leur synthèse, des modifications chimiques souvent régies par des réactions enzymatiques, appelées « modifications post-traductionnelles » ou PTMs (de l'anglais « *Post-Translational Modifications* »). Ce terme désigne également, par extension, les modifications survenant lorsque la chaîne peptidique est encore en cours de synthèse, et qui sont donc à proprement parler des modifications co-traductionnelles. Il existe de nombreuses PTMs différentes, de par la nature chimique des modifications ou leur localisation sur la protéine. Pas moins de 1493 modifications covalentes sont répertoriées dans la base de données de modifications publique Unimod en 2019. Ces modifications incluent, entre autres, le clivage de liaisons peptidiques (peptide signal, peptide C de la forme immature de l'insuline), et la modification covalente d'acides aminés particuliers par des groupements de natures variées, tels des phosphates (phosphorylation), groupements acétyle (acétylation) ou des polysaccharides (glycosylation et glycation). Elles peuvent être réversibles, alors souvent associées à des processus biologiques de régulation ou de signalisation, ou bien être irréversibles et dans ce cas potentiellement associées à des phénomènes de vieillissement cellulaire ou de dégradation (1). La glycosylation est un processus enzymatique (contrairement à la glycation qui est un processus passif) aboutissant à la liaison covalente d'une protéine et d'un mono- ou oligosaccharide. Il s'agit d'une famille de PTMs extrêmement courante retrouvée chez les organismes eucaryotes, pour lesquels on estime que la majorité des protéines sécrétées sont glycosylées (2), mais aussi sur les protéines synthétisées par les organismes du domaine bactérien (3), contrairement à ce qui fut longtemps admis. Ces modifications ont un impact sur la stabilité, la conformation, la fonction et les capacités d'interaction des glycoprotéines.

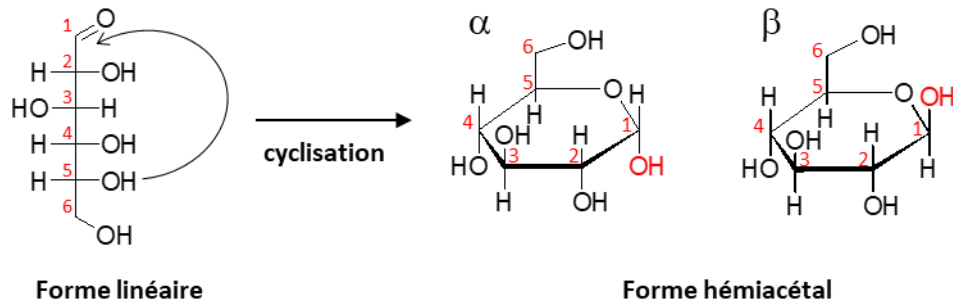


Figure 2 : Formes linéaire et cyclique (hémiacétal) du glucose.

L'orientation du C1 au moment de la cyclisation peut aboutir à deux formes anomériques α ou β selon la position du groupement hydroxyle 1 (en rouge) par rapport au cycle.

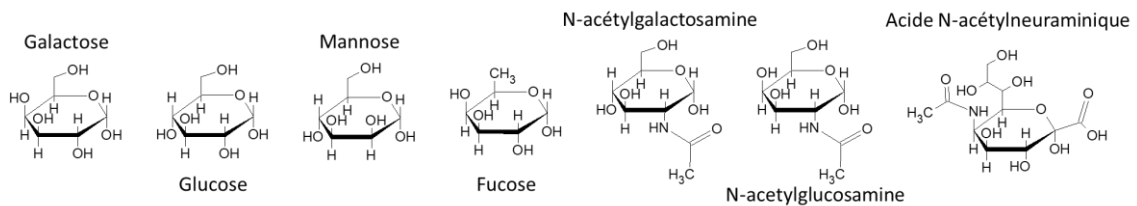


Figure 3 : Formes α -pyranosiques (hétérocycle hémiacétal à 6 atomes) des monosaccharides usuellement rencontrés dans les glycanes des mammifères.

Les monosaccharides existent sous deux formes, linéaire ou cyclique (Figure 2). Lorsqu'ils adoptent une forme cyclique, ce sont des molécules hétérocycliques (leur cyclisation provoque l'intégration d'un hétéroatome d'oxygène au cycle carboné), poly-hydroxylées et chirales. L'orientation des groupements hydroxyles sur chaque carbone stéréogène (à environnement asymétrique) a un tel impact sur les propriétés du monosaccharide que chaque diastéréoisomère possède une identité et un nom qui lui est propre (Figure 3). Chaque monosaccharide sous forme linéaire possède un unique groupement aldéhydique dont la réactivité est conservée lors de la cyclisation (Carbone-1 hémiacétal). Ce carbone étant impliqué dans la liaison de formation des oligosaccharides, la majorité des polymères d'oses n'ont qu'une extrémité réductrice, généralement impliquée dans la liaison à la protéine glycosylée. Un polysaccharide, aussi appelé glycan, est le nom donné au polymère linéaire ou ramifié composé de monosaccharides (ou oses) reliés entre eux par des liaisons covalentes, dites « osidiques ».

Différents types de glycosylation sont décrits en fonction du groupement chimique modifié ou de la molécule intermédiaire entre le polysaccharide à la

protéine glycosylée. Chaque famille de glycosylation utilise des voies de synthèse différentes, aboutissant à des structures osidiques variées mais également à des fonctions biologiques propres. La N-glycosylation et la O-glycosylation sont les formes les plus étudiées et sont les deux formes de glycosylation identifiées sur les hormones gonadotropes, tandis que d'autres types de glycosylation moins courants existent, comme la C-mannosylation, la glypiation et la glycosylation des phosphosérines.

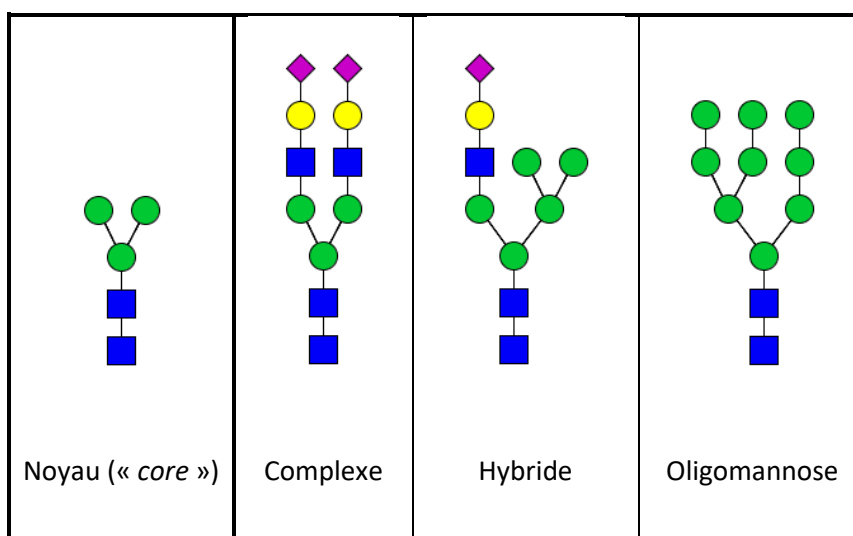


Figure 4 : N-glycosylation, représentations SNFG du noyau (ou « core ») commun aux N-glycanes et des trois familles de N-glycanes.

La représentation SNFG est explicitée au paragraphe 1.2 Nomenclature.

La N-glycosylation chez les organismes eucaryotes est la plus étudiée de toutes les glycosylations. Elle implique la modification d'un résidu asparagine (Asn) présent au sein d'une séquence particulière Asn-X-Ser/Thr par un N-glycane (X étant un acide aminé quelconque sauf une proline). Les N-glycanes ont en commun un noyau pentasaccharidique composé de deux N-acétylglucosamines et trois mannosés (appelé « core »), lesquels peuvent être différemment modifiés et étendus. La famille des oligomannoses (ou « high-mannose ») regroupe les N-glycanes constitués du core et d'un nombre variable de mannosés supplémentaires (au nombre total généralement supérieur à cinq), les formes dites « complexes » sont modifiées par un nombre variable d'antennes trisaccharidiques (composée d'une GlcNAc, d'un Gal et d'un NeuAc), et les formes hybrides sont intermédiaires et sont composées d'au

moins une antenne et d'un mannose sur leur *core* (Figure 4). De nombreuses protéines sont N-glycosylées, les plus étudiées étant les immunoglobulines de par leur utilisation biothérapeutique sous forme notamment d'anticorps monoclonaux ou d'ADC (« *Antibody-Drug Conjugate* » ou médicament couplé à un anticorps).

Le terme « O-glycosylation » désigne la modification d'un résidu thréonine (Thr) ou sérine (Ser) d'une protéine par un O-glycane au niveau du groupement hydroxyle libre de leur chaîne latérale. Contrairement aux sites de N-glycosylation, les sites de O-glycosylation ne nécessitent pas d'être impliqués dans une séquence consensus particulière. Les mucines sont des protéines ubiquitaires fortement O-glycosylées situées à la surface des épithéliums de nombreux organismes ayant un rôle important dans leur fonction de barrière. Leurs O-glycane sont amorcés par une liaison commune O-GalNAc qui constitue la structure commune à tous les noyaux de type mucine, et sont allongés et ramifiés par l'ajout d'un nombre variable de monosaccharides. Le noyau O-GlcNAc est une forme moins courante de O-glycosylation surtout détectée sur les protéines nucléaires, mitochondriales et cytoplasmiques. Cette modification, très dynamique, n'est constituée que du *core* GlcNAc et permet de prévenir la phosphorylation de ces protéines et réguler leur activité. Ainsi, les protéines O-acétylglucosaminées ont souvent un rôle dans l'adaptation au stress cellulaire, la régulation de phénomènes épigénétiques et de la transcription, ainsi que dans la production, la stabilité et la dégradation des protéines.

Les autres familles moins documentées de glycosylation incluent la C-mannosylation, une glycosylation située au niveau d'un atome de Carbone d'un résidu tryptophane (Trp) (la sous-unité α de la protéine C8 du complément est un exemple de protéine C-mannosylée), la glypiation, qui implique l'ancrage d'une protéine à un lipide membranaire par l'intermédiaire d'un glycosylphosphatidylinositol (GPI), et la modification de résidus phosphosérines. La glycation quant à elle est un processus non-enzymatique

impliquant la réaction entre le groupement carbonyle de molécules osidiques avec une fonction amine d'une protéine. L'exemple le plus courant de protéine glyquée est l'hémoglobine, dosée dans les cas de suspicion ou de diabète avéré.

Du fait de la grande hétérogénéité au sein même des glycanes, il n'est pas rare d'identifier plus de 10 structures différentes sur un unique site de glycosylation d'une protéine. De ce fait, le nombre de combinaisons possible pour une protéine multi-glycosylée (comme la hCG qui possède 8 sites de glycosylation, ou la eCG qui en possède 15) peut théoriquement être supérieur à certaines estimations du nombre total de protéoformes dans un type cellulaire (4). Ces considérations doivent guider le développement de méthodes d'analyse qui seront nécessairement capables d'explorer l'immense hétérogénéité de ces modifications tout en prenant en compte l'effet combiné de celles-ci.

1.2 Nomenclature

SHAPE	White	Blue	Green	Yellow	Orange	Pink	Purple	Light Blue	Brown	Red
Filled Circle	Hexose ○	Glc ●	Man ●	Gal ●	Gul ●	Alt ●	All ●	Tal ●	Ido ●	
Filled Square	HexNAc □	GlcNAc ■	ManNAc ■	GalNAc ■	GulNAc ■	AltNAc ■	AllNAc ■	TalNAc ■	IdoNAc ■	
Crossed Square	Hexosamine ◻	GlcN ◻	ManN ◻	GalN ◻	GulN ◻	AltN ◻	AllN ◻	TalN ◻	IdoN ◻	
Divided Diamond	Hexuronate ◇	GlcA ◇	ManA ◇	GalA ◇	GulA ◇	AltA ◇	AllA ◇	TalA ◇	IdoA ◇	
Filled Triangle	Deoxyhexose △	Qui ▲	Rha ▲		6dGul ▲	6dAlt ▲		6dTal ▲		Fuc ▲
Divided Triangle	DeoxyhexNAc ◡	QuiNAc ◡	RhaNAc ◡			6dAltNAc ◡		6dTalNAc ◡		FucNAc ◡
Flat Rectangle	Di-deoxyhexose ▭	Oli ▭	Tyv ▭		Abe ▭	Par ▭	Dig ▭	Col ▭		
Filled Star	Pentose ☆		Ara ★	Lyx ★	Xyl ★	Rib ★				
Filled Diamond	Deoxynonulosonate ◇		Kdn ◇				Neu5Ac ◇	Neu5Gc ◇	Neu ◇	Sia ◇
Flat Diamond	Di-deoxynonulosonate ◇		Pse ◇	Leg ◇		Aci ◇		4eLeg ◇		
Flat Hexagon	Unknown ⬡	Bac ⬡	LDManHep ⬡	Kdo ⬡	Dha ⬡	DDManHep ⬡	MurNAc ⬡	MurNGc ⬡	Mur ⬡	
Pentagon	Assigned ⬠	Api ⬠	Fruc ⬠	Tag ⬠	Sor ⬠	Psi ⬠				

Figure 5 : Nomenclature graphique pour les glycanes, SNFG.
Figure originale disponible à l'adresse <https://www.ncbi.nlm.nih.gov/glycans/snfg.html>

Les glycanes pouvant être ramifiés et de structure complexe, une nomenclature graphique appelée SNFG pour « *Symbol Nomenclature for Glycans* » s'est imposée au fil des années avec sa première édition publiée en 1999 dans la revue « *Essentials of Glycobiology* », révisée en 2004 et en 2015 pour sa troisième édition (5). Elle consiste en l'identification de chaque monosaccharide par des formes géométriques colorées, la forme géométrique indiquant généralement la structure du monomère, tandis que la couleur est une indication de sa stéréochimie (Figure 5). Aussi, les couleurs et les formes sont cohérentes entre les oses les plus communs : un cercle jaune représente un galactose (Gal), un carré jaune une N-acétylgalactosamine (GalNAc) et un carré bleu une N-acétylglucosamine (GlcNAc).

La nomenclature textuelle est aussi définie par le SNFG pour la représentation des structures polysaccharidiques dont les liaisons glycosidiques sont connues. Il s'agit d'une version condensée de l'écriture recommandée par l'IUPAC représentant de façon linéaire ou ramifiée la composition des glycanes et les liaisons glycosidiques entre ses constituants monosaccharidiques, en écrivant dans le sens « extrémité(s) non réductrice(s) » → « extrémité réductrice ». Un trisaccharide linéaire pourra ainsi s'écrire NeuAc α 2-3Gal β 1-3GalNAc. Si ses liaisons ou sa stéréochimie ne sont pas intégralement caractérisées, le sucre sera représenté graphiquement en utilisant les symboles SNFG non colorés ou en détaillant uniquement la composition du sucre, NeuAc₁Gal₁GalNAc₁ ou HexNAc(1)Hex(1)NeuAc(1) dans le cas du sucre précédent.

Les structures des N-glycanes étant très conservées, la notation Oxford peut être utilisée car plus compacte : Pour les formes high-mannose on indique le nombre total de mannoses « **K** » sous la forme MK (par exemple, M9 correspond à un N-glycane à neuf mannoses), tandis que les formes complexes suivent la forme générale « **FAXBGYSZ** », avec « **X** » le nombre d'antennes, « **Y** » le nombre de galactoses et « **Z** » le nombre d'acides sialiques (acide N-acétylneuraminique, NeuAc ou NANA) terminaux. L'utilisation des lettres « **F** » et « **B** » indique respectivement la présence d'un fucose à la base du N-

glycane, ou la bissection du *core* par une N-acétylglucosamine (située sur le mannose lié en β 1-4. Les formes hybrides adoptent une notation similaire, indiquant à la fois le nombre total de mannoses et la composition des antennes (Figure 6).

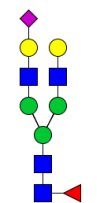
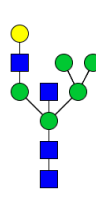
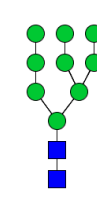
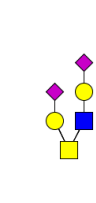
Nomenclature IUPAC	NeuAc α 2-6 Gal β 1-4 GlcNAc β 1-2 Man α 1-3 (Gal β 1-4 GlcNAc β 1-2 Man α 1-6) Man β 1-4 GlcNAc β 1-4 (Fuc α 1-6) GlcNAc	GlcNAc β 1-4 (Gal β 1-4 GlcNAc β 1-2 Man α 1-3) (Man α 1-3 (Man α 1-6) Man α 1-6) Man β 1-4 GlcNAc β 1-4 GlcNAc	Man α 1-2 Man α 1-2 Man α 1-3 (Man α 1-2 Man α 1-3 (Man α 1-2 Man α 1-6) Man α 1-6) Man β 1-4 GlcNAc β 1-4 GlcNAc	NeuAc α 2-3 Gal β 1-3 (NeuAc α 2-3 Gal β 1-4 GlcNAc β 1-6) GalNAc
Représentation graphique				
Famille	N-glycane complexe	N-glycane hybride	N-glycane oligomannose	O-glycane
Composition	HexNAc(4) Hex(5) Fuc(1) NeuAc(1)	HexNAc(4) Hex(6)	HexNAc(2) Hex(9)	HexNAc(2) Hex(2) NeuAc(2)
Nomenclature Oxford	FA2G2S1	M5A1BG1	M9	/

Figure 6 : Nomenclatures graphiques et textuelles, exemples représentatifs de N- et O-glycane.

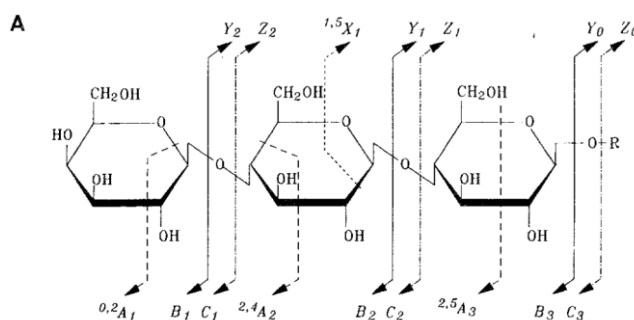


Figure 7 : Nomenclature des fragments des polysaccharides.

Figure issue de « A Systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS/MS Spectra of Glycoconjugates » par Bruno Domon et Catherine E Costello (Glycoconjugate Journal, 1988)

La nomenclature proposée par B. Domon et C. Costello (6) pour désigner les ions produits lors de la fragmentation en spectrométrie de masse en tandem est universellement acceptée par la communauté et sera utilisée dans ce manuscrit (Figure 7). Elle suit des règles analogues à celles de la nomenclature des fragments peptidiques (7) : Les fragments B, C, Y et Z sont des fragments issus de la rupture de la liaison osidique, tandis que les fragments A et X sont issus

de la fragmentation au sein d'un hétérocycle (fragments dits « *cross-ring* »). Les fragments A_i , B_i et C_i désignent les fragments contenant l'extrémité non réductrice du glycanes, et les fragments X_j , Y_j et Z_j , l'extrémité réductrice ; le nombre en indice indique la position relative de la liaison rompue par rapport à l'extrémité du sucre ; le nombre en exposant indique, le cas échéant, les liaisons intrasaccharidiques rompues.

1.3 Fonctions biologiques

Du fait de leur omniprésence sur la majorité des protéines et de leur grande hétérogénéité, les glycanes sont susceptibles de remplir des fonctions extrêmement diverses. Les glycanes peuvent être impliqués dans le repliement des protéines dans le réticulum endoplasmique, le caractère hydrophile et la solubilité des protéines, leur stabilité, l'adhésion cellulaire, ou encore le caractère pro- ou anti- inflammatoire ou immunogène de certaines protéines ou cellules. Ces propriétés biologiques sont généralement liées au caractère hydrophile, à la densité de charge ou à la reconnaissance des glycoconjugués par des récepteurs membranaires.

Les glycoprotéines les plus étudiées pour leur fonction biologique sont probablement les immunoglobulines (de classe IgG) du fait de leur fort intérêt pharmaco-médical. En effet la glycosylation des immunoglobulines a un rôle structural et modulateur des fonctions effectrices de l'anticorps. Les anticorps ont un site de N-glycosylation conservé en position N297 de leur Fc (fragment cristallisable) et parfois un site de glycosylation dans leur région variable, les deux modifications étant retrouvées sur la chaîne lourde de l'immunoglobuline (8). Sur ces glycoprotéines, les N-glycanes sont majoritairement des formes complexes biantennaires, sialylées et fucosylées. La simple absence de N-glycanes est associée à un taux d'agrégation plus important et une sensibilité aux traitements thermiques, chaotropiques (Guanidine-HCl), ou protéasiques (papaine) augmentée, et donc une plus courte période de stabilité (9). La conformation du domaine CH2 des anticorps est particulièrement affectée par leur glycosylation, ainsi que leur affinité pour

les récepteurs anti-Fc et les protéines du complément. Au-delà de la simple présence de N-glycanes, la composition de ceux-ci impacte également les fonctions effectrices des anticorps. En particulier, la présence de GlcNAc en position bissectrice a un effet pro-inflammatoire (activation de la voie ADCC), tandis que la fucosylation du *core* augmente l'activation du complément. Les Gal et NeuAc terminaux sont quant à eux anti-inflammatoires (10). Dans le cadre du design d'anticorps thérapeutiques ou d'ADCs (anticorps couplé à une molécule active), la maîtrise de leur glycosylation permet donc d'optimiser leurs propriétés. (11,12)

L'état anormal de glycosylation de certaines protéines est un promoteur de la progression et de l'évolution de cellules cancéreuses, et donc un excellent biomarqueur de ce type d'anomalie. Les modifications que subissent les N-glycanes et les O-glycanes sont les plus décrites, probablement du fait de la meilleure connaissance de ces familles. Du fait de l'altération des niveaux d'expression des enzymes liées à la glycosylation dans ces cellules, on observe alors des modifications structurales des *cores* O-glycosidiques et de l'état de ramification et de complexité des chaînes polysaccharidiques. Plusieurs motifs semblent conservés et associés à différentes formes de cancer. Notamment, la présence de structures O-glycosidiques particulières (antigènes T, Tn, STn) a été décrite dans des cas de cancers colorectaux (13,14), pour lesquels une augmentation de la concentration des mucines sialylées (α 2-6) a également été observée et associée à la formation de métastases et l'échec des traitements (15,16). L'extension des antennes des N-glycanes complexes par des répétitions Gal-GlcNAc (poly-N-acétyllactosamine, ou polyacNAc) est également un marqueur de cancer pulmonaire et colorectal (17). Les polyacNAc ont en effet une haute affinité pour la galectine-3 qui est connue pour promouvoir la formation de métastases pulmonaires (18). L'état de glycosylation de la β -hCG semble lié au développement de cellules tumorales, les formes hyperglycosylées de la hCG étant plus prépondérantes chez les patients touchés par ces cancers (19). (20)

1.4 Les hormones glycoprotéiques gonadotropes

Quatre hormones hétérodimériques partagent une sous-unité α commune, les hormones pituitaires lutéinisante (LH), folliculostimulante (FSH), la thyroïdostimuline (TSH) et uniquement chez les primates et les équidés la gonadotrophine chorionique (CG) sécrétée directement par l'embryon puis le placenta (Figure 8). La nature et la fonction de la forme dimérique de ces hormones sont déterminées par leur sous-unité β , dont la séquence et le nombre de sites de glycosylation sont spécifiques à chaque sous-unité.

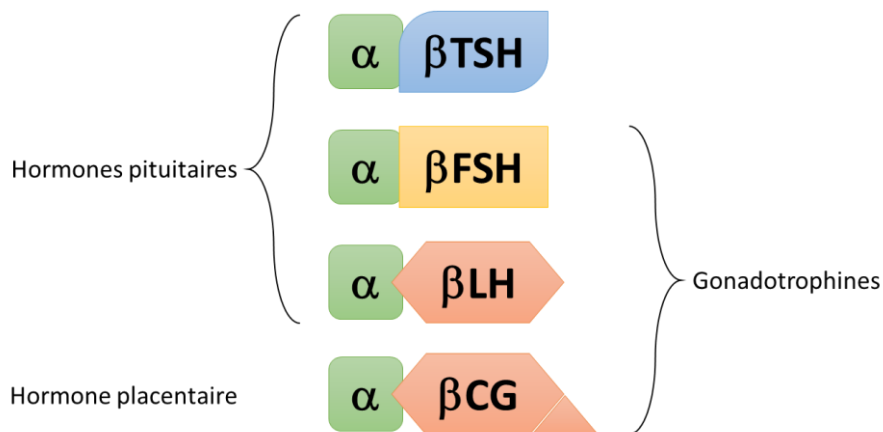


Figure 8 : Les glycoprotéines TSH, FSH, LH et CG.

Elles partagent une même sous-unité alpha. La séquence des sous-unités beta définit l'activité de l'hormone. TSH, FSH, et LH sont exprimées dans l'hypophyse, hCG est exprimée par le placenta.

La grossesse chez les mammifères implique les hormones gonadotropes LH, FSH, et CG (uniquement chez les espèces concernées, soit les primates et les équidés), aux fonctions conservées au sein des différentes espèces. La FSH agit au début du cycle menstruel (après la dégénérescence des cellules utérines en cas de non fécondation) pour stimuler la croissance des cellules folliculaires et préparer la production de l'ovule. L'hormone LH est produite à un niveau basal pour stimuler la production d'œstradiol au début du cycle menstruel puis est sécrétée en très grande quantité environ 14 jours après le début du cycle menstruel pour déclencher l'ovulation. Le corps jaune (follicule après émission de l'ovule) permet la stabilisation des cellules utérines grâce à la sécrétion de progestérone, pour une durée d'environ deux semaines après l'ovulation. En l'absence de fécondation, le corps jaune entre en apoptose et la production de progestérone décroît, déclenchant la menstruation. Dès les premiers jours après

la fécondation chez la femme, le trophoblaste (partie de l'œuf qui évoluera en placenta au cours du développement embryonnaire) sécrète la hCG qui empêche la dégénérescence du corps jaune et maintient la production de progestérone par ce dernier. Une fois le placenta formé, ce dernier assurera la production de progestérone et le maintien de la paroi utérine.

1.4.1 La gonadotrophine chorionique humaine (hCG)

La choriogonadotropine (CG) est l'hormone responsable du maintien du corps jaune, apparue uniquement chez les équidés et les primates (dont l'Homme) au cours de l'évolution. C'est une hormone active uniquement sous sa forme dimérique ($\alpha\beta$) sécrétée par les cellules trophoblastiques rapidement après la fécondation de l'ovule et pendant le développement de l'œuf. Du fait de sa similarité structurale avec la LH, hCG interagit avec le même récepteur (LHCGR) au niveau des cellules du corps jaune pour le maintenir et induire la production de progestérone (21), hormone nécessaire au maintien et au développement du fœtus. La hCG est ainsi impliquée au cours de la grossesse dans la vascularisation de l'endomètre (22) et le caractère invasif du trophoblaste extravilleux (23), ainsi que l'immunotolérance locale de la mère pour l'embryon (24).

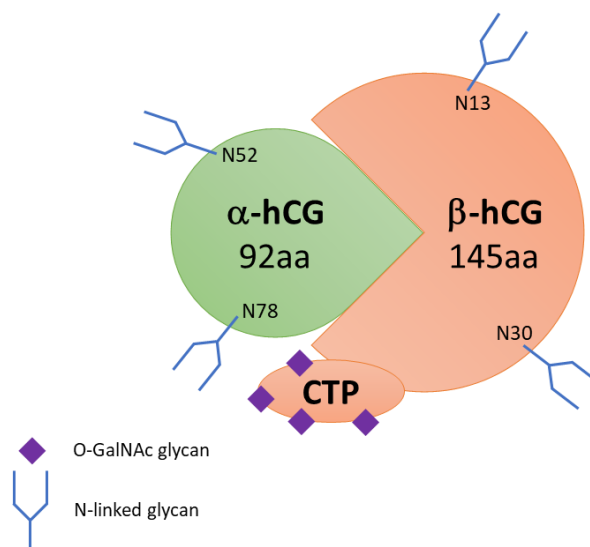


Figure 9 : Gonadotrophine chorionique humaine.

Schéma montrant les N-glycanes et O-glycanes. Les polysaccharides O-GalNAc sont situés sur les résidus S121, S127, S132, S138).

Cette hormone est une glycoprotéine polymorphe hétérodimérique (Figure 9), composée d'une sous-unité α (α -hCG, commune à la LH, la FSH et la TSH) et de la sous-unité β spécifique à chaque hormone (β -hCG). Plusieurs mutations des régions codantes existent naturellement, aboutissant à différentes isoformes (formes tronquées, modification du peptide signal). La portion conservée de ces isoformes après maturation conserve néanmoins un fort taux d'homologie de séquence. La sous-unité α mature (entrée « GLHA_HUMAN » de la banque de données annotée Swiss-Prot) est une protéine de 92 acides aminés qui possède deux asparagines potentiellement N-glycosylées (Asn-52 et Asn-78). La fraction purement protéique de la α -hCG a une masse moléculaire de 10,2 kDa, tandis qu'en fonction de son état de glycosylation la masse de la sous-unité mature peut varier d'environ 13 kDa à plus de 15 kDa, soit une part osidique d'environ 30% de la masse totale. La sous-unité β (entrées « CGB2_HUMAN » ou « CGB3_HUMAN » de Swiss-Prot, l'isoforme 3 correspondant à une mutation ponctuelle A117→D) quant à elle, est constituée de 145 acides aminés pour une masse non glycosylée de 15,5 kDa. Elle possède 2 sites de N-glycosylation (Asn-13 et Asn-30) ainsi que 4 sites de O-glycosylation caractérisés sur son CTP (Ser-121, Ser-127, Ser-132 et Ser-138). Ainsi, la masse de la β -hCG mature est généralement comprise entre 21 kDa et 25 kDa, mais des formes hyperglycosylées peuvent augmenter sensiblement la masse de cette sous-unité.

L'état de glycosylation de la hCG (celui de la sous-unité α en particulier) est lié à l'association correcte des différentes sous-unités des protéines multimériques dans le réticulum endoplasmique (25). La demi-vie des glycohormones dans la circulation sanguine est également impactée par l'abondance des N- et des O-glycanes présents sur la protéine. Il apparaît que la demi-vie de ces hormones est fortement augmentée entre les formes pituitaires (5 à 30 minutes) et les formes placentaires possédant un peptide C-terminal glycosylé (2 jours) (26). En effet, la production de formes recombinantes de ces protéines avec ou sans CTP confirme ce rôle dans l'élimination urinaire des hormones. Enfin, bien que la reconnaissance hormone-récepteur ne dépende pas de la glycosylation de la

hCG, l'activation de son activité adénylate cyclase et la cascade de signalisation qui s'en suit nécessitent la forme glycosylée de la hCG (27). Les formes hyperglycosylées de la hCG (hCG-H) sont cependant associées à une plus faible activation des récepteurs (28) et sont majoritairement sécrétées par les cellules invasives du trophoblaste extravilleux (29).

L'hormone hCG est injectée aux femmes dans le cadre de certains traitements contre l'infertilité ou pour déclencher l'ovulation préalablement à des protocoles de procréation médicalement assistée, ou chez l'homme pour stimuler la spermatogénèse. En France et dans le monde, deux types d'hCG sont utilisés : la forme urinaire purifiée à partir d'urines de femmes enceinte (u-hCG), et la forme recombinante (r-hCG), de séquence protéique identique, exprimée dans des systèmes mammifères comme les cellules ovariennes de hamster *C. griseus* (cellules CHO). Ces deux formes thérapeutiques sont équivalentes en termes d'efficacité de stimulation, et sont utilisées à la discrétion du clinicien en charge des protocoles entrepris (30,31), bien que le produit recombinant permette une diminution des effets indésirables au niveau du site d'injection (32). Cette propriété peut être due à la meilleure pureté de la r-hCG, mais également aux différentes glycoprotéoforèmes produites par l'Homme (u-hCG) et le hamster (r-hCG), que l'on sait être liées aux propriétés pro- et anti-inflammatoires des IgG notamment.

Bien que cette protéine, utilisée pour provoquer l'ovulation dans des traitements de fertilité ou préliminairement à un protocole de fécondation *in vitro*, soit dosée au cours de la grossesse, seule sa concentration est habituellement suivie. En effet, cette dernière est un indicateur du bon déroulement de la gestation puisque la concentration élevée en hCG peut être un biomarqueur de grossesses multiples, ectopiques, ou de syndrome de Down (33). Cependant, la glycosylation de la hCG n'est pas suivie alors que son importance dans l'activité hormonale est connue et associée au développement de cellules cancéreuses et à des grossesses aberrantes (19,34). Certaines études (35,36) suggèrent aussi que la glycosylation de certaines glycoprotéines

puisse être corrélée à l'origine ethnique des donneurs, renforçant le besoin de disposer des outils nécessaires à la caractérisation de l'hormone et de sa glycosylation, à la fois dans un contexte de diagnostic et de design de biothérapeutique.

1.4.2 La gonadotrophine chorionique équine (eCG)

La sous-unité β de la gonadotrophine chorionique équine et de la hCG sont des séquences homologues et diffèrent en particulier au niveau de la séquence de leur extrémité C-terminale. Il en découle que la forme non-glycosylée de la β -eCG est légèrement plus massive que la version humaine, avec une masse moléculaire de 15,7 kDa environ. Mais la différence majeure entre ces deux hormones réside dans la densité des sites de glycosylation, puisque la β -eCG possède un unique site de N-glycosylation (Asn-13) et jusqu'à 12 sites de O-glycosylation potentiels. Ces derniers sont situés sur la partie C-terminale de la sous-unité β et apportent une forte densité de charge et une taille importante à la protéine, cohérente avec sa longue demi-vie dans la circulation (environ 2 à 5 jours) (37).

Au cours de l'évolution, l'hormone chorionique gonadotrope est apparue indépendamment chez les primates (hCG chez l'espèce humaine) et chez les équidés (eCG) (38). Contrairement à la hCG qui est produite dans le placenta par l'expression d'un gène dédié chez l'Homme, la eCG partage avec la LH équine un gène commun dont la transcription produit une hormone ayant les deux activités. Utilisée chez d'autres espèces de mammifères, la eCG possède également une activité FSH. Pour ces raisons, l'hormone est souvent utilisée en reproduction vétérinaire pour le bétail (jument, vache, truie...) et sa production implique des méthodes de récupération *in vivo* récemment dénoncées pour leur éthique contestable. Une meilleure connaissance de l'hétérogénéité structurale des glycanes de la eCG active et de la fonction précise des diverses formes de glycosylation pourrait ainsi permettre de s'affranchir de ces méthodes controversées de production.

1.5 Synthèse in vivo

Les familles de glycoconjugués sont multiples et la synthèse de chacune de celles-ci passe par des voies de synthèses variées et complexes. En effet, la banque de séquences protéines SwissProt recense 153 entrées (enzymes annotées) impliquées dans la voie biologique de la glycosylation des protéines chez l'Homme (banque de données consultée le 8 novembre 2019 à l'adresse www.uniprot.org/uniprot/?query=organism:9606+pathway:460.378).

Les protéines étudiées dans le cadre de ce projet sont N-glycosylées et O-glycosylées, l'accent sera donc mis sur la description des voies métaboliques mises en œuvre dans ces deux familles chez l'Homme et les mammifères, tandis que les autres voies de glycosylation seront décrites succinctement.

1.5.1 N-glycanes

Chez les organismes eucaryotes, la synthèse des N-glycanes a lieu dans le réticulum endoplasmique (RE) et l'appareil de Golgi. Cette voie de synthèse est précédée par la production d'un précurseur oligosaccharidique de 14 résidus GlcNAc(2)Man(9)Glc(3) lié à la membrane du RE par l'intermédiaire d'une molécule de dolichol (lipide isoprénoïde ou polyisoprénol)(Figure 10). La synthèse du dolichol pyrophosphate (Dol-P-P) activé met en jeu des enzymes diverses à la surface cytoplasmique du RE : La polyprénol réductase permet la production de dolichol à partir de son précurseur, une dolichol kinase transforme le dolichol en sa forme activée phosphatée (Dol-P), et la dolichyldiphosphatase prend en charge le recyclage du dolichol pyrophosphate déglycosylé en Dol-P. Des glycosyltransférases de la famille ALG (ALG7, ALG13/14, ALG1, ALG2, ALG11) effectuent les premières étapes de la synthèse du polysaccharide précurseur à partir du Dol-P ancré à la face cytoplasmique de la membrane du RE et de monosaccharides cytoplasmiques activés (UDP-GlcNAc et GDP-Man), puis transporté vers la face interne du RE par l'action de l'enzyme RFT1 (translocation membranaire). Une fois exposé à la lumière du RE, la synthèse du polysaccharide continue avec l'action d'autres enzymes transmembranaires (ALG3, ALG9, ALG12, ALG6, ALG8, ALG10) qui

catalysent la liaison du glycane-P-P-Dol et des monosaccharides ancrés (Glc-P-Dol et Man-P-Dol) afin de produire la forme finale du précurseur des N-glycanes $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-Dol}$. Ce glycane précurseur est transféré du lipide vers le résidu asparagine situé au sens de la séquence consensus **Asn-X-Ser/Thr** par des complexes enzymatiques appelés OST (oligosaccharyltransférase), OSTA dans le cas d'un transfert co-traductionnel ou OSTB si le transfert se fait sur la protéine intégralement traduite (ces complexes sont alors formés de sous-unités différentes).

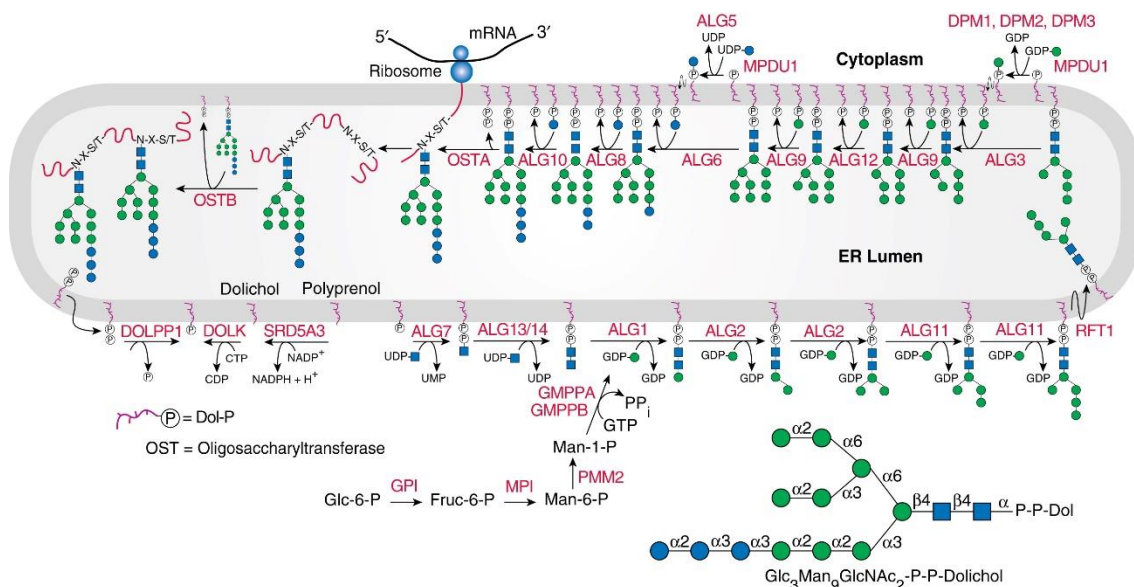


Figure 10 : Synthèse du précurseur $\text{Dol-P-P-GlcNAc}_2\text{Man}_9\text{Glc}_3$ à la surface externe puis dans le réticulum endoplasmique.

Les codes des enzymes catalysant chaque réaction sont indiqués en rouge. Figure issue du livre « Essentials of Glycobiology, Third Edition, Chapter 9, Figure 3 ».

A la suite du transfert sur la protéine du précurseur oligosaccharidique, les glucoses terminaux de ce dernier sont progressivement éliminés par des glucosidases du RE (Figure 11). Ce processus a lieu de concert avec le repliement de la protéine immature, sous contrôle des protéines chaperons du RE (Calnexine, Calreticuline) qui se chargent d'adresser la glycoprotéine correctement repliée au Golgi ou de la confier à l'enzyme EDEM (Endoplasmic Reticulum Degradation-enhancing α -Mannosidase I-like) pour sa dégradation si sa conformation est incorrecte. Une fraction des glycoprotéines normalement repliées peut aussi échapper à la maturation dans le Golgi et être sécrétée sous forme oligomannosylée.

La maturation des N-glycanes vers des formes complexes ou hybrides a lieu dans l'appareil de Golgi (Figure 11). Les N-glycoprotéines sont prises en charges par différentes glycosidases (principalement dans les saccules d'entrée, ou *cis*-Golgi) et glycosyltransférases (principalement dans les saccules de sortie, ou *trans*-Golgi). Plus précisément, le sucre $\text{Man}_8\text{GlcNAc}_2$ est dégradé par les mannosidases MAN1A1 et MAN1A2 (α 1-2 mannosidases I) jusqu'à la production de $\text{Man}_5\text{GlcNAc}_2$ dans le *cis*-Golgi. La formation de la première antenne des N-glycanes complexes ou hybrides est alors initiée par le transfert d'une GlcNAc (à partir d'une UDP-GlcNAc importée du cytoplasme) sur le mannose en α 1-3 du core par MGAT1 (α 1,3-Mannosyl-Glycoprotein 2- β -N-AcetylglucosaminylTransferase).

La N-glycoprotéine subit dans le *med*-Golgi (saccules médians de l'appareil de Golgi) l'action de MAN2A1 et MAN2A2 (α 1-2 mannosidases II) qui retirent jusqu'à deux mannoses du N-glycane pour former $\text{GlcNAcMan}_{3-5}\text{GlcNAc}_2$. Les formes $\text{GlcNAcMan}_{4-5}\text{GlcNAc}_2$ qui échappent à l'activité des MAN2A servent de précurseur aux formes hybrides des N-glycanes. La forme $\text{GlcNAcMan}_3\text{GlcNAc}_2$ peut alors être transformée par MGAT2 (α 1,6-Mannosyl-Glycoprotein-2- β -N-Acetylglucosaminyl-Transférase) qui lui ajoute une GlcNAc supplémentaire et initie la formation d'une deuxième antenne pour un N-glycane complexe biantennaire.

Les dernières étapes de la maturation des N-glycanes ont lieu dans le *trans*-Golgi, dans lequel les antennes du sucre sont galactosylées par la β 1-4-galactosyltransférase (isoformes 4, 3 et 2) et sialylées par des sialyltransférases (il en existe un grand nombre en fonction de l'ose recevant l'acide sialique et de la liaison osidique). Les N-glycanes peuvent aussi être fucosylés sur leur core par FUT8 (α 1,6-fucosyltransférase), bissectés par une GlcNAc fixée sur leur premier mannose par MGAT3 (β 1,4-mannosyl-glycoprotein 4- β -N-acetylglucosaminyltransférase), et leur nombre d'antennes peut être augmenté (jusqu'à 6 chez certains organismes) par d'autres enzymes de la famille des acetylglucosaminyltransférases (MGAT4A, 4B, 4C, 5, 5B, 6). Dans le *trans*-

Golgi, tous les oses sont utilisés sous leur forme activée nucléotide-monosaccharide (UDP-GlcNAc, GDP-Fuc, UDP-Gal ou CMP-NeuAc) importés depuis le cytoplasme par des transporteurs spécifiques.

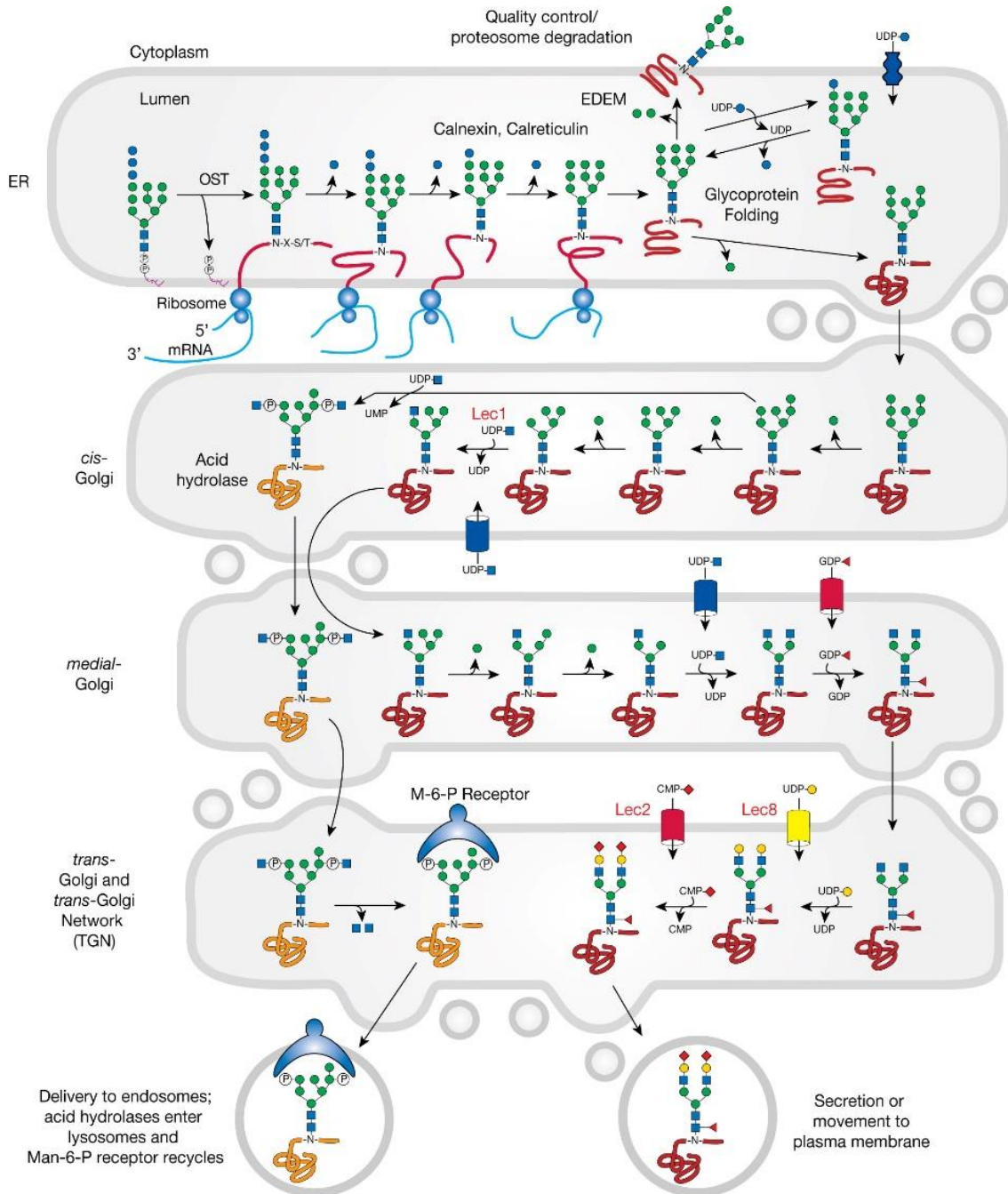


Figure 11 : Maturation d'un N-glycane dans le RE et l'appareil de Golgi.
Figure issue du livre « Essentials of Glycobiology, Third Edition, Chapter 9, Figure 4 ».

Les enzymes impliquées dans ce processus, et donc la composition des N-glycane présents sur la protéine mature, sont variables selon l'espèce, le type cellulaire, la protéine modifiée et le site de glycosylation. Une fois la protéine

correctement glycosylée, elle est exportée à la membrane cellulaire ou sécrétée par l'émission de vésicules golgiennes.

1.5.2 O-glycanes (O-GalNAc)

La synthèse des O-glycanes (O-GalNAc) a lieu principalement dans l'appareil de Golgi mais diffère de celle des N-glycanes sur un certain nombre de points. En effet, l'ajout de O-glycanes sur les protéines ne passe pas par des intermédiaires glycolipidiques mais uniquement par des monosaccharides activés (UDP-GlcNAc, UDP-Gal, CMP-NeuAc, GDP-Fuc) importés du cytoplasme. Les enzymes impliquées dans la O-glycosylation sont des enzymes transmembranaires exposées à la face interne des saccules de l'appareil de Golgi, concentrées dans le *cis*-Golgi pour celles ayant un rôle précurseur et dans le *trans*-Golgi pour les enzymes finalisant la production des O-glycoprotéines.

L'initialisation de cette synthèse nécessite la formation d'une liaison osidique entre le carbone anomérique du monosaccharide activé UDP-GalNAc et les résidus Thr ou Ser de la protéine en cours de maturation. Le site ainsi formé (GalNAc-O-Ser/Thr) constitue l'antigène Tn (39). Les enzymes responsables de ce transfert sont de la famille des GALNT (polypeptide GalNAc-transférase), localisée principalement dans le *cis*-Golgi. Les enzymes ayant cette activité sont nombreuses (20 gènes codant pour une GALNT chez l'Homme), et ont généralement des affinités et des spécificités variables pour différentes séquences peptidiques, expliquant la grande variabilité de séquence des O-glycoprotéines. Si aucune séquence consensus n'est identifiée pour la détermination des sites de O-glycosylation, la forte concentration de sérines, thréonines, prolines, glycines et alanines au voisinage immédiat des résidus O-glycosylés (Ser et Thr) semble indiquer que le faible encombrement stérique provoqué par ces acides aminés et la torsion imposée à la chaîne polypeptidique par les résidus Pro est favorable à l'action des GALNT (40).

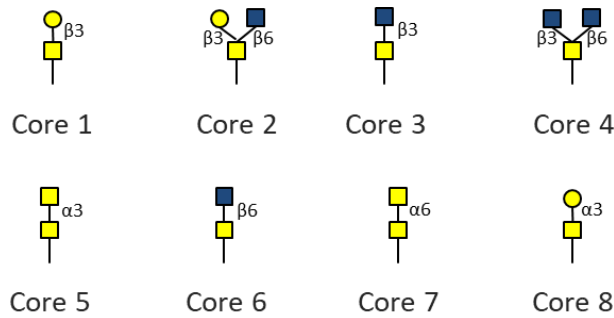


Figure 12 : Noyaux des O-glycanes de type mucine.

L'antigène Tn est progressivement modifié dans les saccules intermédiaires par l'ajout d'un à deux monosaccharides neutres (Gal, GalNAc, GlcNAc) pour former un des 8 *cores* communs aux O-GalNAc-glycanes (Figure 12). Le *core 1* (aussi appelé antigène T ou TF) est constitué du disaccharide Gal β 1-3GalNAc-O-Ser/Thr. L'enzyme C1GALT1 (*Core 1* β 1-3 galactosyltransférase 1) catalyse la formation du *core 1* à partir de l'antigène Tn dans l'appareil de Golgi, et l'activité de cette enzyme requiert la présence de la protéine chaperon C1GALT1C1 (C1GALT1 chaperon 1) qui contrôle la conformation de la transférase au moment de sa synthèse dans le RE. Ce *core* est ubiquitaire et semble constituer la majorité des O-glycoprotéines caractérisées (avec le *core 2*) (41). La synthèse du *core 3* est principalement limitée aux muqueuses, où est exprimée la B3GNT6 (acétylgalactosaminyl-O-glycosyl-glycoprotein β -1,3-N-acétylglucosaminyl-transférase) catalysant la réaction de transfert d'une GlcNAc en β 1-3 de l'antigène Tn. Le transfert d'une GlcNAc en β 1-6 de la base du *core 1* par les enzymes GCNT (β 1,3-galactosyl-O-glycosyl-glycoprotein β 1,6-N-acétylglucosaminyltransférases) forme le *core 2*. Les enzymes GCNT1 et GCNT4 sont responsable de la formation du *core 2* uniquement, tandis que la GCNT3 (exprimée dans les tissus à mucines, comme l'intestin ou le poumon) peut également cibler le *core 3* pour son activité transférase, permettant la synthèse du *core 4* (Figure 13). Les *cores 5* à 8 sont moins décrits et souvent très localisés : les O-glycanes formés à partir du *core 5* sont principalement retrouvés à la surface des adénocarcinomes ou dans le méconium (premiers déchets biologiques du nouveau-né), le *core 6* dans le méconium et la κ -caséine, le *core 7* dans les muqueuses sous-maxillaires bovines, et le *core 8* dans la muqueuse

respiratoire (42). En revanche, les enzymes et voies de synthèse de ces *cores* ne sont pas parfaitement connues, bien que les enzymes responsables de leur formation soient probablement exprimées dans les types cellulaires concernés par ces O-glycanes (43,44).

Les *cores* sont ensuite élongés ou ramifiés par l'action de GlcNAc-transférases, Gal-transférases, Sialyltransférases, Fuc-transférases pour former des O-glycanes des protéines matures. Certaines de ces enzymes sont communes aux voies de synthèses des O-glycanes, des N-glycanes et parfois des glycolipides. Ces enzymes incluent par exemple la B4GALT1 (β 1,4-galactosyltransférase 1), la B3GALT5, ayant une affinité pour le *core* 3 et certains glycolipides (β 1,3-galactosyltransférase 5), la B3GNT3 (N-acétyllactosaminide β 1,3-N-acétylglucosaminyltransférase 3), la ST3GAL1 qui catalyse la synthèse de l'antigène sialyl-T à partir du *core* 1 (CMP-N-acétylneuramate- β -galactosamide- α 2,3-sialyltransférase 1) et la ST6GALNAC1 (α N-acétylgalactosaminide α 2,6-sialyltransférase 1) qui catalyse la sialylation de l'antigène Tn.

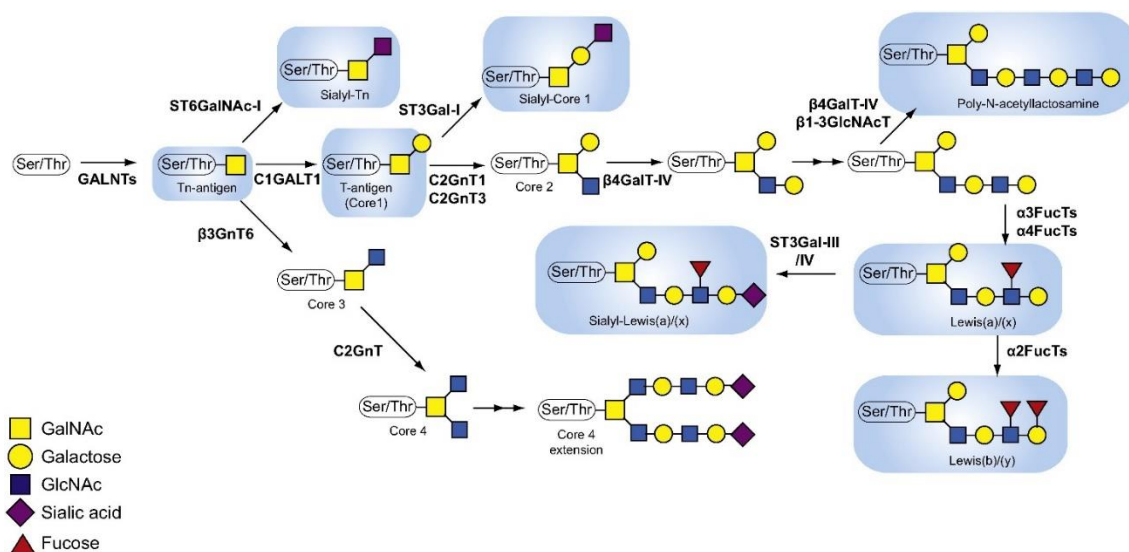


Figure 13 : Voies de synthèse des *cores* O-GalNAc 1 à 4 et formes de Lewis.

Figure issue de « Short O-GalNAc glycans: regulation and role in tumor development and clinical perspectives, Figure 1 », par J. Chia et al. (BBA, 2016).

1.5.3 Autres glycoconjugués

Les protéines ancrées (*GPI-anchored proteins*) sont synthétisées de façon similaire aux premières étapes de synthèse des N-glycanes, le

phosphatidylinositol ancré à la membrane externe du RE est d'abord modifié par un GlcNAc puis est mûri par l'action de glycosyltransférases et autres enzymes à la face interne du RE. Après le transfert d'une protéine sur le GPI, ses fractions lipidiques et osidiques sont modifiées et réarrangées dans le RE et l'appareil de Golgi jusqu'à l'obtention de la forme mature de la protéine-GPI, exportée par les vésicules golgiennes à la membrane externe de la bicouche lipidique.

Dans le cas de la C-mannosylation, le transfert d'un mannose depuis une molécule de Dol-P-mannose sur le C-2 d'un tryptophane a lieu dans le RE. Ce résidu semble être inclus dans une séquence consensus Trp-X-X-Trp (rarement Trp-X-X-Phe, avec X un acide aminé quelconque) et la réaction est catalysée par des C-mannosyltransférases. Ces enzymes sont assez méconnues chez l'Homme, bien que de récentes publications scientifiques semblent indiquer que la protéine DPY19L3 possède cette activité envers le résidu Trp-156 de la R-spondine 1 humaine (45).

Enfin, au-delà des voies « classiques » de glycosylation des protéines, de nombreuses autres liaisons protéine-glycane existent (au nombre de 34 couples « acide aminé-ose ») et sont plus ou moins bien caractérisées. Celles-ci aboutissent à la synthèse de glycoconjugués excrétés, exposés à la membrane externe, mais aussi parfois composants de la matrice extracellulaire (collagènes, protéoglycanes), cytoplasmiques (glycogénine), ou potentiellement dans tous les compartiments cellulaires.

2. Chimie analytique et glycoprotéomique

Les glycanes présents sur les glycoprotéines, en particulier sur les hormones chorioniques gonadotropes hCG et eCG, ont un rôle proéminent dans leur durée de vie et leur élimination, ainsi que dans leurs fonctions effectrices résultant de leur interaction avec les récepteurs membranaires LH/CGR. Ces paramètres chimiques et biologiques sont impactés par la nature exacte des structures polysaccharidiques, c'est-à-dire leur composition, leur stéréoisométrie, leur site de liaison à la protéine, et la microhétérogénéité de ces glycohormones. La diversité des glycoformes présentes au sein d'un individu pourrait aussi avoir un rôle dans l'activité globale des sécrétions hormonales.

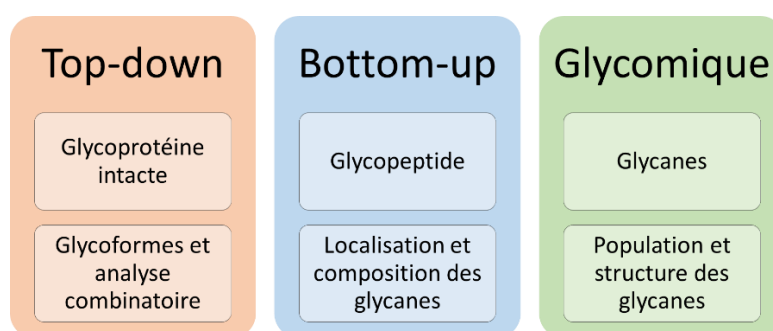


Figure 14 : Trois approches complémentaires pour l'étude des glycoprotéines, permettant l'obtention d'informations exclusives.

On distingue généralement trois niveaux d'analyse des glycoprotéines (Figure 14), selon qu'elle soit faite sur la molécules intacte (approche *top-down*), sur les glycopeptides issus de la dégradation chimique ou enzymatique de la protéine (approche *bottom-up*), ou uniquement sur les sucres après rupture de leur liaison à la protéine (glycomique). En effet, la dégradation des glycoprotéines permet de réduire la taille des molécules analysées et diminue ainsi leur complexité structurale et augmente les performances des outils analytiques disponibles. Cependant, ces étapes de traitement visant à diminuer la complexité structurale des échantillons provoque également la perte d'information, potentiellement biologiquement pertinente, quant à l'effet combiné de ces différentes glycosylations entre elles, ou avec d'autres types de modifications. Il semble donc judicieux de conserver ces approches

complémentaires qui sont toutes sources d'avancées dans le domaine de l'étude des glycoprotéines.

Bien que représentant un pan important de la glycoprotéomique, les études biologiques ou pharmacologiques des glycanes et de leurs dérivés ne seront pas approfondies dans le présent manuscrit. En effet, ces techniques très informatives permettant en particulier d'explorer la fonctionnalité des glycoconjugués sont généralement dissociées d'analyses focalisées sur la caractérisation structurale des glycanes. Plusieurs approches analytiques sont capables d'identifier les glycanes et glycoconjugués sur la base de la caractérisation des éléments structuraux de ces molécules.

Les lectines sont une famille de protéines présente chez tous les organismes reconnaissant certains traits structuraux des glycanes. Ces protéines et les immunoglobulines sont souvent utilisées comme outil pour la caractérisation des glycanes grâce à leur très grande spécificité et à leur grande diversité. Conjuguées à des éléments détectables (fluorochromes...), les différentes espèces de lectines permettent d'identifier les familles de N-glycanes et *cores* O-glycosidiques, la présence d'acides sialiques terminaux, de GlcNAc bisectrice ou de fucose, et donc virtuellement la plupart des glycanes. L'utilisation de protéines reconnaissant les motifs glycosylés est d'intérêt majeur mais est parfois limitée par la faiblesse de l'interaction glycane-lectine (46) ou par l'indisponibilité de protéine dirigée vers des formes de glycanes moins communes (47).

La résonance magnétique nucléaire (RMN) permet une analyse stéréochimique, non-destructive et sans connaissance préalable de la structure étudiée. Elle permet la caractérisation structurale au niveau de glycoprotéines intactes (48), ainsi que pour l'étude d'interactions glycane-récepteur (49,50). De plus, cette technique est relativement tolérante vis-à-vis d'une large diversité de molécules en termes d'acidité ou d'hydrophobie et est intrinsèquement quantitative. La RMN nécessite cependant des échantillons faiblement hétérogènes, rendant souhaitable l'utilisation de méthodes de séparation ou de

purification. Elle est limitée en sensibilité à des quantités de l'ordre de la nanomole pour l'analyse de glycanes.

La spectrométrie de masse (MS) permet de contourner certaines de ces limitations. Elle propose une meilleure sensibilité que la RMN et permet la détection d'un nombre important de structures réparties sur une grande gamme de concentrations, tout en pouvant être couplée à des méthodes de séparations en phase liquide ou gazeuse. L'utilisation de la spectrométrie de masse en tandem permet la caractérisation de la composition et de la séquence des glycanes, de leur site de liaison à la protéine ainsi que l'identification de celle-ci. Mais la spectrométrie de masse fait également face à certains verrous, en particulier quant à la nature des composés analysables, nécessairement ionisables et souvent hydrophobes. De plus, la quantification en spectrométrie de masse nécessite des méthodes de normalisation relativement complexes à mettre en œuvre.

L'étude des glycanes et des glycoconjugués par spectrométrie de masse est un domaine de recherche en constante évolution, tant au niveau instrumental que d'un point de vue méthodologique. L'objectif de cette partie est de présenter l'état des avancées techniques dans ce domaine afin de resituer la pertinence du travail de recherche entrepris.

2.1 Introduction à la spectrométrie de masse

La spectrométrie de masse (MS) est une technique de chimie analytique développée au long du XX^{ème} siècle, basée sur la mesure du rapport masse / charge (m/z) de molécules ionisées à l'état gazeux par un spectromètre de masse. On peut généralement schématiser ces appareils par la combinaison de quatre éléments : une source d'ionisation, un (ou plusieurs) analyseur(s), un système de détection et d'enregistrement du signal (Figure 15). Un spectromètre de masse peut être utilisé indépendamment, généralement pour des molécules purifiées ou des mélanges peu complexes, ou être couplé à des méthodes séparatives, en couplage direct (*online*) ou indirect avec collecte de fractions (*offline*).

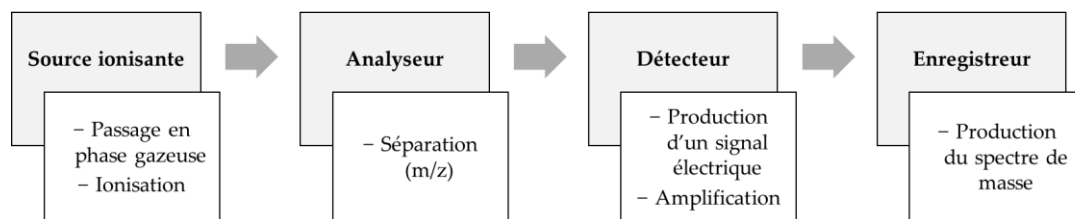


Figure 15 : Représentation schématique des éléments constituant un spectromètre de masse et leurs fonctions principales.

Le rôle de la source d'ionisation est notamment d'apporter une charge aux analytes neutres afin de les manipuler avec des champs électriques ou magnétiques. Les molécules ionisables peuvent être introduites dans différents types de sources en fonction de l'état initial de l'échantillon (solide, liquide, gazeux) ou du mode de transfert de la charge (ionisation électronique, chimique, assistée par matrice, par plasma, électrospray...). Tous les modes d'ionisation sont différents de par la quantité d'énergie initialement transmise, conduisant (ionisation dure) ou non (ionisation douce) à des fragmentations des analytes en source. Cependant, ils sont également conditionnés par leur capacité à transférer une charge aux analytes d'intérêt, il est notamment moins aisé d'ioniser positivement les glycanes qui ne possèdent pas nativement de résidu basique comme c'est le cas pour les peptides ou les protéines. Du fait de l'effet compétitif ou antagoniste de certaines molécules sur l'ionisation des analytes d'intérêt, et des différences d'efficacité d'ionisation pour des analytes différents, la spectrométrie de masse n'est pas intrinsèquement quantitative. Les ions ainsi formés en phase gazeuse sont soumis à un champ électrique ou magnétique et guidés vers l'analyseur.

Tout comme pour les sources d'ionisation, il existe une grande diversité d'analyseurs, à géométries et fonctionnements variables, destinés à séparer les ions selon leur rapport m/z. Il s'agit d'éléments dans lesquels les ions sont soumis à différentes combinaisons de champs électriques et/ou magnétiques. Ainsi, les ions sont soumis à des forces régies par les lois de Lorentz (I) et de Newton (II), transcrites en une équation différentielle (III) décrivant le mouvement des ions en tout point et à tout instant en fonction de leur rapport

m/z , et de paramètres instrumentaux et expérimentaux (intensité des champs électrique \vec{E} et magnétique \vec{B} , vecteurs vitesse \vec{v} et accélération \vec{a} , et \vec{F} la force appliquée à un ion) :

$$(I) \quad \vec{F} = z\vec{E} + z\vec{v} \times \vec{B}$$

$$(II) \quad \vec{F} = m\vec{a}$$

$$(III) \quad (m/z)\vec{a} = \vec{E} + \vec{v} \times \vec{B}$$

La mesure de la vitesse de ces ions implique leur détection et leur amplification. Le détecteur est parfois indépendant de l'analyseur comme c'est le cas des systèmes de dynodes, ou constituent un seul élément capable de mesurer le courant image provoqué par le passage d'ions. Le courant électrique produit par ces détecteurs est enregistré sous la forme de spectres ayant pour dimensions la quantité de signal (ou d'ions) en fonction du rapport m/z mesuré par l'analyseur.

Les performances des différentes combinaisons source-analyseur(s)-détecteur sont variables en termes de résolution (capacité à séparer deux molécules ayant des valeurs de m/z proches), de précision et justesse (distribution des mesures de m/z centrée étroitement sur la valeur vraie respectivement), de vitesse d'acquisition (de l'ordre de 10^{-1} Hz à au-delà de 10^2 Hz), de gamme de m/z (limites en termes de m/z mesurables) et de gamme dynamique (écart d'intensités relatives détectables simultanément).

À cette première dimension d'analyse est presque systématiquement associée une autre dimension m/z sur les appareils récents. En effet pour discerner des molécules isobares (mêmes masses nominales mais masses exactes différentes), la fragmentation de la molécule permet de séquencer et donc d'identifier des fragments spécifiques. On parle dans ce cas de spectrométrie de masse en tandem (MS/MS, MS², MS₂) qui consiste en deux analyses de masse successives. La séparation des ions est spatiale (deux analyseurs distincts et parfois différents) ou temporelle (un analyseur sépare les ions parents puis les

produits de fragmentation). On utilise dans ce cas différents modes d'activation des ions (collision avec un gaz neutre, excitation par des photons, impact ou transfert d'électrons...), leur apportant une énergie interne suffisante pour provoquer leur dissociation. Chaque mode de fragmentation met en place des mécanismes physiques différents aboutissant à des fragmentations différentes et parfois complémentaires. Ainsi il y a parfois un intérêt à combiner ou varier les méthodes de dissociation afin d'obtenir des informations de séquences plus complètes ou élucider des motifs structuraux précis.

Enfin, ces techniques sont souvent complétées par des dimensions de séparation supplémentaires, comme des méthodes chromatographiques en phase liquide ou gazeuse pour les composés volatils. Cette dimension est nécessaire en particulier dans le cas d'échantillons complexes pour être en accord avec la vitesse d'acquisition en MS, mais également pour résoudre efficacement des glycanes et glycoconjugués isobares ou isomères. Dans ce cas, le choix des phases et du mode de séparation est critique et doit être adapté à la question analytique.

Deux approches seront particulièrement étudiées, la N-glycomique et la glycoprotéomique.

2.2 N-glycomique

La glycomique est l'étude de l'ensemble des molécules saccharidiques d'un organisme, d'un tissu, d'un biome, *etc.* On parle de N-glycomique pour lorsqu'on se limite à l'étude des N-glycanes issus de la déglycosylation d'une glycoprotéine.

Cette approche demande donc nécessairement de libérer les glycanes des protéines à l'aide de traitements chimiques ou enzymatiques. Les produits de déglycosylation doivent être purifiés avant l'analyse MS du fait de leur plus faible efficacité d'ionisation que les protéines et peptides.

Les glycanes sont des molécules hydrophiles et non chargées ou acides, différentes stratégies sont employées pour permettre leur ionisation et leur

analyse. Le mode d'ionisation MALDI (désorption et ionisation laser assistée par matrice) permet la co-cristallisation avec une molécule aromatique (la matrice) capable d'absorber les photons UV et servant d'intermédiaire au transfert d'énergie du laser vers l'analyte. De fait de la rapidité d'acquisition des analyseurs de type TOF et les possibilités de parallélisation de la préparation d'échantillon en MALDI, cette approche a un fort potentiel d'analyses à haut débit, particulièrement compatible avec des analyses de cohortes. Alternativement, les sucres peuvent être modifiés pour être compatibles avec les sources de type électrospray (ESI), souvent couplées à des méthodes de chromatographie liquide. Ce couplage permet une meilleure séparation des analytes aidant à leur identification mais augmentant significativement le temps nécessaire à l'analyse d'un échantillon.

Quelle que soit la méthode de détection utilisée, une des limitations majeures de l'approche glycomique tient du fait de son concept, les glycanes étant analysés indépendamment de leur protéine d'origine. Ainsi, si l'échantillon consiste en un mélange de glycoprotéines ou en une unique protéine multi-glycosylée, les identifications porteront sur l'ensemble des N-glycanes libérés et ne pourront être attribuées à un unique site ou une unique protéine. L'étude des N-glycanes issues d'une ou d'un ensemble de glycoprotéines par MALDI-TOF (*N-glycan MALDI fingerprinting*) permet donc d'envisager la glycosylation comme un phénomène global.

2.2.1 Production des N-glycanes

2.2.1.1 Déglycosylation des protéines

La première question qui se pose lors de la mise au point de l'analyse par une approche glycomique est la méthode de récupération des glycanes à partir de glycoprotéines.

La dégradation totale des glycanes par hydrolyse acide permet d'en analyser la composition, tandis que des exoglycosidases (sialidases, mannosidases, *etc.*) permettent d'hydrolyser une liaison osidique terminale et ainsi « rogner » les extrémités non-réductrices des polysaccharides afin d'en réduire la complexité.

Cependant, dans le but de disposer du N-glycane intact et d'en analyser la structure, il est nécessaire de rompre la liaison N-osidique entre la première GlcNAc et l'asparagine modifiée. L'élimination des N-glycanes est généralement effectuée par des glycosidases hydrolysant la liaison N-osidique entre le polysaccharide et la protéine.

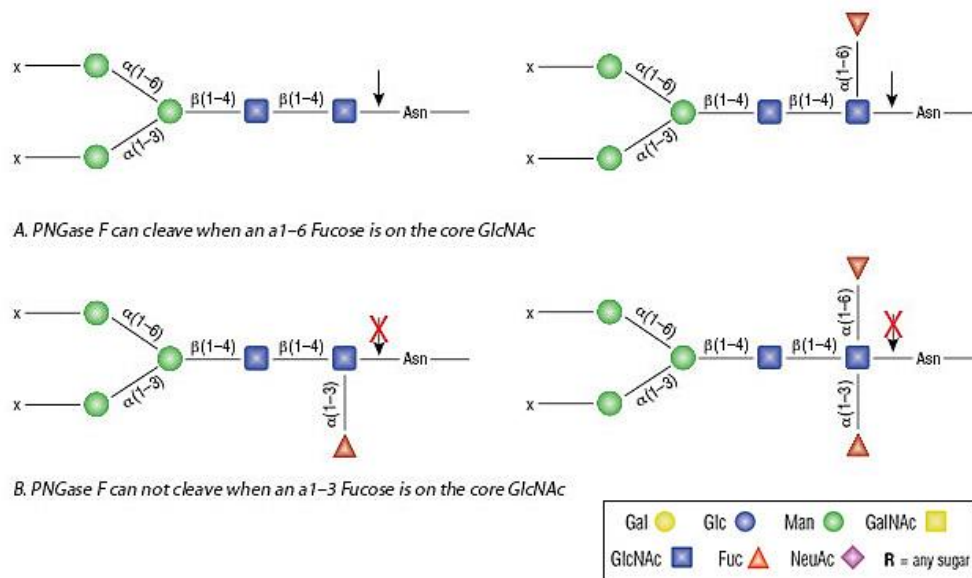


Figure 16 : Déglycosylation par la PNGase F, spécificité des sites de clivage. La PNGase F, contrairement aux PNGases A, F-II et H^r, est inactive sur les N-glycanes α 1-3fucosylés. Source : New England BioLabs Inc.

Parmi ces enzymes, les molécules de la famille des peptide-N4-(N-acétyl- β -glucosaminyl)asparagine amidases (ou PNGase) sont couramment utilisées. Elles catalysent la réaction d'hydrolyse de la liaison asparagine-glycane, modifiant l'asparagine déglycosylée en résidu aspartate. Plusieurs types de PNGase ont été utilisés pour l'étude des N-glycanes en fonction de certaines spécificités structurales de ces derniers. La PNGase A tient son nom de son isolation à partir d'extrait d'amandes (51), et possède une activité amidase indépendante de la structure du N-glycane. Cependant, la PNGase A possède une faible efficacité d'hydrolyse sur les glycoprotéines et est préférentiellement utilisée sur les glycopeptides (naturels ou issus de la digestion de glycoprotéines). Cette glycosidase est elle-même une glycoprotéine, et son autodigestion tendrait à compromettre les analyses glycomiques.

Une autre PNGase, d'origine procaryote, a été découverte puis caractérisée (52,53), elle propose l'avantage d'être plus efficace que la PNGase A (54), non glycosylée et est capable d'être produite sous forme insoluble par transformation bactérienne dans des systèmes hétérologues *E. coli*. La PNGase F (pour *Flavobacterium meningosepticum*, l'espèce produisant l'enzyme, maintenant dénommée *Elizabethkingia meningoseptica*) fonctionne aussi bien sur les glycopeptides que les glycoprotéines, permettant de s'affranchir d'étapes de protéolyse coûteuses en matériel biologique. Cette PNGase F est aujourd'hui la forme la plus communément utilisée en glycomique. Néanmoins, l'activité de la PNGase F est inhibée par la présence d'un fucose en α 1-3 du *core* du N-glycane (forme très représentée chez les plantes et les insectes) (55,56), limitant son utilisation à des structures qui en sont dépourvues, notamment les glycoprotéines de mammifères (Figure 16).

Plus récemment, d'autres enzymes ayant une activité PNGase ont été caractérisées dans des organismes procaryotes : une seconde PNGase F (F-II) a été isolée d'une souche du même organisme *E. meningoseptica* (57), tandis qu'une PNGase active en conditions acides a été découverte chez la bactérie *Terriglobus roseus* (PNGase H⁺) (58). Ces deux enzymes ont l'avantage, comme la PNGase A, d'être actives même en cas de fucosylation du *core* en α 1-3. La PNGase F-II permet la déglycosylation des N-glycoprotéines préférentiellement dénaturées et son activité PNGase même en cas de α 1-3Fuc sur la base du sucre indique un mécanisme d'action distinct de la PNGase F.

La PNGase H⁺ a une activité maximale en conditions acides (pH optimal de 2,6) offrant de nouvelles possibilités analytiques, mais augmentant la dégradation de certains résidus dans ces conditions, notamment les acides sialiques terminaux usuellement hydrolysés à faible pH (59). D'autres enzymes à activité PNGase sont évidemment exprimées dans différents organismes, certaines isolées à partir de plantes, champignons, mammifères (60–62), mais ne sont pas toujours adaptées à une utilisation *in vitro*. La N-déglycosylation par la PNGase F en gel après une séparation électrophorétique SDS-PAGE (63) ou sur

membrane à la suite d'un transfert (Western Blot) (64) permet de limiter la déglycosylation à une glycoprotéine isolée d'un mélange. Différentes stratégies ont également permis d'améliorer l'efficacité et la vitesse de la déglycosylation, en immobilisant la PNGase (forme F uniquement) (65), en assistant la digestion par micro-ondes (66) ou ultrasons (67). Plusieurs endoglycosidases permettent également d'hydrolyser la liaison entre les deux GlcNAc du *core* des N-glycanes. Adaptées à certaines stratégies analytiques particulières, l'utilisation de telles glycosidases ne permet cependant pas d'analyser les glycanes intacts.

Les méthodes d'analyse des N-glycanes par spectrométrie de masse sont généralement transposables à la O-glycosylation. Cependant, la déglycosylation de tous les O-polysaccharides n'est pas possible par voie enzymatique, aucune O-glycanase efficace sur tous les *cores* n'étant connue. L'hydrolyse des liaisons osidiques entre un glycane et un résidu Ser/Thr par voie chimique en conditions alcalines a été également décrite. Cette réaction de β -élimination consiste en l'attaque du C α de la sérine modifiée par les ions hydroxyde et la libération consécutive du O-glycane (68). Des conditions réductrices sont généralement utilisées pour empêcher le O-glycane de subir lui aussi une dégradation en conditions alcalines. Cependant la réduction du O-glycane empêche tout marquage ultérieur au niveau de son extrémité réactive. De ce postulat a découlé l'utilisation de conditions non réductrices pour l'élimination des O-glycanes et leur dérivation chimique (69).

2.2.1.2 Modification chimique des glycanes

Les N-glycanes sont nativement hydrophiles de par l'abondance de groupements hydroxyles et souvent neutres et peu ionisables, à l'exception notable des formes sialylées chargées négativement. De plus, les glycanes sont constitués de monosaccharides très proches structurellement les uns des autres, et sont de ce fait des molécules très symétriques dont les liaisons sont difficiles à caractériser précisément. Différentes stratégies ont été développées afin de modifier chimiquement les sucres pour améliorer leur rétention ou leur fragmentation.

Afin notamment d'augmenter l'hydrophobie des glycanes et d'augmenter leur volatilité nécessaire à leur transition en phase gazeuse dans un spectromètre de masse, les hydrogènes des groupements -OH, -NH₂ et -COOH peuvent être remplacés par des méthyles. Un des avantages principaux de la perméthylation par rapport aux autres modifications des groupes hydroxyle (acétylation, trifluoroacétylation...) est le faible adduit de masse généré (-H → -CH₃ soit +14 Da), qui permet de limiter la masse des glycanes modifiés. Cette particularité est nécessaire afin d'opérer dans des gammes où la plupart des spectromètres de masse sont les plus sensibles et résolutifs. Ce traitement implique généralement une réaction avec l'iodométhane (CH₃I) en conditions alcalines (NaOH) et anhydres (DMSO, diméthylsulfoxyde), et a été développé sous différents formats (70,71). Bien que largement utilisée en combinaison avec des méthodes de séparation liquide en phase inverse, la perméthylation manque de reproductibilité et nécessite plusieurs manipulations de l'échantillon et étapes de purification limitant l'aspect quantitatif de l'analyse glycomique sur de tels sucres (72). De plus, le développement des méthodes de chromatographie compatibles avec les glycanes natifs (HILIC, *etc.*) diminue l'intérêt de tels traitements à l'échelle conventionnelle. Des approches miniaturisées pourraient néanmoins permettre d'augmenter leur rendement.

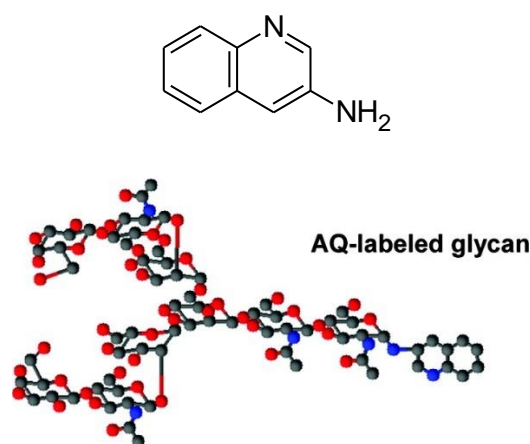


Figure 17 : Structure de la 3-aminoquinoline.

La fonction amine primaire libre participe à la liaison à l'extrémité réductrice du N-glycane (bas).

La réactivité de l'extrémité réductrice des N-glycanes a souvent été exploitée pour y greffer des groupements rapporteurs, en particulier fluorescents, afin

d'augmenter la sensibilité de leur détection. La GlcNAc réductrice terminale et les autres monosaccharides constituant des N-glycanes ont des réactivités différentes mais peuvent néanmoins être isobares. De ce fait, les fragments produits en MS/MS peuvent être de masses identiques et donc peu informatifs quant à la structure des glycanes, comme le montre la Figure 18.

Des stratégies de dérivation ont ainsi été développées pour des applications en spectrométrie de masse. Une des plus courantes est l'amination en conditions réductrices, consistant à faire réagir une molécule aminée avec le groupement carbonyle du glycanes, formant une base de Schiff peu stable qui peut ensuite être réduite en amine secondaire stable. Les amines les plus communément utilisées pour ces dérivations sont la 2-aminobenzamide (2-AB) et la 2-aminopyridine (2-AP). Bien que couramment utilisées pour l'analyse de polysaccharides en HPLC-ESI-MS, leur intérêt n'a été que partiellement démontré avec les sources MALDI. La limitation majeure de ces méthodes est généralement la nécessité de purifier les glycanes après leur marquage et la conversion non quantitative du sucre en son/ses dérivés (73). Utilisée en tant que matrice pour la première fois en 1994, la 3-aminoquinoline (3-AQ)(Figure 17) n'a pas immédiatement présenté un intérêt évident pour l'analyse de glycoconjugués du fait de la formation de base de Schiff (forme instable). Cette situation a été exploitée par Rohmer *et al.* (74) en utilisant des conditions favorisant la conversion quantitative du sucre en sa base de Schiff associée. De ce fait, l'utilisation de 3-AQ combine les bénéfices de la dérivation pour l'analyse structurale des sucres, mais agit également en tant que matrice pour l'ionisation MALDI, éliminant la nécessité d'étapes de lavages supplémentaires et permettant la réaction directement sur cible MALDI.

Les acides sialiques sont notablement instables en spectrométrie de masse du fait du caractère labile de leur proton acide. Plusieurs stratégies visant à modifier le groupement acide en un groupement neutre stable ont émergé. La formation d'esters (de méthyle principalement) est une méthode couramment employée pour stabiliser les fonctions acides, d'autant plus qu'elle permet sous

certaines conditions expérimentales d'obtenir des informations de liaison des acides sialiques. En effet, l'utilisation de chlorure de 4-(4,6-diméthoxy-1,3,5-triazin-2-yl)-4-méthylmorpholinium (DMT-MM) permet l'estérification des acides sialiques liés en α 2-6 tandis que les formes liées en α 2-3 sont converties en lactones, introduisant un delta de masse de 32 Da détecté en spectrométrie de masse (75).

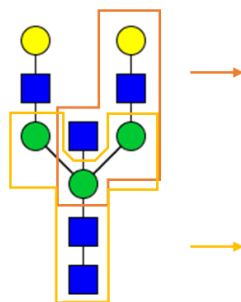


Figure 18 : Représentation de deux fragments isobares issus de la fragmentation en MS/MS du sucre A2G2.

La masse des deux fragments est identique (910,3278 Da).

2.2.1.3 Purification des sucres

L'utilisation de méthodes séparatives (chromatographie phase inverse, interaction hydrophile, affinité, *etc.*) remplit deux rôles principaux dans l'étude des glycanes. D'abord, lorsqu'ils sont issus de la déglycosylation de protéines, il est nécessaire de purifier les sucres par ces méthodes séparatives afin de minimiser les effets de matrice et la suppression du signal ionique dans la source d'ionisation en MS. De plus, certaines méthodes séparatives permettent de résoudre les espèces isomères n'étant pas discernables en MS puisque de même masse. Cependant, du fait de la grande diversité structurale des glycanes (charge, taille, masse, hydrophobie...) il n'existe pas de méthode universelle permettant une purification quantitative des glycanes (76). Les techniques d'électrophorèse capillaire sont également utilisées pour la séparation des glycanes et glycoconjugués chargés (nativement ou non) mais ne seront pas détaillées ici. La séparation en phase gazeuse par spectrométrie de mobilité ionique (IMS, *Ion Mobility Spectrometry*) ne sera pas discutée ici mais son intérêt

est indéniable et mérite d'être utilisée dans le but d'ajouter une dimension de séparation supplémentaire.

Dans l'optique de purifier les glycanes, les différences de solubilité des protéines et des oligosaccharides dans des solvants organiques peuvent être exploitées. Plusieurs stratégies ont été développées, et doivent être évaluées et adaptées en fonction des protéines étudiées et de leur solubilité. La précipitation des protéines sériques humaines après N-déglycosylation a été réalisée dans une solution d'éthanol à 80% à basse température (-75°C) (77). Les protéines insolubles sont séparées des N-glycanes et des sels en solution par simple centrifugation. De façon similaire, la précipitation des N-glycanes et des protéines en acétone (80%) est possible (78). Dans ce cas, le culot de précipitation est lavé afin de le débarrasser des sels ayant un effet négatif sur l'ionisation en MS. Les sucres sont ensuite récupérés par des lavages successifs avec une solution de méthanol à 60%. Un rendement maximal de purification est obtenu dans des conditions légèrement acides (pH 5,5). Dans chacune de ces méthodes, les glycanes sont récupérés dans un solvant partiellement organique et volatile (méthanol, éthanol), permettant une concentration et une reprise de l'échantillon dans un solvant adéquat après évaporation.

Bien que la solubilité des glycanes dans ces solvants soit très probablement variable, ces méthodes visent avant tout à collecter la majorité des glycanes et ne permettent pas une sélectivité particulière vis-à-vis des différents glycanes. Les lectines sont une famille de protéines (introduites au paragraphe 2. Chimie analytique et glycoprotéomique) reconnaissant des molécules de nature osidique avec une grande spécificité. Immobilisées (par couplage à des billes d'agarose, de silice, *etc.*), les lectines sont couramment utilisées en chimie analytique pour purifier les glycanes et glycoconjugués (79,80). Ces méthodes chromatographiques sont regroupées sous le terme LAC (*Lectin Affinity Chromatography*). Elles permettent de purifier les glycoprotéines avant leur déglycosylation ou de purifier les glycopeptides / glycanes issus de leur protéolyse / de leur déglycosylation. Des études récentes montrent l'utilisation

de LAC miniaturisée au format capillaire, permettant notamment la réduction de la durée totale du protocole, une meilleure sensibilité et une possible automatisation, bien que la répétabilité de telles méthodes soit perfectible (81,82). Plusieurs lectines peuvent être immobilisées sur le même substrat pour combiner leurs spécificités et étendre la gamme de molécules enrichies (MLAC, *Multi-Lectin Affinity Chromatography*) (83). Le séquençage de lectines d'origine bactérienne permet désormais d'exprimer ces protéines par génie génétique, et de modifier leur site actif et donc leur interaction avec les glycanes par mutagenèse dirigée (84). L'utilisation des lectines dans un cadre analytique est un domaine également porté par le développement de micropuces à lectine (ou *lectin microarray*) permettant une analyse comparative et quantitative à haut débit de la glycosylation d'un grand nombre d'échantillons (85). En tant que méthode de purification, les méthodes LAC sont principalement utilisées en analyse ciblée pour isoler une catégorie particulière de glycanes.

Les méthodes de chromatographie liquide sont les stratégies les plus utilisées, puisqu'elles permettent de séparer des glycanes natifs ou modifiés ainsi que de séparer certains isomères de position. Utilisées indépendamment des méthodes de détection (*offline*), elles servent principalement à dessaler ou purifier l'échantillon polysaccharidique, et sont également utilisées en couplage à la MS pour une meilleure résolution des glycoformes. Les glycanes ont une composante hydrophile forte du fait de la grande quantité de groupements hydroxyle sur chaque monosaccharide permettant une rétention en phase normale (NPLC, *Normal-Phase Liquid Chromatography*). De plus, leur masse carbonée non négligeable et la possible modification chimique des groupements hydroxyle (perméthylation) et / ou de l'extrémité réductrice des sucres permet de moduler leur hydrophilie et d'améliorer leur rétention sur les phases hydrophobes des méthodes de chromatographie à polarité de phase inverse (RPLC, *Reversed-Phase Liquid Chromatography*).

Ces dernières et notamment les phases de type octadécylrique C18 sont les phases les plus utilisées en protéomique et en couplage avec la spectrométrie de masse du fait de leur polyvalence, de leur stabilité et de la compatibilité des solvants utilisés avec la MS. Les glycanes natifs sont faiblement retenus sur ce type de phase de par leur hydrophilie et limitent l'utilisation des phases C18 aux glycanes perméthylés et généralement modifiés au niveau de leur extrémité réductrice (86,87). La RPLC (souvent sur phase C18) de par sa large utilisation est une méthode éprouvée et très robuste, cependant, les chromatographies liquides HILIC (*Hydrophilic Interaction Liquid Chromatography*) et PGC (*Porous Graphitized Carbon*) permettent une meilleure résolution des glycanes isomères et isobares. La RPLC peut toutefois être couplée à ces autres dimensions séparatives afin d'en augmenter la résolution.

La chromatographie HILIC est une méthode de choix couramment utilisée en glycomique puisqu'elle permet la rétention de glycanes chargés et/ou très hydrophiles, et donc pas nécessairement modifiés par dérivation chimique. Elle est originellement définie comme une chromatographie de partage dans laquelle les analytes interagissent avec une couche riche en solvant aqueux en équilibre autour de la phase stationnaire hydrophile, et sont élués par un mélange d'eau et de solvant organique (88). Les mécanismes de rétention et d'éluion mettent en réalité en jeu plusieurs types d'interactions (hydrogène, ionique, dipôle-dipôle) et la séparation des analytes se fait en fonction de leur polarité (89). De par les phases mobiles (eau, acétonitrile) et stationnaires (hydrophiles) utilisées, ainsi que les mécanismes de rétention mis en jeu, on positionne parfois la chromatographie HILIC à l'interface des chromatographies RPLC, NPLC et ionique. Plusieurs phases et conditionnements ont été décrits pour la purification des N-glycanes. On peut notamment citer l'utilisation de phase solide à base de polymère osidique (cellulose, agarose) (90,91), de silice amine (3-aminopropyle) (92,93) ou amide (94), ainsi que l'introduction relativement récente de phase zwitterionique (ZIC, *zwitterionic chromatography*) (95,96). Ces méthodes de purification « *offline* » ont été implémentées à des couplages LC-MS afin de

réduire les pertes de matériel et les biais de sélectivité et permettre la résolution des différents glycanes. La chromatographie HILIC présente un intérêt majeur puisqu'elle permet non seulement d'enrichir les glycanes mais également d'en résoudre certains isomères structuraux (acides sialiques α 2-3 ou α 2-6, fucose situé sur le *core* ou sur une antenne...) (95,97).

L'enrichissement et la séparation des glycanes natifs ou dérivés est également couramment effectuée sur des phases carbonées (PGC) seules ou couplées à d'autres phases (98–101). Les mécanismes de rétention sont multiples, le plus évident étant une rétention proche de la chromatographie en phase inverse (RPLC C18) liée à la masse carbonée des analytes (interactions de van der Waals) (102). Cependant et bien que le PGC soit idéalement un réseau de Carbone parfaitement plan et apolaire, plusieurs mécanismes proposés indiquent la formation d'une région polarisée à la surface de la phase PGC en réaction à la proximité d'un soluté polaire (103) ou à cause de la délocalisation des électrons sur le réseau carboné (104). Il est généralement admis que c'est ce mécanisme de rétention polaire qui domine dans la chromatographie sur phase graphitique.

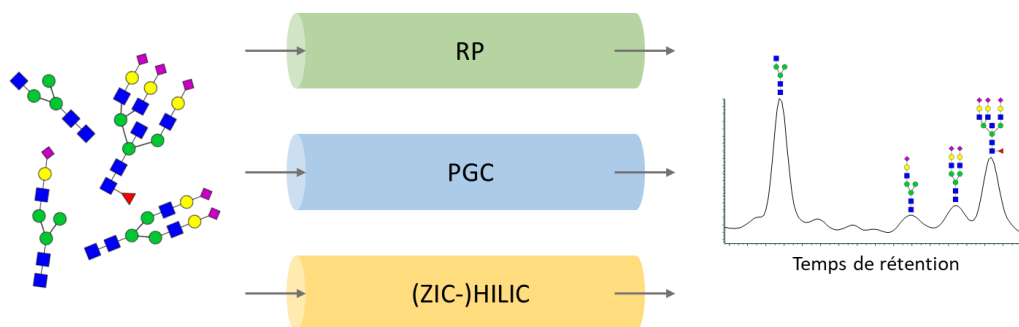


Figure 19 : Illustration de la séparation chromatographique des glycanes.

La séparation à polarité de phase inverse (RP) nécessite la dérivation des groupements hydroxyle libres des sucres. Bien qu'impliquant des mécanismes de rétention très différents (polarité, charge, hydrophobie, etc.), la rétention des N-glycanes les plus massifs et les plus sialylés est généralement plus importante.

Pour toutes ces méthodes, le choix de la phase mobile est également critique, en particulier en ce qui concerne le choix de l'acide et du contre-ion utilisé. Il ne semble pas y avoir de consensus quant à la phase stationnaire la plus adaptée à la séparation et l'analyse des N-glycanes (105–109) et la forte orthogonalité

des méthodes indique que dans leur état actuel, ces technologies présentes toutes un intérêt pour l'étude des glycanes (Figure 19).

2.2.2 Etude structurale des N-glycanes par MS

2.2.2.1 Matrices d'ionisation MALDI solides et liquides

Un des modes d'ionisation les plus utilisés en spectrométrie de masse dans le domaine de la glycobiochimie structurale est l'ionisation *Matrix-Assisted Laser Desorption/Ionization*, ou MALDI. Elle permet le transfert d'énergie aux analytes et leur ionisation par l'intermédiaire de l'excitation de composés aromatiques, appelés « matrice », par une source de lumière ultra-violette (Laser). Ce procédé permet notamment d'ioniser des composés nativement neutres et hydrophiles donc peu volatils, notamment les glycanes.

Plusieurs matrices ont été utilisées pour l'analyse de glycanes et ont donné des résultats probants. L'acide α -cyano-4-hydroxycinnamique (*α -cyano-4-hydroxycinnamic acid*, CHCA) est une matrice communément utilisée pour l'analyse de peptides ou de sucres. Elle a l'avantage de co-cristalliser avec les analytes de façon homogène, en formant une fine couche uniforme de cristaux rhomboédriques, et permet ainsi une analyse reproductible. Le signal produit est intense et « clair », avec peu de bruit, et les fragments produits en MS/MS CID (fragmentation par collision avec un gaz neutre) sont principalement des ruptures de la liaison glycosidique, donc relativement peu informatifs mais simples à interpréter. La matrice CHCA présente certains désavantages pour l'analyse des sucres, car c'est une matrice dite « chaude » qui induit des fragmentations spontanées après ionisation du fait de la grande quantité d'énergie interne acquise par les sucres (PSD, *Post-Source Decay*). Les glycanes sont souvent perméthylés avant l'analyse en MS pour minimiser ce phénomène, ce qui implique des étapes supplémentaires de manipulation et de lavage de l'échantillon et donc une perte de matériel et de sensibilité. L'utilisation d'acide 2,3-dihydroxybenzoïque (DHB) est également répandue en glycomique, du fait de son caractère plus « froid », induisant moins de fragmentation de type PSD. La fragmentation CID est aussi plus informative

car les sucres peuvent se fragmenter au niveau des liaisons osidiques mais aussi produire des ions issus de fragmentations « *cross-ring* » (fragmentations au sein du cycle d'un monosaccharide) ou internes. Un des principaux désavantages de la matrice DHB est qu'elle forme des cristaux en aiguille initiés à la périphérie des dépôts. Ce mode de cristallisation induit nécessairement une faible homogénéité du dépôt et la formation de zones riches en signal, limitant la reproductibilité et l'automatisation des mesures de masse. (110,111)

L'utilisation de matrices liquides (ILM, *Ionic Liquid Matrix*) s'est développée ces vingt dernières années et notamment pour l'analyse des sucres. Les ILMs sont un mélange de composés aromatiques solides (matrice) et d'un sel, qui reste fluide sous vide, pouvant prendre la forme d'un liquide ou d'un gel, parfois incrustés de cristaux en fonction des molécules qui composent le mélange (112). Du fait de leur état liquide, les matrices liquides sont plus homogènes que les matrices classiques, moins susceptibles de présenter des zones préférentielles ou « zones chaudes » et donc plus à même de permettre une approche quantitative et automatisable (113). Cette homogénéité des dépôts permet également une plus grande précision de masse, qui peut être affectée par les variations d'épaisseur inhérentes à la lente formation de cristaux de matrice avec les analyseurs de type Temps de Vol (TOF). Cependant, plusieurs études ont montré que les ILMs nécessitent des énergies Laser plus importantes que les matrices classiques, conduisant nécessairement à une perte de résolution. Ce paramètre doit également être suivi attentivement pour l'étude des sucres qui sont généralement affectés par des phénomènes de fragmentation spontanée en source.

L'utilisation de ces matrices liquides donne des résultats variables selon la nature des analytes (peptides, oligosaccharides...), mais des matrices constituées de sels d'acide para-coumarique (pCA, *p-Coumaric acid*) ont permis d'améliorer sensiblement la détection de N-glycanes (114), avec une sensibilité augmentée de deux à trois ordres de grandeur par rapport à la matrice DHB.

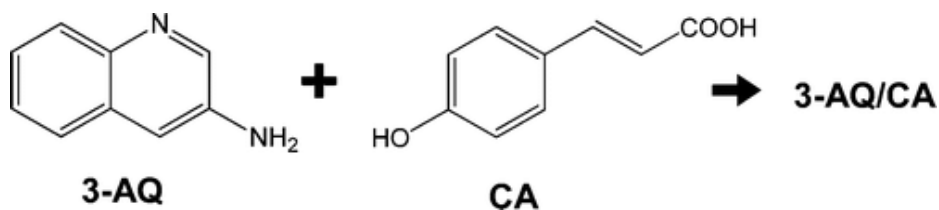


Figure 20 : Composition de la matrice liquide 3-AQ/CA, telle décrite par Y. Fukuyama et al.

Les fonctions les plus labiles des glycanes, comme les acides sialiques ou les groupements sulfate, sont stabilisées lorsque la matrice utilisée contient du pCA, diminuant également l'impact négatif de ces dissociations sur la sensibilité. Cette propriété semble trouver son origine dans plusieurs facteurs, d'abord l'affinité du pCA pour les groupements hydrophiles, sa faible ionisation qui tend à limiter la dissociation des groupements labiles et la fragmentation en source des glycanes, et le pH neutre ou légèrement acide des ILMs à base de pCA favorisant le maintien des groupement labiles sensibles aux pH acides. De plus il semble que la fragmentation spontanée des glycanes ionisés en MALDI soit un phénomène relativement lent et donc ayant une faible incidence sur des analyseurs rapides comme le TOF, mais pouvant être problématique lorsque la source MALDI est couplée à des analyseurs de type « piège à ions ». L'utilisation combinée de pCA et d'aminoquinoline (3-AQ) forme une matrice liquide, proposée initialement par Y. Fukuyama (115) pour l'analyse des glycanes, glycopeptides et phosphopeptides (Figure 20). Elle y est décrite comme permettant un gain en sensibilité et une diminution de la dissociation des groupements labiles.



Figure 21 : Comparaison visuelle de la forme des cristaux formés par différentes matrices MALDI. (gauche) CHCA, cristaux rhomboédrique, (milieu) DHB, cristaux en aiguilles, (droite) 3-AQ/pCA, matrice formant un gel à pression atmosphérique. (Photos CHCA et DHB issues de la publication de Helen M Karlsson et al. « Two-Dimensional Gel Electrophoresis and Mass Spectrometry in Studies of Nanoparticle-Protein Interactions »)

La matrice 3-AQ/pCA permet une dérivation de l'extrémité réductrice des sucres sans lavage, l'excès de 3-AQ agissant comme une matrice MALDI (voir paragraphe 2.2.1.2), et propose également un gain en sensibilité, en homogénéité des échantillons et en précision de masse du fait de l'état liquide de la matrice (Figure 21). Elle semble être peu reprise dans la littérature, essentiellement par le groupe qui en est à l'origine (116–118), pour l'étude par MALDI-TOF de N-glycanes. Ses propriétés en font pourtant une matrice susceptible d'améliorer la robustesse et la sensibilité des stratégies d'analyse glycomique.

2.2.2.2 Approches ESI et fragmentation en MS/MS

Comme évoqué précédemment, les glycanes sont nativement hydrophiles et généralement neutres ou négativement chargés (charge apportée par les acides sialiques des N-glycanes complexes et des O-glycanes sialylés). Outre l'ionisation MALDI, les N-glycanes sont également analysés par spectrométrie de masse avec ionisation « électrospray » (ESI-MS). La méthode ESI permet d'ioniser un échantillon à l'état liquide et est donc souvent couplée à des méthodes séparatives (LC, CE, GC...), permettant la résolution temporelle des différents glycanes. De nombreuses publications montrent l'intérêt de cette stratégie analytique pour la glycomique (119–122). Cependant, et plus encore que par analyse en MALDI-MS, l'étude des glycanes par ESI-MS nécessite de modifier les sucres par perméthylation et/ou en marquant leur extrémité réductrice.

Afin de répondre aux défis proposés par l'analyse structurale des glycanes, d'autres stratégies analytiques se sont développées de concert avec les avancées technologiques en spectrométrie de masse. Deux domaines semblent particulièrement actifs et l'innovation dans ce domaine y est forte : les méthodes de fragmentation (MS/MS) et la séparation par mobilité ionique couplée à la spectrométrie de masse (IM-MS).

La fragmentation des sucres en MS/MS permet d'élucider la composition et la structure des glycanes. Parmi les méthodes de fragmentation les plus utilisées,

on peut former trois groupes : les méthodes de dissociation par collision avec un gaz neutre (p. ex. CID, *collision-induced dissociation*), les méthodes d'excitation par photons (p. ex. IRMPD, *infrared multiphoton dissociation*), et les méthodes de fragmentation à base de transfert d'électrons (p. ex. ECD, *electron capture dissociation*). Ces méthodes de dissociation produisent deux types de fragments, les fragments de la liaison osidiques (liaison entre les monosaccharides) et les fragments dits « *cross-ring* » (voir paragraphe 1.2 Nomenclature). Les premiers sont des indicateurs de la composition et de l'enchaînement des constituants monosaccharidiques d'un glycanes tandis que les fragments *cross-ring* permettent d'identifier les atomes impliqués dans les liaisons osidiques. L'importance de l'ion apportant une charge au sucre dans sa fragmentation a été étudiée. En effet par fragmentation CID et IRMPD, il semble que les glycanes protonés se fragmentent majoritairement au niveau de la liaison peptidique tandis que les formes ionisées par des métaux (principalement Na⁺ et K⁺) produisent également des fragments *cross-ring* (123). Certains monosaccharides ont naturellement une forte affinité pour les ions Na⁺, de telle sorte qu'il n'est pas rare de ne pas détecter la forme protonée des glycanes lors d'expériences utilisant le mode d'ionisation MALDI sans purification (des matrices et de l'échantillon). Selon le mécanisme de fragmentation, certains groupements instables peuvent être perdus ou transloqués, en particulier lors de la fragmentation de glycanes natifs en mode CID (124). Ce phénomène peut être atténués par la perméthylation des glycanes ou l'utilisation d'autres méthodes de fragmentation. Les méthodes à base de capture ou de transfert d'électron sont plus récentes et semble permettre d'affiner l'étude structurale des glycanes, en particulier en ne provoquant pas la perte de groupements labiles des glycanes sulfatés (125) et en permettant une excellente couverture de la totalité des fragments (liaison osidique et *cross-ring*) (126). Ces modes de fragmentation sont très informatifs mais ne sont pas forcément disponibles sur tous les instruments commerciaux. Surtout, la trop grande abondance de fragments de toutes sortes risque à terme de complexifier l'identification des glycanes si les suites logicielles ne sont pas

capables de prendre en charge une telle quantité de données, en augmentant le risque de faux positifs.

2.2.2.3 Mobilité ionique

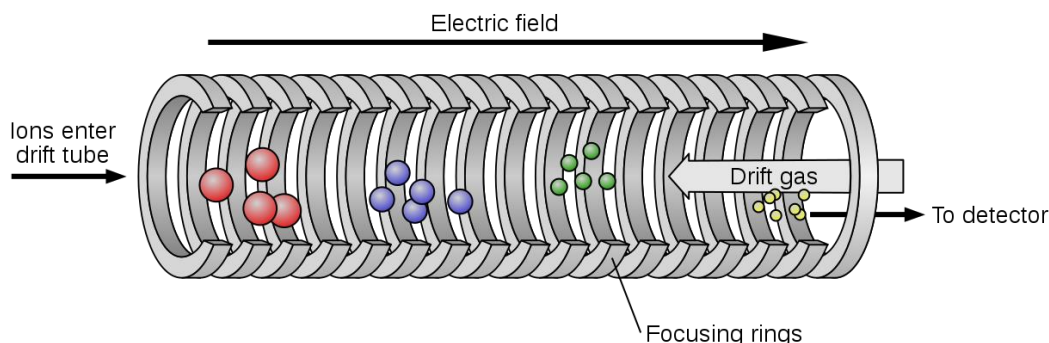


Figure 22 : Diagramme d'une cellule de spectrométrie de mobilité ionique. Les sens de la pression de gaz de dérive (drift gas) et de la tension électrique sont opposés. Les molécules les moins freinées sont celles, pour une masse donnée, ayant la plus petite section efficace de collision.

Image originale : Jeff Dahl.

Le terme de spectrométrie de mobilité ionique (IMS) regroupe en réalité plusieurs designs instrumentaux différents mais qui reposent tous sur la séparation des analytes en fonction de leur mobilité dans un flux de gaz, et donc en fonction de leur section efficace de collision (ou plus grossièrement, en fonction de leur taille)(Figure 22). Cette dimension supplémentaire de séparation permet de séparer des molécules isomères et de produire des spectres de fragmentation uniques et spécifiques de chacun de ces isomères (127–129). Les constructeurs exploitent avantageusement cette propriété puisque la plupart des appareils récents destinés à la protéomique sont désormais proposés avec une forme de séparation par mobilité ionique. Cependant, les suites logicielles dédiées à la glycomique ne sont pas au point et nécessitent d'être mises à niveau pour exploiter pleinement ces nouvelles technologies.

2.2.2.4 Quantification

La quantification des glycanes fait face aux défis inhérents à la quantification par spectrométrie de masse. L'efficacité d'ionisation des analytes dépend de la nature de l'analyte (la présence de groupements ionisés ou ionisables, l'hydrophobie, la masse...) ce qui exclut de ce fait toute quantification absolue

ou la comparaison d'espèces différentes. Le plus souvent, la quantification des glycanes est faite par marquage fluorescent et détection après une dimension séparative (HPLC par exemple). Par spectrométrie de masse, la relative labilité des acides sialiques et des sulfates, les rendements de marquage chimique des sucres, la présence de différents états de charge en ionisation ESI ou la cristallisation non-homogène des matrices en MALDI rend presque impossible la quantification absolue des glycanes en MS (130). De ce fait, des approches de quantification relative sont le plus souvent utilisées, et permettent de comparer les intensités relatives des signaux de certains glycanes ou certaines familles de glycanes entre différents échantillons.

Cependant il faut soulever plusieurs problèmes majeurs à ces approches. D'abord, la quantification de glycanes seuls pose nécessairement des difficultés du fait de l'omission de la nature glycoprotéique de l'analyte initial. En effet, il est nécessaire de normaliser la quantité de glycoprotéine à déglycosyler pour que les intensités mesurées pour les sucres ne reflètent pas uniquement un changement dans la quantité de glycoprotéine initiale. Lorsque la matrice initiale est composée de plusieurs protéines, cette tâche peut être un frein à la quantification des sucres. Un problème se pose également dans le cas où une protéine possède plusieurs sites de glycosylation potentiels, la quantification des sucres pouvant alors être le reflet d'une altération de l'occupation de ces sites ou d'une maturation (clivage) de la protéine.

Des tentatives de marquage isotopique des sucres ont été entreprises (131,132). Elles ont pour but de normaliser la préparation et l'ionisation des N-glycanes mais ne répondent pas à toutes les problématiques soulevées précédemment.

2.3 Glycoprotéomique bottom-up

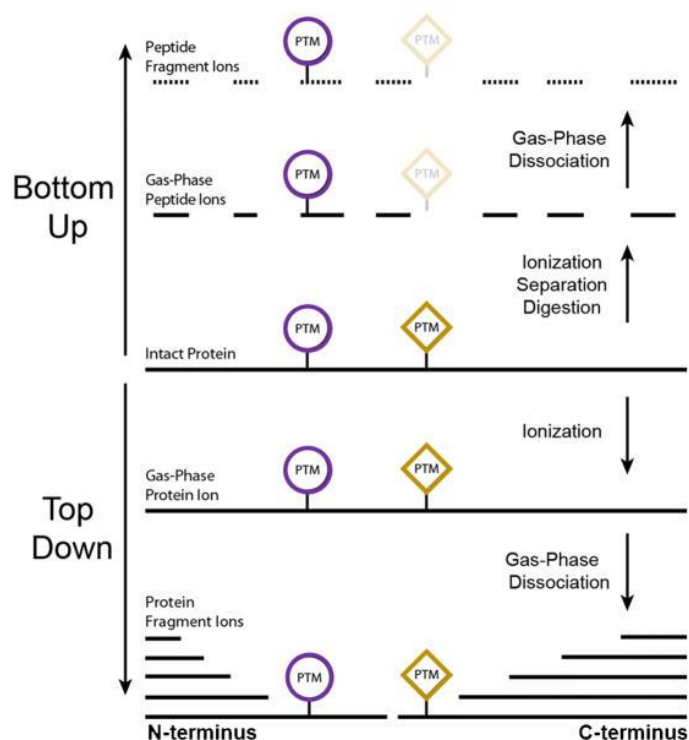


Figure 23 : Schématisation des approches bottom-up et top-down en protéomique.
La couverture partielle des peptides détectables et des fragments obtenus en MS/MS est mise en évidence.

Image originale : Catherman A. D., 2014

L'approche *bottom-up* en protéomique consiste à dégrader les protéines par protéolyse, généralement enzymatique, et à identifier les peptides protéolytiques produits (Figure 23). Cette approche permet de contourner les verrous technologiques actuels en ce qui concerne la séparation, l'ionisation et la fragmentation des protéines intactes. Les avancées dans la résolution des méthodes séparatives, ainsi que la vitesse, la résolution et la sensibilité des spectromètres de masse ont permis au couplage RPLC-MS/MS (chromatographie liquide à polarité de phase inverse couplée à la spectrométrie de masse en tandem) de s'imposer comme méthode de choix pour la protéomique *bottom-up*. Les peptides fragmentés en MS/MS sont séquencés et permettent d'identifier les protéines présentes initialement. Bien que cette approche permette d'identifier efficacement des protéines ainsi que des acides aminés modifiés, ces derniers étant isolés du reste de la protéine et des autres sites de modification éventuels, l'information combinatoire est perdue.

2.3.1 Généralités

2.3.1.1 Préparation de l'échantillon

Toute approche de protéomique nécessite avant tout de s'interroger sur la prise en charge de l'échantillon. La gamme dynamique des spectromètres de masse étant limitée, la surabondance d'une espèce (albumine ou immunoglobulines dans la matrice sanguine, collagène cutané...) masquera souvent la présence d'espèces en plus faible concentration par suppression spectrale. Ainsi certaines applications peuvent nécessiter différentes purifications ou déplétions protéiques en fonction de l'application recherchée. Il serait difficile de décrire de façon générale les méthodes de préparation de l'échantillon tant elles sont nombreuses et souvent spécifiques à chaque application.

On peut néanmoins s'attarder sur la purification des gonadotrophines chorioniques équine et humaine. La hCG a depuis longtemps été purifiée par des méthodes de chromatographie liquide (phase inverse, échangeuse d'ions, exclusion stérique...) permettant d'obtenir un bon enrichissement d'un point de vue biologique avec une augmentation de l'activité stimulatrice de l'hormone (133,134). Ces approches ont été par la suite complétées, notamment par l'utilisation de méthodes de purification par affinité, en particulier dirigées contre les sucres (concanavaleine A et autres lectines)(135) ou contre des épitopes de la β hCG (135,136).

Les protéines sont dénaturées avant leur digestion par réduction de leurs ponts disulfures afin de réduire l'encombrement stérique aux alentours des sites d'action des protéases. Les agents réducteurs les plus courants en protéomique sont le dithiothréitol (DTT) et le β -mercaptoéthanol (β ME), et la dénaturation des protéines est généralement suivie d'un blocage de la réactivité des cystéines libres par un agent alkylant, typiquement l'iodoacétamide (IAM) ou l'acide iodoacétique (IAA). Ces traitements sont pratiqués en amont de la séparation en milieu liquide mais peuvent également être effectués *in gel* dans le cas d'une séparation par SDS-PAGE.

2.3.1.2 Protéolyse enzymatique

Une étape cruciale d'un protocole de protéomique bottom-up consiste à générer des peptides par protéolyse enzymatique. Intuitivement, dégrader la ou les protéine(s) d'intérêt en peptides aurait tendance à augmenter le nombre d'analytes, mais permet en réalité de simplifier leur caractérisation. Bien souvent, une protéine est en fait un mélange de plusieurs isoformes, et les protéines issues d'un milieu donné sont très hétérogènes (masse, pKa...). Leur dégradation en peptides permet de réduire l'étendue de ces caractéristiques et de concentrer cette hétérogénéité sur un ensemble de molécules aux propriétés biochimiques plus homogènes.

Les protéases utilisées sont des enzymes qui hydrolysent la liaison peptidique entre deux acides aminés, avec une sélectivité plus ou moins grande selon l'application (Figure 24). L'utilisation de la trypsine, une protéase à sérine, est depuis longtemps le *gold-standard* de l'analyse protéomique *bottom-up*, car elle produit statistiquement des peptides de taille idéale pour les spectromètres de masse (1000-1500 Da) porteurs d'un résidu basique (Arginine ou Lysine) à leur extrémité C-terminale. D'autres enzymes spécifiques sont également utilisées pour des applications particulières en fonction de la séquence des protéines d'intérêt ou de la localisation des modifications à caractériser. L'augmentation de la vitesse et de la sensibilité des spectromètres de masse, ainsi que la diversification des méthodes de fragmentation, permettent d'envisager l'utilisation de protéases aspécifiques, produisant des peptides de nature variable. De telles enzymes permettent de nouvelles approches pour l'étude de certaines modifications post-traductionnelle, en particulier les glycosylations, et seront discutées ci-après. (137)

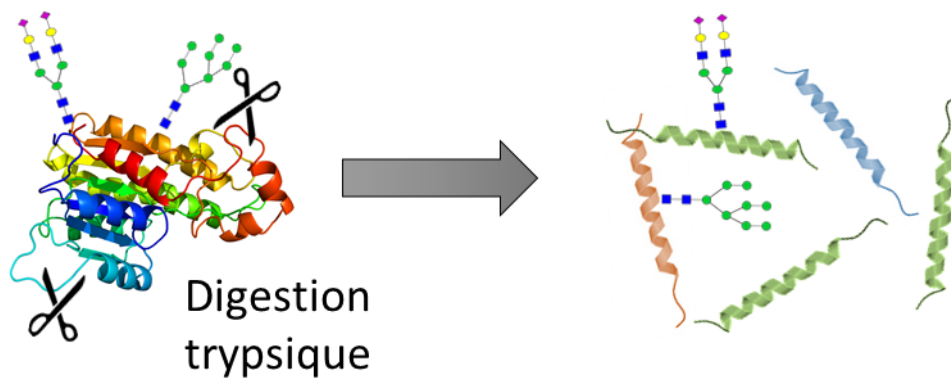


Figure 24 : La digestion d'une glycoprotéine par une protéase (par exemple, la trypsine) produit un mélange de peptides modifiés ou non.

La digestion en solution, dans laquelle l'enzyme et l'échantillon sont incubés dans un milieu tampon, reste l'approche la plus simple à mettre en place. Des variantes à cette méthodologie, en particulier des modifications de paramètres physiques, existent néanmoins et permettent généralement d'améliorer la cinétique de réaction mais également parfois la stabilité de l'enzyme face à son autolyse. En plus des méthodes de digestion sous ultrasons ou micro-ondes qui provoquent une dénaturation des protéines par échauffement local, il est possible d'immobiliser les protéases sur support solide. De tels dispositifs peuvent être miniaturisés et réutilisés, augmentant l'efficacité de la réutilisabilité des enzymes (137). Les protéines séparées par électrophorèse en gel (SDS-PAGE) peuvent être digérées à l'intérieur du gel, permettant ainsi une approche ciblée sur un nombre réduit de protéines ou de protéoformes (138).

2.3.1.3 LC-MS/MS

Le nombre de peptides générés par une digestion enzymatique est d'autant plus grand qu'il ne dépend pas uniquement du nombre et de la taille des protéines digérées. En effet, la microhétérogénéité générée par les modifications « innées » (PTMs) ou les artefacts de manipulation de l'échantillon (dégradation, oxydation etc...) multiplie grandement les isoformes de chaque peptide. Ce nombre est encore augmenté lors du processus d'ionisation qui peut aboutir à la formation de différents états de charge et donc différentes valeurs de m/z .

Les performances actuelles des spectromètres de masse ne permettent pas d'identifier plusieurs dizaines ou centaines de milliers de peptides différents injectés et fragmentés simultanément. Des méthodes séparatives sont utilisées *offline* ou *online* afin de résoudre temporellement les différents analytes et permettre leur caractérisation. Ainsi, on parle souvent de LC-MS/MS pour décrire le couplage d'une méthode séparative en chromatographie liquide avec une méthode de spectrométrie de masse en tandem.

En protéomique *bottom-up* « classique », la chromatographie liquide à polarité de phase inverse sur phase C18 est le plus souvent utilisée de par sa capacité de pic pour l'analyse des peptides. Les peptides sont élués par ordre croissant de leur hydrophobie. Dans ce cas, la phase mobile est constituée d'un mélange d'eau et d'un liquide organique (acétonitrile, méthanol...) en conditions acides. Si une élution isocratique qui utilise une phase mobile de composition constante est convenable pour des mélanges peu complexes, pour des mélanges très complexes, comme fréquemment en protéomique, une élution avec un gradient croissant de phase organique est souvent employée, et permet une meilleure résolution sur une plus large gamme d'analytes.

Le fonctionnement des spectromètres de masse moderne peut être schématisé en deux étapes : une première mesure de masse des ions introduits par la source (MS1 *full scan*), puis, après sélection et fragmentation des précurseurs sélectionnés, une seconde mesure de masse des fragments ainsi produits (MS2 *full scan*). Un des modes d'analyse le plus courant pour l'identification protéique sans a priori utilise une sélection adaptative des précurseurs à fragmenter, ou « *data dependent acquisition* » (ou ddMS2).

Ainsi, les données obtenues en LC-MS/MS comportent au moins trois dimensions, la séparation temporelle des analytes par chromatographie liquide, la mesure de masse des analytes élués et introduits dans le spectromètre de masse, et la mesure de la masse des fragments issus de chaque précurseur sélectionné. À ces dimensions s'ajoutent parfois, selon les

instruments, une séparation par mobilité ionique (IM-MS) ou des fragmentations et analyses de masses supplémentaires (MS3 ou MSⁿ).

2.3.1.4 Séquençage peptidique et identification protéique par MS

La fragmentation des peptides en spectrométrie de masse peut aboutir à la rupture de la liaison peptidique, des liaisons du carbone α de chaque acide aminé, ou de leur chaîne latérale. La nomenclature utilisée pour la dénomination des fragments peptidiques a été introduite par Roepstorff et Fohlman (7)(Figure 25).

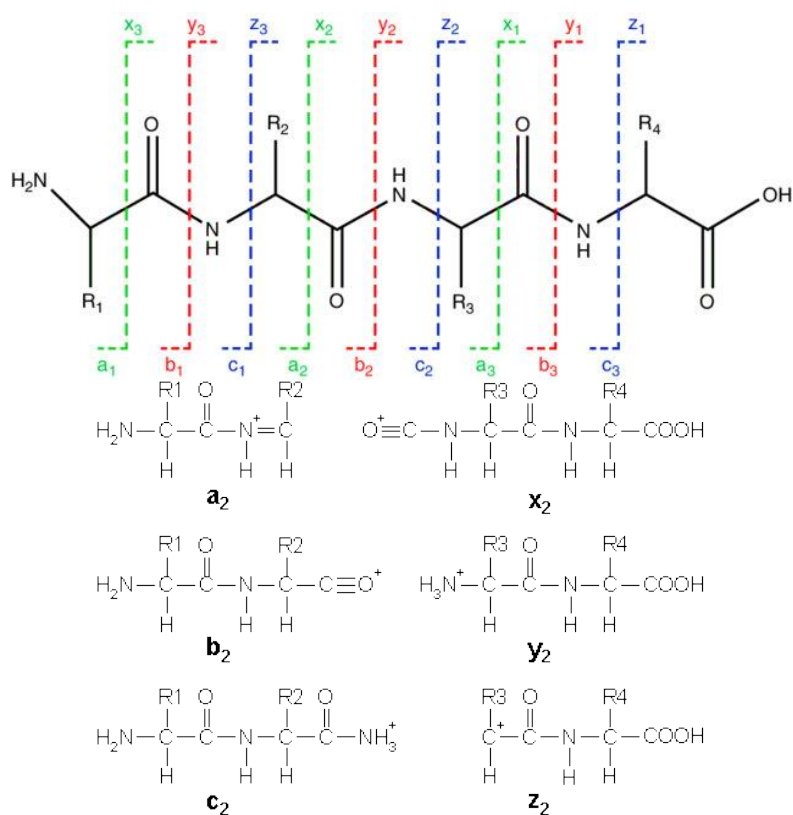


Figure 25 : Nomenclature simplifiée des fragments peptidiques en spectrométrie de masse.
(bas) Structure des fragments communément produits lors de la fragmentation des peptides en spectrométrie de masse. Sources : Zhiqi Hao et al. ; Matrix Science.

Les fragments réellement produits lors de la fragmentation des peptides dépendent de leur séquence et de la méthode de fragmentation utilisée. Par exemple, la fragmentation HCD produira principalement des ions b et y tandis que l'ETD produira des ions c et z. La comparaison des spectres de masse expérimentaux avec des spectres théoriques issus de digestion *in silico* permet de séquencer chaque chaîne peptidique. Les peptides ainsi identifiés

permettent à leur tour d'identifier les protéines présentes dans l'échantillon. Ce traitement bio-informatique nécessite donc l'utilisation de banque de séquences protéiques ainsi que des algorithmes de comparaison et de calcul statistique.

2.3.2 Glycoprotéomique *bottom-up*

La glycoprotéomique *bottom-up* vise à répondre à plusieurs questions, dont l'identification des sites de glycosylation d'une protéine, la détermination de la composition et de la structure desdits glycanes, ainsi que la quantification de chaque glycoforme. Bien que l'analyse de glycoprotéines par approche *bottom-up* soit basée sur les mêmes principes que l'analyse de protéines, la nature et la taille des sucres impose de développer des méthodes spécifiques à l'étude de protéines glycosylées.

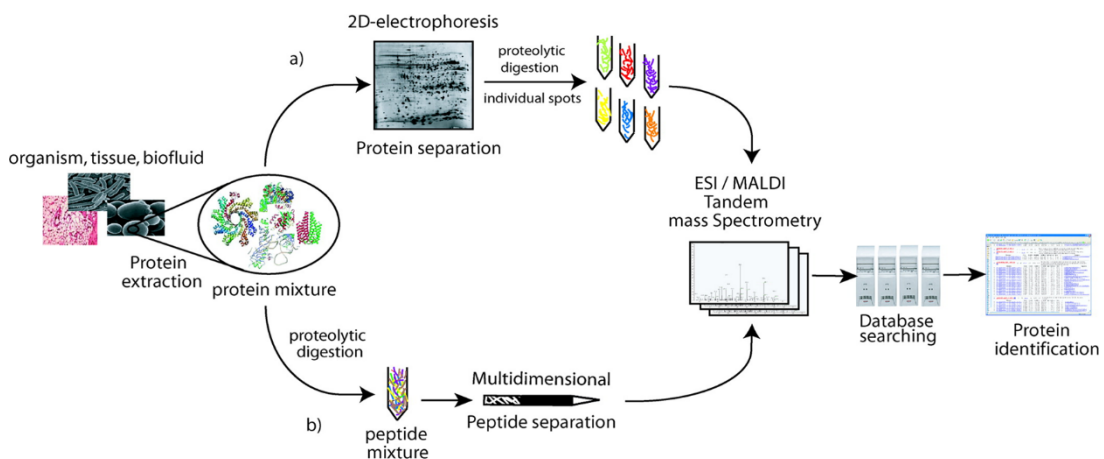


Figure 26 : Etapes classiques d'un protocole de protéomique *bottom-up* (shotgun). (139)

Les N-glycanes forment un adduit de masse de 1500 Da à 2000 Da en moyenne et peuvent dépasser les 3000 Da par polysaccharide, nécessitant d'adapter les méthodes de préparation de l'échantillon et les protocoles de digestion à ces problématiques. La nature chimique des sucres oblige également à adapter les méthodes LC-MS/MS afin de résoudre des molécules faiblement retenues en phase inverse et peu ionisables. Les spectres de fragmentation obtenus nécessitent un traitement particulier tenant compte de tous les niveaux d'information de séquence et de localisation du peptide et de son glycanes.

2.3.2.1 Traitement de l'échantillon et protéolyse

Si la majorité des protéines sont glycosylées, l'hétérogénéité des structures saccharidiques implique souvent que chaque glycoprotéoforme puisse être en réalité bien plus faiblement concentrée. Dans le cadre d'une approche *bottom-up*, la digestion d'une glycoprotéine produit donc généralement un grand nombre de glycopeptides faiblement concentrés et quelques peptides non-modifiés, dont l'efficacité d'ionisation est également bien supérieure à celle des glycopeptides. Ce problème est d'autant plus préoccupant que la glycoprotéine d'intérêt est parfois elle-même minoritaire au sein d'un mélange d'autres protéines.

La plupart des méthodes de purification des glycanes sont généralement adaptables à des approches de glycoprotéomique *bottom-up*, puisque le glycane modifie de manière significative les propriétés physico-chimiques du peptide modifié. Ainsi, plusieurs groupes ont montré l'intérêt de telles méthodes d'enrichissement des glycopeptides pour l'étude des glycoprotéines, basées sur des extractions solide-liquide HILIC (cellulose ou amide), ZIC-HILIC, ERLIC, phase graphitique, oxyde de titane, par enrichissement basé sur des interactions covalentes (hydrazide et acide borique), ou par affinité avec des lectines(140). Les stratégies les plus communément utilisées incluent généralement des enrichissement en mode HILIC ou ZIC-HILIC, sur phase PGC, ou utilisant une ou plusieurs lectines (140–144). Aucune étude ne semble s'accorder sur une approche universelle plus performante et reproductible dans toutes les situations (145,146). Il faut noter que le choix du mode d'enrichissement doit être adapté à chaque application, du fait notamment des biais de sélection de certaines méthodes en faveur de certains glycanes (86,140,145). Ainsi, les lectines sont évidemment spécifiques de certains motifs glycosylés, et les enrichissements sur oxyde de titane sont régulièrement utilisés pour les glycopeptides sialylés. Au contraire, les méthodes d'enrichissement HILIC, PGC, et sur acide boronique sont des méthodes peu ou non-biaisées qui permettent la rétention de tout type de glycopeptide.

Les glycanes sont des modifications de taille relativement élevée. Ainsi, les glycanes d'une protéine glycosylée peuvent compter pour plus de 30 % de sa masse, et un unique glycane peut être plus massif que le peptide protéolytique qui le porte, particulièrement dans le cas des N-glycanes. De ce fait de telles extensions peuvent parfois inhiber l'action enzymatique des protéases par simple encombrement stérique si le site cible se situe au voisinage proche du site modifié. Les glycanes augmentent donc sensiblement la gamme de masse des glycopeptides, de par leur masse propre non négligeable mais également de par l'encombrement stérique qu'ils impliquent. Les spectromètres de masse permettent de détecter, de sélectionner et de fragmenter de manière efficace des peptides de masse proche de 3 kDa.

Selon les applications, une digestion classique par la trypsine peut produire les meilleurs résultats, en particulier dans les cas où l'accessibilité de sites de protéolyse n'est que peu affectée par la présence du sucre, et si la masse des glycopeptides produits reste dans la gamme d'analyse des appareils de mesure. Cependant, la glycosylation de certaines glycoprotéines comme l'hormone chorionique gonadotrope humaine (hCG) est parfois difficile à élucider simplement par la trypsine, comme le montrent les digestions théoriques ci-dessous (Figure 27).

Ces séquences permettent de mettre en évidence deux phénomènes : sur la sous-unité α , un site de N-glycosylation suit directement un site de coupure de la trypsine, et deux sites de O-glycosylation sont présents sur le même peptide tryptique. Il est probable que de telles configurations ne permettent pas la digestion correcte de la région glycosylée sur la séquence (a) et que la double glycosylation d'un peptide ne soit pas correctement interprétable en LC-MS/MS.


```

(a)  1 APDVQDCPEC TLQENPFFSQ PGAPILQCMG CCFSRAYPTP
      41LRSKKTMLVQ KNVTSESTCC VAKSYNRVTV MGGFKVENHT
      81ACHCSTCYYH KS

(b)  1 SKEPLRPRCR PINATLAVEK EGCPVCITVN TTICAGYCPT
      41 MTRVLQGVLP ALPQVVCNYR DVRFESIRLP GCPRGVNPV
      81 SYAVALSCQC ALCRRSTTDC GGPKDHPLTC DDPRFQDSS
      121SKAPPPSLPS PSRLPGPSDT PILPQ

```

Figure 27 : Séquences protéiques des formes annotées des sous-unités α (GLHA_HUMAN)(a) et β (CGB3_HUMAN)(b) de la hCG.

Les sites de reconnaissance de la trypsine sont indiqués en gras et les sites de glycosylation sont mis en évidence en vert (N-glycosylation) et en bleu (O-glycosylation).

De telles configurations justifient l'utilisation de méthodes de digestion alternatives permettant de limiter la masse des glycopeptides protéolytiques à des gammes analysables en MS, et limitant à un unique glycane présent sur un peptide. Pour cela, d'autres enzymes spécifiques ont parfois été utilisées mais leur intérêt dépend évidemment, comme dans le cas de la trypsine, de la séquence des glycoprotéines étudiées. Nous avons récemment montré l'intérêt d'immobiliser les protéases dans des réacteurs de digestion miniaturisés afin d'augmenter sensiblement le rapport enzyme/substrat et donc l'efficacité de digestion des glycoprotéines (147). Cet article présente également la valeur ajoutée apportée par l'utilisation de plusieurs enzymes, ici la trypsine et la pepsine (protéase à plus large spectre de site de protéolyse), pour l'identification de glycoformes. Dans une approche plus poussée, plusieurs groupes ont développé des méthodes utilisant des protéases non-spécifiques (148–153), c'est-à-dire hydrolysant virtuellement toutes les liaisons peptidiques d'une protéine, comme la protéinase K ou la pronase E. Ces dernières hydrolysent sans distinction les liaisons entre les résidus mais leur activité est toutefois affectée par l'encombrement stérique présent sur la chaîne polypeptidique. Ces protéases aspécifiques permettent alors la déplétion des peptides non modifiés d'une protéine, qui sont dégradés en acides aminés ou petits oligopeptides, et ce qui induit un pseudo-enrichissement en séquences modifiées (et en particulier glycosylées) des protéines. Les glycopeptides ainsi

produits conservent une taille analysable, avec jusqu'à 5 acides aminés de part et d'autre du site de glycosylation.

Ces approches non-spécifiques ne sont toutefois pas sans défaut. En effet, les peptides tryptiques ont l'avantage de posséder un site basique et donc une charge positive à leur extrémité C-terminale, qui permet une ionisation et une fragmentation efficace et d'interprétation plus aisée. Ce n'est pas le cas pour les peptides issus de digestion non-spécifique qui nécessiteront des méthodes de MS et de traitement informatique des données adaptées. Les peptides ainsi produits sont également de taille variable, fonction de plusieurs paramètres expérimentaux, comme le temps d'incubation, l'immobilisation de la protéase, ou l'accessibilité de la région digérée. Cette variabilité provoque une perte de sensibilité par dilution du signal associé à une glycoforme sur plusieurs peptides différents. Il est possible de minimiser cette variabilité en optimisant l'efficacité de digestion afin d'obtenir un peptide de séquence constante et de taille minimale. Cependant, cette solution apporte elle aussi un autre problème, puisque des peptides de quelques acides aminés de long peuvent ne pas fournir d'information de séquence suffisamment spécifique pour localiser la modification sur un site d'une protéine sans a priori. Malgré ces limitations, l'approche *bottom-up* utilisant des enzymes non-spécifique a prouvé son utilité pour l'étude des glycoprotéines, en particulier dans le cas d'échantillons purifiés contenant plusieurs protéoformes d'une même séquence protéique.

2.3.2.2 Méthodes séparatives

Comme en protéomique classique, la méthode de référence pour la séparation des glycopeptides préalablement à une analyse en spectrométrie de masse est la chromatographie liquide à polarité de phase inverse sur phase stationnaire C18 (RPLC-C18). Les glycopeptides sont retenus principalement par l'interaction de leur squelette peptidique avec les chaînes alkyles de la phase stationnaire. Les glycanes que portent ces peptides ont un impact faible à modéré sur la force de l'interaction et donc sur le temps de rétention des peptides. Les sucres neutres (hexose, fucose) ont tendance à augmenter

l'hydrophilie des peptides et à réduire légèrement leur temps de rétention en RPLC, tandis que les acides sialiques (NeuAc) augmentent de façon plus sensible l'hydrophobie et le temps de rétention (147,154). Ces caractéristiques peuvent servir à confirmer ou infirmer une identification en LC-MS/MS et minimiser le taux de faux positifs (86). De récentes études semblent montrer la possibilité de séparer des isomères de position par RPLC à haute température (155). Cependant de telles conditions diminuent également la rétention des glycopeptides les moins hydrophobes, et nécessite de ce fait de considérer d'autres approches.

Certains glycopeptides (les plus courts et les plus hydrophiles) ne sont généralement que peu ou pas retenus sur phase inverse. Des méthodes de chromatographie (ZIC-)HILIC/ERLIC (sur phase stationnaire polaire) ou PGC (phase stationnaire graphitique) ont été développées et utilisées pour la séparation de ces glycopeptides. Ces méthodes ont également permis de séparer les isomères de position, en particulier d'acides sialiques mais également de fucose (156,157).

Différents couplages ont également permis de combiner des mécanismes de séparation orthogonaux et ainsi augmenter le recouvrement d'un plus grand nombre de glycopeptides. En particulier, les glycopeptides de la RNase B ont été soumis à un préfractionnement en mode HILIC couplé à une séparation en PGC-LC afin de profiter de la grande sélectivité de ces deux phases. Les glycopeptides ainsi purifiés ont été soumis à une déglycosylation enzymatique (PNGase F) afin d'élucider finement la structure de ces sucres (158). D'autre part, une double séparation (2D-LC) sur phases C18 et PGC a été utilisée par le groupe de M. Wührer (100,159) pour la séparation de glycopeptides issus de digestion non spécifique. Les glycopeptides sont alors retenus et séparés, sur la première dimension (C18) pour les plus hydrophobes, ou sur la seconde dimension (PGC) pour les glycopeptides non retenus en RPLC classique. La séparation de tels isomères de position a également été étudiée par d'autres méthodes, telles que l'électrophorèse capillaire (160,161) ou la spectrométrie de

mobilité ionique (129,162,163). Cette dernière, du fait des récents développements instrumentaux, semble prendre une place importante dans les études actuelles et à venir.

2.3.2.3 Fragmentation

En protéomique classique, les modes de fragmentation classiquement utilisés (CID et HCD) permettant la fragmentation des ions moléculaires par collision avec un gaz neutre (p. ex. diazote, argon, hélium). Ces modes de fragmentation sont idéaux puisqu'ils favorisent la fragmentation robuste des liaisons peptidiques et la production de fragments b et y majoritairement. Ces fragments permettent de séquencer de façon relativement simple les peptides fragmentés, et donc d'identifier les protéines dont ils sont issus.

Cependant, à la différence des peptides, les glycopeptides ne suivent pas les mêmes mécanismes de fragmentation. En effet, la présence d'un N-glycane sur un peptide modifie considérablement la nature et l'intensité des fragments produits lors d'une analyse MS/MS. De plus, une méthode de fragmentation n'affecte pas la portion peptidique d'un glycopeptide de la même façon que sa portion polysaccharidique (164). La fragmentation CID favorisera par exemple la fragmentation du polysaccharide, dont les liaisons glycosidiques sont plus fragiles, aboutissant à la formation d'ions Y informatifs sur la séquence du glycane. Les fragmentations à base d'électrons, comme l'ETD, favorisent quant à elles la formation de fragments du squelette peptidique (en particulier c et z), plus utiles pour la localisation du site de modification. Il est généralement admis que la fragmentation HCD est intermédiaire et permet, en variant l'intensité de l'énergie de collision, l'obtention de fragments du squelette peptidique (b et y) ainsi qu'une série de fragments glycopeptidiques Y et d'ions oxoniums (Figure 28) permettant l'identification de la structure du glycane et sa localisation (165). De plus, les méthodes d'activation à base de transfert d'électron sont généralement plus lentes que les méthodes basées sur la collision avec un gaz, compliquant leur utilisation dans le cadre d'analyse *bottom-up* à haut débit. D'autres méthodes de fragmentation moins courantes, car

disponibles optionnellement sur un nombre réduit d'appareils commerciaux, peuvent également conduire à des fragmentations moins conventionnelles et informatives. Si aucune méthode ne permet la fragmentation informative du glycopeptide dans son intégralité, un consensus semble émerger en faveur de la combinaison de plusieurs méthodes de fragmentation (137,164–167). Ceci nécessite en revanche des appareils « faits-maison » ou des solutions commerciales coûteuses.

La fragmentation HCD est une méthode de fragmentation basée sur la collision des ions moléculaires avec un gaz dans une chambre de fragmentation multipolaire. L'énergie apportée aux ions dans cette cellule augmente leur mouvement et l'excitation vibrationnelle produite par l'impact des précurseurs et du gaz. Il est donc possible de moduler l'ampleur des fragmentations en faisant varier l'énergie de collision appliquée (réglage des tensions du multipôle). Dans les appareils de type QqOrbitrap, les ions sont stockés dans une *C-trap* avant d'être envoyés dans la cellule de fragmentation, puis les fragments sont de nouveau stockés dans la *C-trap* puis éjectés vers l'analyseur de masse principal (Orbitrap). Ce montage permet un mode de fragmentation graduel, dit *stepped-HCD*, dans lequel une partie des ions précurseurs peut être envoyée dans la cellule HCD pour être fragmentée à une énergie donnée. On peut, par cette approche, utiliser jusqu'à trois énergies de collision différentes sur trois tiers des ions accumulés dans la *C-trap*. Tous les fragments ainsi produits sont injectés simultanément dans l'Orbitrap et produisent un unique spectre de masse tandem, qui englobe donc les fragments issus de trois fragmentations à des énergies différentes. L'avantage principal de cette approche est de promouvoir la production de différents fragments et ainsi augmenter la couverture de fragmentation et la qualité de l'identification du glycopeptide. Cette approche a d'abord été utilisée pour l'étude de petites molécules, mais son intérêt pour l'élucidation structurale des glycoprotéofomes est évident. Récemment (153,164), plusieurs études ont montré l'intérêt de la fragmentation *stepped-HCD* dans le cadre de l'analyse de

glycopeptides pour sa capacité à produire des fragments peptidiques, glycopeptidiques, et oxoniums.

Figure 28 : Spectre de fragmentation HCD type (nanoLC-MS/MS) du glycopeptide CamCamCRPIN(+A2G2S1)ATLAVEK.

(page suivante) Les nomenclatures utilisées sont décrites précédemment (cf. 1.2 Nomenclature). Lors de l'existence possible de formes isobares, une seule forme est représentée. Par soucis de clarté, l'intensité des signaux supérieurs à 400 m/z est multipliée par 4. À faible m/z, les ions oxoniums (mono- et oligosaccharides) ainsi que leurs fragments (entre parenthèses), permettent la validation de la composition du glycopeptide. Les ions peptidiques (fragments b en bleu, y en rouge) permettent la localisation du site de modification. Les fragments peptidiques ayant conservé un fragment du glycane sont notés « ~b_i » ou « ~y_i ». Leur structure détaillée est fournie dans le Tableau 1 : Fragmentation HCD type (nanoLC-MS/MS) du glycopeptide CamCamCRPIN(+A2G2S1)ATLAVEK.

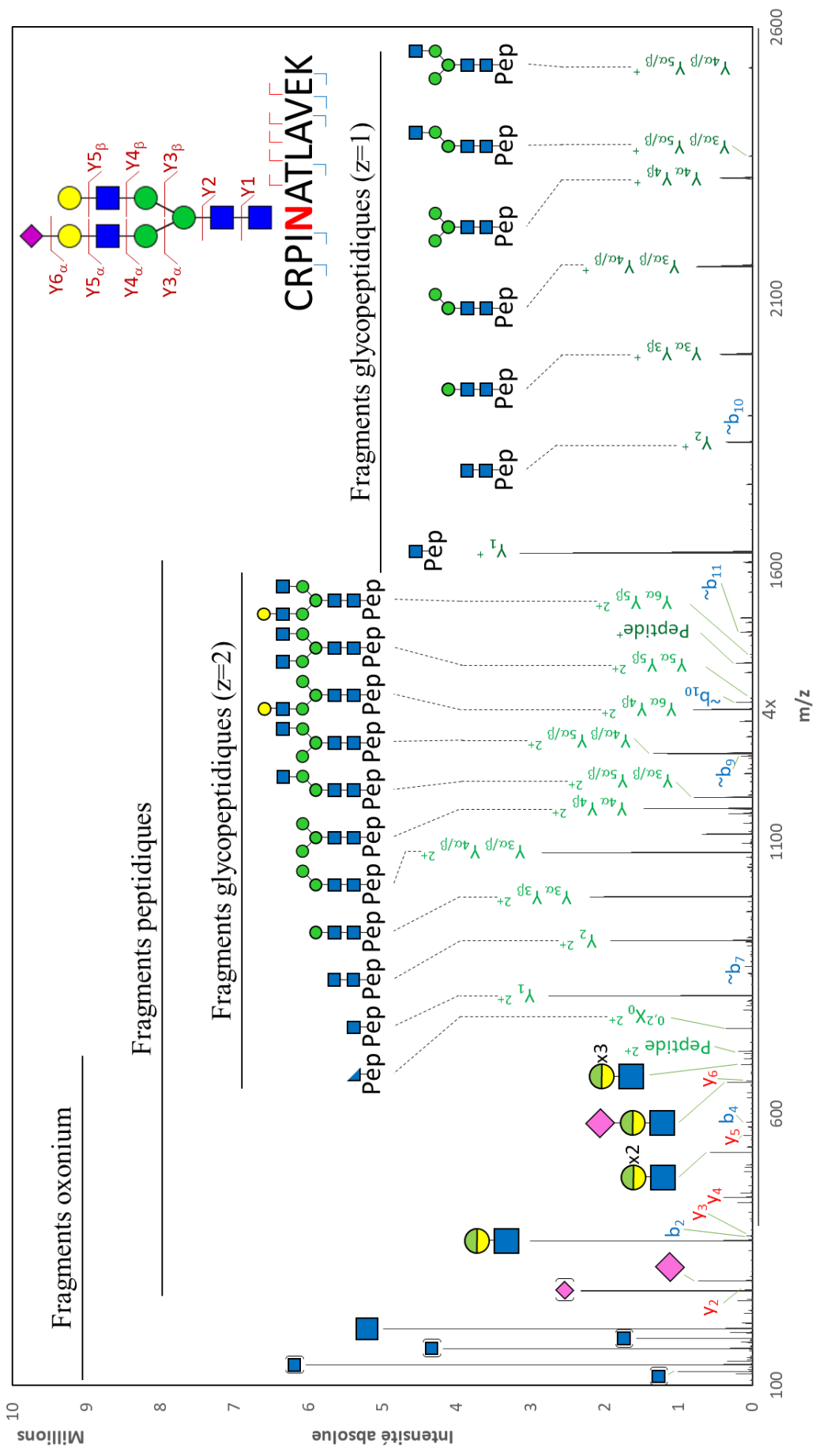




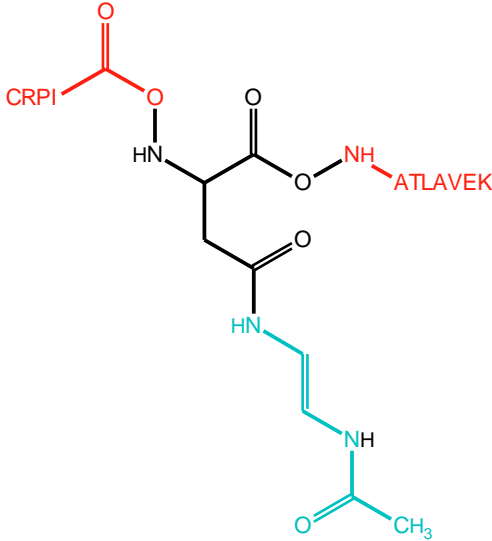


Tableau 1 : Fragmentation HCD type (nanoLC-MS/MS) du glycopeptide
CamCamCRPIN(+A2G2S1)ATLAVEK.

Structures des fragments glycopeptidiques hybrides ayant subi des ruptures de liaison sur leur partie peptidique et leur partie saccharidique. La famille de fragments $\sim b_i$ correspond à la fragmentation d'une ou plusieurs liaisons osidiques, d'une liaison peptidique et de la rétention de charge par le fragment N-terminal. L'ion $^{0,2}X_0$ correspond à la fragmentation « cross-ring » de la première GlcNAc et du réarrangement électronique subséquent. La structure détaillée de l'ion $^{0,2}X_0$ utilisé le code couleur suivant : en bleu le fragment de GlcNAc, en noir l'asparagine glycosylée, en rouge les acides aminés constituant le glycopeptide.

m/z (erreur)	Fragment	Structure
1257,625 (0,3 ppm)	$\sim b_9$	
1356,699 (3,4 ppm)	$\sim b_{10}$	
1485,740 (2,5 ppm)	$\sim b_{11}$	
1883,871 (4,2 ppm)	$\sim b_{10}$	
756,404 (0,8 ppm)	$^{0,2}X_0^{2+}$	

2.3.2.4 Quantification des glycoformes

La glycosylation est une des PTMs les plus abondantes trouvées sur les protéines, en particulier des organismes eucaryotes. De ce fait, ces modifications sont impliquées dans un nombre important de mécanismes cellulaires et de pathologies. Il est cependant avéré que la simple détection (ou

l'absence de détection) d'une glycoforme ne peut pas toujours, à elle seule, constituer une indication sur l'état physiologique de l'organisme étudié. Il est donc nécessaire de disposer de méthodologies permettant la quantification de glycoformes, au niveau des glycoprotéines ou de leurs produits de digestion.

Deux informations quantitatives peuvent être recherchées : le taux de modification d'un site (*site-occupancy* en anglais), et la quantité d'une ou plusieurs glycoformes. Le taux d'occupation d'un site particulier peut avoir une signification biologique majeure, notamment lorsque la glycosylation de ce site est dynamique et remplit un rôle de régulation de l'activité de la protéine (p. ex. dynamique Ser-O-GlcNAc / Phosphosérine) ou lorsque la glycosylation est nécessaire à la fonction de la protéine (O-GalNAc des mucines, N-glycosylation des hormones gonadotropes). La quantification des peptides N-glycosylés et non-glycosylés est généralement effectuée grâce à l'utilisation de PNGase F, qui hydrolyse la liaison peptide-glycane et convertit le résidu Asn en Asp (déamidation, correspondant à un adduit de masse de 1Da). Ainsi le taux de glycosylation peut théoriquement être calculé grâce au dosage du rapport entre les formes déamidée et non-déamidée. Cependant, cette modification survient également spontanément sur les asparagines en solution (168), augmentant artificiellement la quantité de site déamidé et donc le taux de glycosylation calculé (169). Pour contourner ce problème, plusieurs méthodologies ont été proposées. La quantification du peptide non-glycosylé et de la protéine totale (par l'intermédiaire de peptides rapporteurs non modifiables et de peptides standards marqués) permet de s'affranchir du phénomène de déamidation spontanée (170), mais reste tributaire de la connaissance de toutes les isoformes des peptides utilisés pour la quantification. Une déglycosylation partielle, utilisant l'enzyme EndoH (hydrolyse la liaison osidique située entre les deux GlcNAc des N-glycanes oligomannose ou hybrides) puis la PNGase F, permet de quantifier le taux d'occupation des sites de N-glycosylation. La déglycosylation par EndoH laisse un résidu GlcNAc sur le peptide, ajoutant 203 Da à sa masse, tandis que la PNGase F (en eau « lourde » H₂¹⁸O) provoque un adduit de masse de 3 Da par

rapport au peptide non glycosylé. Ces trois signaux permettent de différencier et de quantifier l'occupation des sites de N-glycosylation par des formes *high-mannose*/hybrides ou des formes complexes (171). L'utilisation de standards internes marqués isotopiquement (incorporation d'acides aminés lourds) permet une meilleure appréciation de la quantité absolue mesurée pour chaque peptide et permet d'augmenter la répétabilité des dosages. (86)

Le dosage de la N-glycosylation totale sur les sucres libérés a déjà été évoqué précédemment (*cf.* paragraphe 2.2.2.4 Quantification) et reste problématique sur les échantillons comportant plusieurs protéines glycosylées et/ou plusieurs site de glycosylation par protéine. La quantification à l'échelle des glycopeptides propose une approche plus ciblée et plus significative biologiquement, mais également plus compliquée à mettre en œuvre.

La quantification des glycoformes peut être divisée en deux approches, qu'elle soit obtenue par des quantifications absolue ou relative. La proportionnalité entre l'intensité du signal associé à un glycopeptide donné et sa concentration (dans une matrice identique) permet, par le calcul de l'aire des pics chromatographiques, de quantifier les variations de concentration relative d'un même glycopeptide dans plusieurs échantillons. Cette dernière approche, bien que plus facile à mettre en place, ne permet pas d'identifier totalement des altérations de la glycosylation, et donc complique généralement une transposition au diagnostic clinique qui requiert des données de concentration absolue. La quantification absolue de la glycosylation est idéale mais nécessite de prendre en compte les différences d'expression des protéines modifiées, l'efficacité d'ionisation des différentes glycoformes et leur dépendance aux effets matrice. Nécessitant souvent l'utilisation de standards synthétiques marqués, la quantification absolue de glycoprotéines est limitée à des approches très ciblées.

Les stratégies analytiques peuvent nécessiter la mise en place de méthodes de marquage métaboliques ou chimique, ou utiliser une approche sans marquage (*label-free*)(Figure 29). La méthode sans marquage repose sur la normalisation

du signal de chaque glycopeptide par rapport à une glycoforme ou plus généralement la somme des signaux de toutes les glycoformes considérées. Bien que normalisée, dans une optique comparative cette approche nécessite une excellente maîtrise de la répétabilité de toutes les étapes de préparation et d'analyse des échantillons. Le marquage isotopique permet de différencier deux peptides identiques en leur ajoutant des groupements aux masses différentes en conservant des propriétés physicochimiques quasi identiques. En effet, les isotopes stables « lourds » du Carbone, de l'Azote ou de l'Oxygène utilisés dans ces marquages (^{13}C , ^{15}N , ^{18}O), ont un effet négligeable sur le comportement chromatographique ou l'ionisation des peptides par rapport à leur équivalent « léger » (^{12}C , ^{14}N , ^{16}O). Ainsi, on peut marquer des acides aminés intégrés à la protéine néosynthétisée (culture cellulaire), ou sur les glycopeptides protéolytiques, puis les mélanger afin de minimiser les biais introduits lors de toute manipulation de l'échantillon ultérieure au mélange des espèces marquées. Un exemple de méthode de quantification par marquage chimique des glycopeptides est la méthodologie TMT (*tandem mass tag*) qui consiste à dériver les peptides, généralement sur leur extrémité aminée, par des molécules de masse constante mais qui produisent des fragments différents. Ainsi un unique précurseur produira des fragments rapporteurs identifiables en spectrométrie de masse en tandem MS/MS, dont les intensités permettent de calculer l'intensité relative du glycopeptide dans chaque échantillon. Cette approche a notamment été employée pour l'identification et la quantification de la N- et O-glycosylation de produits pharmaceutiques de hCG.

Le fonctionnement des spectromètres de masse et le mode de sélection des précurseurs pour leur fragmentation a également été étudié. Deux grands types de déclenchement de la fragmentation existent pour des expériences sans *a priori* (en anglais, « *discovery experiment* ») : soit l'acquisition est dépendante du signal mesuré en MS1 (*data-dependent acquisition*, ou DDA), soit les fenêtres d'isolation des précurseurs sont définies à l'avance (*data-independent acquisition*, ou DIA). Cette dernière approche nécessite des appareils plus performants, rapides, résolutifs, et des algorithmes de recherche des bibliothèques de

spectres efficaces. Néanmoins cette approche permet la sélection et la fragmentation de tous les peptides, permettant une exploration complète du protéome. L'approche DIA limite le risque de saturer la capacité de sélection des précurseurs des appareils, et permet la fragmentation systématique des mêmes précurseurs et donc une meilleure reproductibilité de la quantification. En plein essor et en régulière évolution, beaucoup de travail est fourni pour permettre l'utilisation à grande échelle du DIA, mais ces méthodologies ne sont pas encore couramment utilisées en recherche clinique. A l'inverse, le DDA consiste en la sélection dynamique et la fragmentation uniquement des signaux détectés en MS. Pour l'identification sans *a priori* de nouveaux biomarqueurs, la maturité des méthodologies DDA (analyseurs, méthodes de fragmentation, traitement de données) permet généralement d'obtenir de meilleurs résultats qui sont mieux acceptés par la communauté (172). Quant à elles, les approches ciblées consistent en la sélection et la fragmentation d'un nombre réduit et prédéfini de peptides. Le dosage de glycoformes connues par ces méthodes (SRM, PRM...) permet une sensibilité inégalée parmi les techniques de spectrométrie de masse grâce haute spécificité apportée par la double sélection d'un précurseur et d'un ou plusieurs fragments.

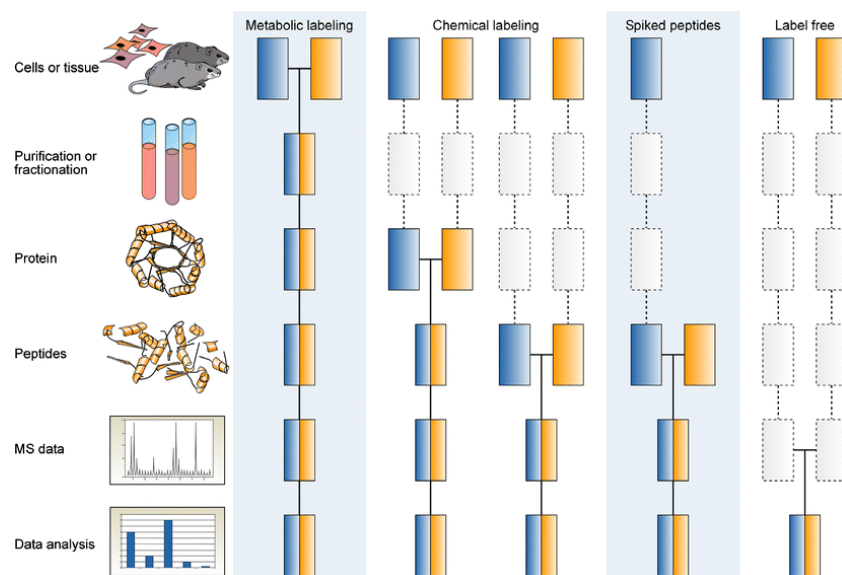


Figure 29 : Différentes approches de protéomique quantitative, avec marquage métabolomique ou chimique, ou sans marquage.

Les étapes précédant le marquage (et le mélange des échantillons) se doivent d'avoir une répétabilité maximale pour minimiser les biais introduits pendant la manipulation de l'échantillon. Un marquage précoce est généralement préférable quand il est possible. (173)

2.4 Approches intacte et *top-down*

La digestion protéolytique ou la déglycosylation des glycoprotéines permet, en étudiant des glycopeptides ou des glycanes, d'obtenir des informations précieuses sur la structure et la localisation de ces modifications. Ces méthodes ont été développées continuellement depuis plusieurs années et sont aujourd'hui très utilisées en *discovery* pour leurs performances (sensibilité, méthodes de séparation et de fragmentation). Cependant, la digestion dégrade la qualité de l'information en décorrélant physiquement les modifications isolées sur des peptides distincts. Ainsi des ensembles de modifications exclusives, c'est-à-dire une situation où la présence d'une modification empêche une autre, ou inclusives, autrement dit deux modifications étant toujours simultanément absentes ou présentes, ne peuvent être analysées que lorsqu'elles sont conservées sur une unique molécule. Pour ces applications, toute digestion est à proscrire. Il est également nécessaire de prendre en compte la perte possible de certains peptides modifiés, du fait de leur non-rétention lors des procédures d'enrichissement ou de séparation, de leurs propriétés d'ionisation ou bien de leur masse inférieure à la limite basse de détection par spectrométrie de masse. L'analyse de protéoformes par MS et MS/MS, dites descendante (ou approche *top-down*) permet de contourner ces limitations. Bien que ce manuscrit ne se focalise pas sur l'analyse de protéoformes, ce domaine est en plein développement pour l'analyse de glycoprotéoformes, entre autres, et mérite d'être développé.

Depuis quelques années, les avancées techniques et technologiques permettent l'étude de ces protéoformes par spectrométrie de masse à haute résolution. En effet, l'étude de protéines intactes induit nécessairement des besoins matériels et logiciels particuliers : méthode de séparation résolutive, haute résolution en masse, sélectivité de l'isolation du précurseur, méthodes de fragmentation alternatives, grande vitesse d'acquisition, *etc.*

Une protéine a tendance à s'ioniser avec un nombre important de charges (jusqu'à plusieurs dizaines pour des protéines de taille moyenne). Ainsi, l'écart

de m/z entre deux isotopes successifs d'une même espèce se retrouve fortement réduit. Si elle n'est pas nécessaire à la détermination de l'état de charge, la résolution des isotopes est tout de même très utile, et permet également de calculer la masse des ions détectés. La séparation entre deux espèces de masses proches est dans ce cas également fortement diminuée et nécessite de minimiser la fenêtre de sélection afin de limiter la co-sélection et co-fragmentation de multiples précurseurs. Contrairement aux peptides dont la fragmentation est relativement aisée et maîtrisée aujourd'hui, les protéoformes sont généralement multi-modifiées par des groupements aux caractéristiques très hétérogènes (glycosylation, phosphorylation, oxydation, acétylation...). Cette grande diversité intramoléculaire combinée à la masse importante des précurseurs protéiques conduit à un nombre de mécanismes de fragmentation important. La multitude de fragments ainsi produits dilue d'autant plus la quantité de signal obtenu et bride la sensibilité des approches *top-down*. L'augmentation du temps d'accumulation et d'acquisition peut améliorer le rapport signal/bruit et la sensibilité mais cette diminution de la vitesse de scan peut être délétère. Les méthodes de purification et de séparation (chromatographie liquide...) n'étant que peu résolutive pour les protéoformes souvent structurellement proches, les analyses LC-MS *top-down* souffrent d'une co-élution de multiples espèces, qui nécessite donc une vitesse de scan rapide.

Plusieurs modifications ou améliorations des méthodologies et instruments ont été introduites aux cours des dernières années pour permettre le développement de la protéomique *top-down*.

Les méthodes séparatives sont un des points d'amélioration critique de toute analyse *top-down*. En effet les glycoprotéines ont des solubilités plus faibles et moins prédictibles que les glycopeptides, par exemple (174). Du fait de leur forte hydrophobie, les protéines de masse moléculaire importante ne sont généralement pas séparées en RPLC sur phase stationnaire C18, car le risque de rétention irréversible est important, mais sur des phases moins hydrophobes (C4, C8...) et avec une résolution plus faible. Les modifications affectant les

protéines n'ont généralement que peu d'impact sur les propriétés globales de celles-ci du fait de la masse négligeable de ces modifications face à celle des protéines entières. Il semble que le multiplexage de techniques de séparation orthogonales permette d'augmenter le nombre de protéines séparables et identifiables (Figure 30). L'utilisation de méthodes de préfractionnement est également souhaitable lorsque l'échantillon est un protéome complexe.

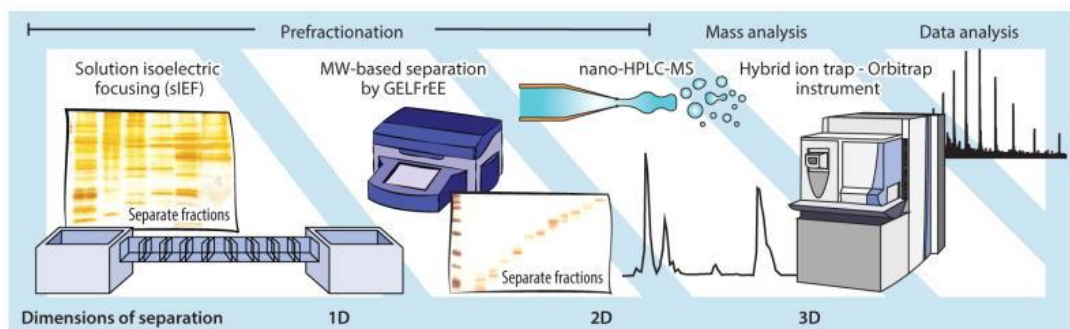


Figure 30 : Séparation multi-dimensionnelle en protéomique top-down.

La combinaison de plusieurs approches complémentaires et/ou orthogonales permet d'augmenter la couverture du protéome d'une expérience. Image originale : « Progress in Top-Down Proteomics and the Analysis of Proteoforms », Timothy K. Toby (2017).

Du point de vue des spectromètres de masse, les appareils à très haute résolution (FT-ICR ou Orbitrap principalement) ont progressé de manière importante durant cette décennie. Le pouvoir résolutif peut dépasser le million pour des temps d'acquisition de plus en plus courts (aujourd'hui de l'ordre de la seconde pour de telles résolutions extrêmes). La sensibilité des méthodes *top-down* peut également être augmentée en moyennant plusieurs spectres dans le but de réduire le bruit électronique, au détriment d'un temps de scan total plus élevé. La plupart des logiciels permettent de faire cette moyenne post-acquisition, et certains appareils permettent l'exécution automatique de cette opération en cours d'analyse. L'utilisation de géométries de quadrupôles et de tensions optimisées permettent également l'utilisation de gammes de m/z étendues à des valeurs plus élevées. La détection d'ions aux états de charge moindre est donc possible et permet d'augmenter les écarts entre les différents signaux (isotopes et espèces différentes). Les méthodes de dissociation employées en protéomique *top-down* sont les plus ou moins les mêmes que celles utilisées dans des approches *bottom-up*, mais utilisent des paramètres

différents, comme des tensions légèrement différentes et adaptées à la fragmentation de plus grosses molécules. Les méthodes « classiques », HCD, ETD, et ECD souffrent parfois de la production de fragments trop peu informatifs, en plus d'être relativement lentes pour les méthodes à base de transfert d'électron. Le transfert non dissociatif d'électrons, causé par la cohésion moléculaire par des liaisons faibles, est une des causes de ces limitations. Une amélioration proposée est la fragmentation HCD des produits de fragmentation ETD, aboutissant à la production de fragments plus informatifs sur la séquence de la protéine (175). Cette méthode a été appliquée très récemment à l'étude des PTMs (dont la glycosylation) à l'échelle des protéoformes (176,177). L'utilisation de fragmentation à base de radiations ultra-violettes (UVPD) permet également de fragmenter efficacement les glycoprotéines intactes, et nécessite un temps d'activation plus court que les activations par transfert d'électron (ETD, EThcD).

Le traitement de données *top-down* est, sur le principe, identique à l'analyse de spectres peptidiques. Cependant, la complexité des spectres MS/MS de protéines qui contiennent de multiples fragments à hauts états de charge et la multitude de mécanismes de fragmentation utilisés compliquent considérablement le séquençage protéique. La quantité toujours plus importante de données collectées, le nombre de modifications recherchées et le nombre de leurs combinaisons décuplent les temps de calcul informatique.

C. Deuxième Partie :

*N-glycomique pour l'étude des
choriogonadotropines*

1. Développement d'une méthode de N-glycomique sensible pour l'étude de la glycosylation des glycohormones

1.1 Introduction

L'importance de la glycosylation des protéines n'est plus à démontrer, puisqu'elle est impliquée dans de nombreux processus physico-chimiques et biologiques (voir Paragraphe B.1. Glycosylation des protéines et glycohormones). La glycosylation de la hCG en particulier a été corrélée à sa durée de vie dans la circulation sanguine, à l'activation de la cascade de signalisation du complexe hCG-LHCGR, à la reconnaissance de la hCG par le récepteur membranaire TGF β , ou encore à la réponse immunitaire utérine (178,179).

Certains de ces processus biologiques ne semblent pas toujours liés à des structures particulières, mais parfois à une caractéristique globale. La densité de charges (négatives) va par exemple permettre la diminution de la vitesse d'élimination de la hCG (180), tandis que la reconnaissance de glycanes mannosylés a été proposée pour expliquer l'action de la hCG sur la production de macrophages et leucocytes utérins(181).

Il semble donc judicieux d'évaluer la glycosylation de ces protéines dans leur ensemble et pas seulement pour chaque site de glycosylation (approche développée dans la troisième partie de ce manuscrit). L'approche *top-down* consiste à fragmenter la glycoprotéine intacte et d'identifier les fragments pour caractériser la molécule fragmentée. Les spectres produits sont cependant rarement interprétables puisqu'ils contiennent l'information de 6 sites de glycosylation, dans le cas de la hCG β . Les techniques disponibles à ce jour ne permettent que d'effleurer la complexité et l'hétérogénéité de telles protéines hautement glycosylées. La glycomique semble être une alternative convenable pour étudier la glycosylation d'un échantillon dans sa globalité. Dans le cas d'une protéine purifiée, l'approche glycomique permet notamment d'évaluer l'ensemble de ses structures N-glycosidiques, et d'obtenir une image moyenne de la population des glycoprotéofomes de cette protéine.

Cependant, certaines propriétés propres aux glycanes sont délétères à leur caractérisation par spectrométrie de masse, en particulier leur hydrophilie et leur faible volatilité, leur efficacité d'ionisation inférieure aux peptides et protéines, ainsi que la labilité de certains monosaccharides conduisant à leur fragmentation en source. Différentes dérivations chimiques des glycanes ont été développées depuis des années pour permettre d'augmenter la sensibilité des méthodes de détection (fluorescence, MS, *etc.*). Ces approches ont été plus longuement discutées dans la partie B.2.2.1.2 Modification chimique des glycanes, et comprennent notamment la modification de l'extrémité réductrice des N-glycanes, permettant d'augmenter leur efficacité d'ionisation et rendant leur fragmentation en MS/MS plus informative.

L'équipe de Madame Yuko Fukuyama a développé une matrice composée de 3-aminoquinoline et d'acide *para*-coumarique pour l'étude des glycanes (et des glycopeptides) en MALDI-MS (115). Cette matrice cumule plusieurs avantages notables, l'aminquinoline en excès permet notamment une conversion totale des glycanes en leur base de Schiff, et agit également comme un intermédiaire pour l'excitation laser en MALDI, tandis que l'acide *para*-coumarique permet la suppression de la dissociation des acides sialiques et sulfates des glycanes. Elle conserve également un état liquide et homogène sous vide, permettant une meilleure répétabilité des mesures. De plus, l'utilisation de cette matrice dans un protocole de N-glycomique permet la dérivation des glycanes et leur « co-cristallisation » avec la matrice en une seule étape, puisque le marquage est effectué sur la cible MALDI servant à l'analyse MS. Ces caractéristiques pourraient permettre d'intégrer l'utilisation de cette matrice à des protocoles hospitaliers, afin de permettre une caractérisation à haut-débit des N-glycanes potentiellement corrélés à une pathologie ou à un niveau d'activité de la hCG. Cette approche a été introduite au laboratoire dans le but d'étudier les N-glycanes des choriogonadotropines.

1.2 Développement de la méthode

L'utilisation du mélange 3-aminoquinoline/acide *para*-coumarique en tant que matrice MALDI pour les glycanes a fait l'objet de plusieurs publications ces dernières années (115,117). Afin de valider les performances de cette approche par rapport à l'utilisation d'une matrice DHB, nous avons d'abord reproduit le protocole tel qu'il a été publié initialement.

1.2.1 Reproduction du protocole publié

1.2.1.1 Matériel et méthode

Une reproduction du protocole publié a été entreprise sur les N-glycanes de la RNase B (*bovine pancreas*, Sigma-Aldrich), protéine commerciale glycosylée par des structures majoritairement de la famille des *high-mannose*. Elle a été dé-N-glycosylée suivant un protocole habituellement utilisé au laboratoire. La protéine solubilisée a d'abord été soumise à un traitement dénaturant afin de réduire ses ponts disulfures et augmenter l'accessibilité des glycanes à la PNGase. Chaque échantillon a été dilué dans une solution de bicarbonate d'ammonium (50mM) dans des proportions 1 : 1 (v : v) puis incubé à 56°C pendant une durée de 30 minutes en présence de DTT (20 mM).

Un traitement par 2 unités de PNGase F (Roche) pendant 3 heures à 37°C a permis la déglycosylation de la RNase B. La purification des N-glycanes a été effectuée par précipitation des protéines par ajout de 4 volumes d'éthanol absolu et leur incubation à -80°C pendant une durée minimale de trois heures. Les glycanes, solubles, sont alors séparés des protéines précipitées par centrifugation (15 000g, 4°C, 30 minutes). Afin de se replacer dans les conditions publiées, le solvant de la solution de N-glycanes a été évaporé et remplacé par de l'eau désionisée.

Le protocole de préparation de la matrice 3-AQ/*p*-CA est identique à celui publié par l'équipe de Y. Fukuyama (dénommé « protocole de référence » ci-après) (115). Une solution de phosphate de monoammonium (ADP, *Ammonium Dihydrogen Phosphate*) est préparée à une concentration de 2 mM dans un mélange eau/acétonitrile (v/v). Cette solution sert à solubiliser 10 mg de *p*-CA

dans un volume de 600 μL , qui sert à son tour à la solubilisation de la 3-AQ (20 mg/150 μL). Cette matrice « concentrée » est diluée 10 fois en ADP (2 mM, eau/acétonitrile). Les glycanes sont mélangés volume à volume avec cette matrice, puis 1 μL du mélange est déposé sur une cible MALDI (méthode de la goutte séchée).

Les analyses MALDI-TOF ont été effectuées sur un appareil AB Sciex MALDI TOF-TOF 5800 dans des conditions classiquement utilisées en glycoprotéomique. L'appareil opère en mode réflectron et la polarité est réglée sur le mode positif pour ces glycanes *high-mannose*. La gamme de masse de 900 à 4000 m/z couvre les masses usuelles des N-glycanes. L'intensité laser a été ajustée extemporanément de façon à obtenir un rapport signal/bruit correct. Chaque spectre est produit à partir de 10000 tirs laser, opérant avec un déplacement continu.

1.2.1.2 Résultats

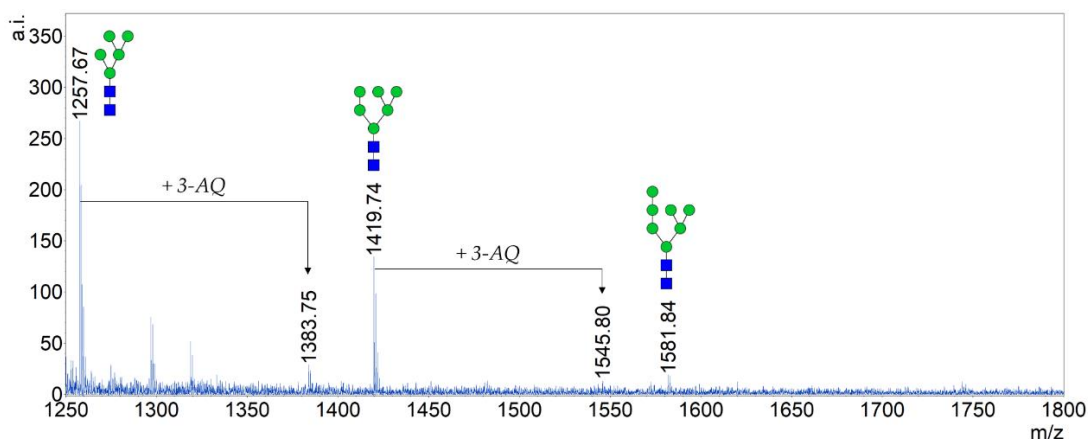


Figure 31 : Spectre MALDI-TOF MS, N-glycanes issus de la déglycosylation enzymatique de la RNase B, protocole de référence.

Les structures SNFG des glycanes natifs sont représentés en vis-à-vis des pics identifiés. L'adduit covalent de 3-AQ aux extrémités réductrices des glycanes est indiqué par une flèche (+ 126 Da).

La reproduction du protocole publié (Figure 31) n'a pas permis, dans les conditions expérimentales propres à ce projet, d'obtenir des résultats concluants. Le rapport signal/bruit est en effet assez faible et ne permet pas, dans les conditions de l'analyse, de détecter les formes à plus de 7 résidus mannose. La RNase B porte pourtant des N-glycanes *high-mannose* jusqu'à 10

résidus mannose. De façon plus préoccupante, le taux de conversion des glycanes en leur base de Schiff (dérivation de l'extrémité réductrice) est très faible, le rapport entre les intensités des glycanes marqués et natifs n'étant que de 1 : 10.

Plusieurs différences entre la publication de Y. Fukuyama *et al.* peuvent probablement expliquer la qualité de ces spectres. En effet, avec l'objectif de pouvoir transposer ce protocole à l'étude de la N-glycosylation de la hCG, nous avons utilisé des glycanes issus de la déglycosylation enzymatique de la RNase B, et non pas des N-glycanes issus de synthèse chimique. De ce fait, les échantillons ont pu contenir des résidus de nature peptidique n'ayant pas été éliminés lors de l'étape de précipitation, mais également des ions carbonate et ammonium. Ces derniers, notamment, peuvent interférer avec la réaction entre l'aminoquinoline et le groupement carbonyle terminal des glycanes. De plus, l'utilisation de cible MALDI μ Focus, permettant notamment de concentrer l'échantillon dans une surface plus réduite, n'a pas ici été entreprise.

Afin d'augmenter la sensibilité de la détection et l'efficacité de marquage des N-glycanes, plusieurs paramètres ont été sélectionnés pour optimiser le protocole de préparation des échantillons. Deux variables ont été suivies pour déterminer les performances desdits protocoles, d'abord le rapport entre les intensités des formes natives et marquées d'un glycanes donné, reflétant l'efficacité de la réaction de dérivation, puis le rapport signal/bruit des N-glycanes détectés et donc la sensibilité de la méthode.

L'optimisation de ce protocole a été alors réalisée au laboratoire.

1.2.2 Optimisation de la méthode

1.2.2.1 Protocoles

Des glycoprotéines commerciales ont été utilisées pour la mise au point du protocole. En particulier, la RNase B (*bovine pancreas*, Sigma-Aldrich) et la Fétuine (*fetal bovine serum*, Sigma-Aldrich) ont été utilisées, car notamment glycosylées par des N-glycanes *high-mannose* et complexes, respectivement. De

plus, la masse moyenne de la RNase B (15 kDa) est proche de celle de la sous-unité α -hCG, tandis que la Fétuine est une glycoprotéine de 64 kDa environ, porteuse de 3 sites de N-glycosylation et 6 sites de O-glycosylation documentés, la rapprochant structurellement de la hCG dimérique. Ces deux glycoprotéines ont donc semblé être de bons modèles pour la mise au point d'un protocole MALDI en vue de l'analyse de la N-glycosylation de la hCG.

Dans le but de déplacer l'équilibre de la réaction de dérivation vers la formation de l'imine (base de Schiff), la quantité de 3-AQ et/ou de *p*-CA ont été augmentées dans la matrice. De plus, certaines applications semblent montrer qu'une activation thermique permet d'augmenter l'efficacité du marquage (182), nous avons donc ajouté une étape de chauffage au protocole initial. Enfin, l'utilisation d'acide permettant d'augmenter l'efficacité d'ionisation des N-glycanes co-cristallisés avec une matrice DHB, nous avons également évalué l'utilisation d'acide acétique 1% comme solvant de re-solubilisation des sucres après déglycosylation et purification (Figure 32).

<i>Matrice</i>	<i>Matrice de référence</i>	<i>Double AQ (40 mg/150 μL)</i>	<i>Matrice 10X</i>
<i>Solvant de reprise des N-glycanes</i>	<i>Eau désionisée</i>		<i>Acide acétique 1%</i>
<i>Mode de dépôt (matrice/échantillon)</i>	<i>1 : 1</i>		<i>3 : 1</i>
<i>Activation thermique sur cible MALDI</i>	<i>Aucun traitement</i>		<i>Traitement 1h à 60°C</i>

Figure 32 : Paramètres testés pour l'optimisation du protocole d'analyse des N-glycanes par MALDI-TOF avec une matrice liquide 3-AQ/*p*-CA.

1.2.2.2 Résultats

Les performances des protocoles sont évaluées sur deux variables quantifiables, le rapport signal/bruit du N-glycane M5-AQ et M6-AQ (pour les échantillons issus de la RNase B) ou du glycane A2G2S2-AQ (pour les N-glycanes de la fétuine), et le ratio entre les intensités de leur forme marquée et

leur forme native. Ces sucres ont été choisis car, de masse relativement faible, la probabilité que les formes natives et dérivées soient simultanément détectées est plus grande, permettant effectivement de calculer un rapport et évitant de devoir s'accommoder aux valeurs manquantes.

Tableau 2 : Conditions de préparation de la matrice 3-AQ/p-CA testées.

Les N-glycanes utilisés sont issus de la déglycosylation de la RNase B. R : matrice de référence ; A ; Matrice enrichie en 3-AQ ; X : Matrice non diluée 10X ; H⁺ : N-glycanes resolubilisés en acide acétique 1% ; E : N-glycanes resolubilisés en eau. Les conditions en bleu sont issues de la méthode de référence.

Condition	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Matrice	R	R	A	A	R	R	A	A	X	X	R	R	A	A	R	R	A	A	X	X
Solvant	H ⁺	H ⁺	H ⁺	H ⁺	E	E	E	E	E	H ⁺	H ⁺	H ⁺	H ⁺	H ⁺	E	E	E	E	E	H ⁺
Matrice : échantillon (x : 1)	1	3	1	3	1	3	1	3	1	1	1	3	1	3	1	3	1	3	1	1
Activation thermique	60°C pendant 1h										Aucune									

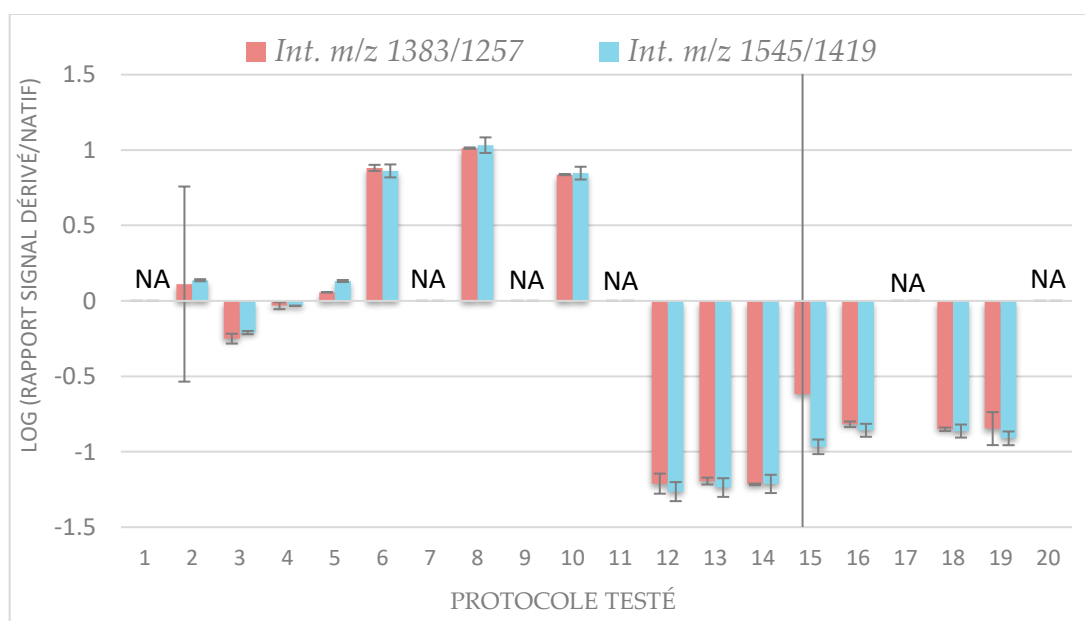


Figure 33 : Optimisation du protocole, efficacité de dérivation.

Ratios (linéarisés \log_{10}) entre l'intensité du signal de la forme marquée (M5-AQ, m/z 1383, ou M6-AQ, m/z 1545) sur celle de la forme native (M5, m/z 1257, ou M6, m/z 1419) pour chacune des conditions testées. Les barres d'erreur représentent l'intervalle de confiance $[M \pm (\sigma/M)]$. Le ratio M5-AQ/M5 dans la condition 15 est aberrant.

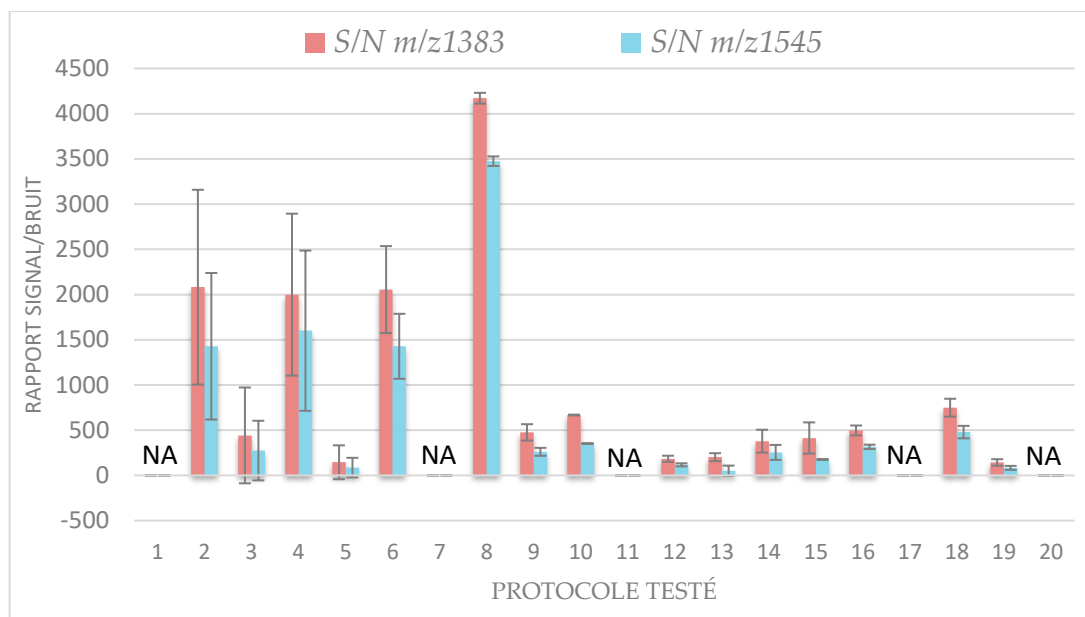


Figure 34 : Optimisation du protocole, sensibilité.
 Valeurs de signal/bruit des glycanes dérivés M5-AQ (m/z 1383) et M6-AQ (m/z 1545). Les barres d'erreur représentent l'intervalle de confiance $[M \pm \sigma]$.

L'analyse des résultats (Figure 33 & Figure 34) montre avec évidence que l'activation thermique est nécessaire et augmente fortement le taux de marquage et donc l'intensité (et le rapport signal/bruit) des glycanes dérivés recherchés (M5-AQ et M6-AQ). Ce résultat était attendu et recherché. Il apparaît également qu'une augmentation de la proportion de matrice et d'aminoquinoline dans l'échantillon permette une meilleure efficacité de marquage, en augmentant la quantité de 3-AQ dans la matrice, en ne diluant pas la matrice avant son utilisation, ou bien en augmentant le rapport matrice/échantillon. L'utilisation d'un solvant acide pour la re-solubilisation des N-glycanes après leur purification n'a pas donné de résultats satisfaisants, notamment du point de vue de l'efficacité de dérivation.

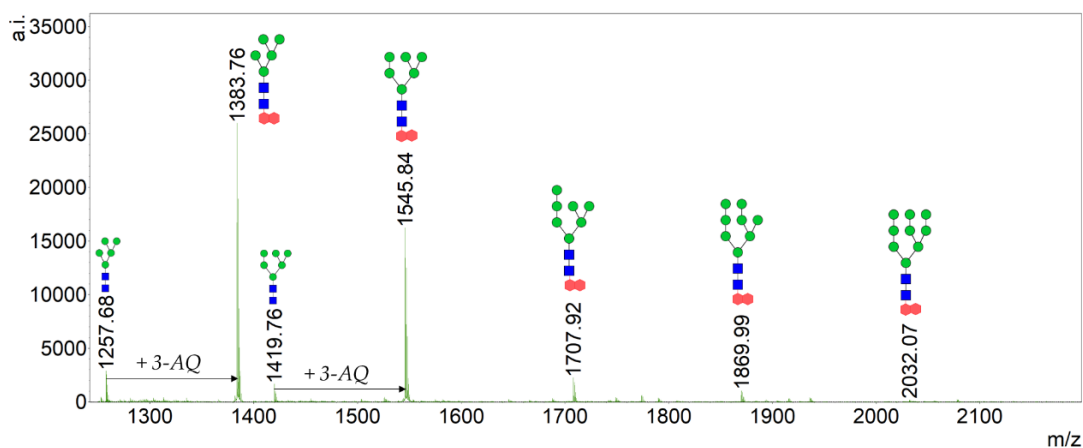


Figure 35 : Spectre MALDI-TOF MS, N-glycanes issus de la déglycosylation enzymatique de la RNase B, protocole optimisé.

Les structures SNFG des glycanes natifs sont représentés en vis-à-vis des pics identifiés. L'adduit covalent de 3-AQ aux extrémités réductrices des glycanes est indiqué par une flèche ainsi que par deux hexagones rouges (+ 126 Da).

Le protocole n°8 est le protocole donnant les meilleurs résultats (Figure 35) : il consiste à augmenter la quantité de 3-AQ dans la matrice (soit 40mg/150µL) ainsi que le rapport matrice/échantillon à une valeur de 3 : 1. Dans ce cas les glycanes solubilisés en eau désionisée et mélangés à la matrice sont déposés sur la cible métallique et chauffés à 60°C pendant une heure avant analyse MALDI-TOF.

Ces modifications ont permis la multiplication du rapport signal/bruit par un facteur 19 entre le signal initial de l'espèce M5 et le signal du glycanes dérivé M5-AQ après optimisation, ce dernier n'étant que peu détecté dans les conditions initiales. Le taux réel de marquage n'est pas calculable par spectrométrie de masse du fait de la différence d'efficacité d'ionisation entre les espèces natives et marquées, mais l'intensité de leur signal permet de remarquer le déplacement de l'équilibre de réaction en faveur de la formation du dérivé. En effet, le rapport entre l'intensité du signal de M5-AQ sur celle du glycanes native M5 passe de 1 : 10 dans les conditions initiales, à plus de 9 : 1 dans les conditions optimisées.

1.2.2.3 Performances de la matrice 3-AQ/p-CA optimisée

Ce nouveau protocole a nécessairement dû être comparé au protocole en place au laboratoire, c'est-à-dire l'utilisation d'une matrice DHB (Figure 36).

Une très forte augmentation du signal ($\times 25$) est observée, mais doit cependant être relativisée. En effet, le bruit est également plus important, aboutissant à des rapports S/N (signal/bruit) multipliés par 2 à 5 environ selon les analytes. Cela signifie notamment que les espèces les plus faiblement détectées avec la matrice DHB, le plus souvent les N-glycanes les plus massifs ($> 3\ 000$ Da), sont bien mieux distinguées du bruit en utilisant la matrice 3-AQ/p-CA. La largeur des pics a été légèrement augmentée par le changement de matrice, avec des résolutions de 13900 en DHB et de 11600 en AQ/CA pour le N-glycane A2G2 natif ou dérivé, respectivement. Cependant, cette résolution reste cohérente avec les performances attendues pour une analyse sur un appareil MALDI-TOF fonctionnant en mode réflectron. Elle est surtout suffisante pour atteindre la résolution isotopique et ainsi déterminer correctement la masse monoisotopique des espèces détectées.

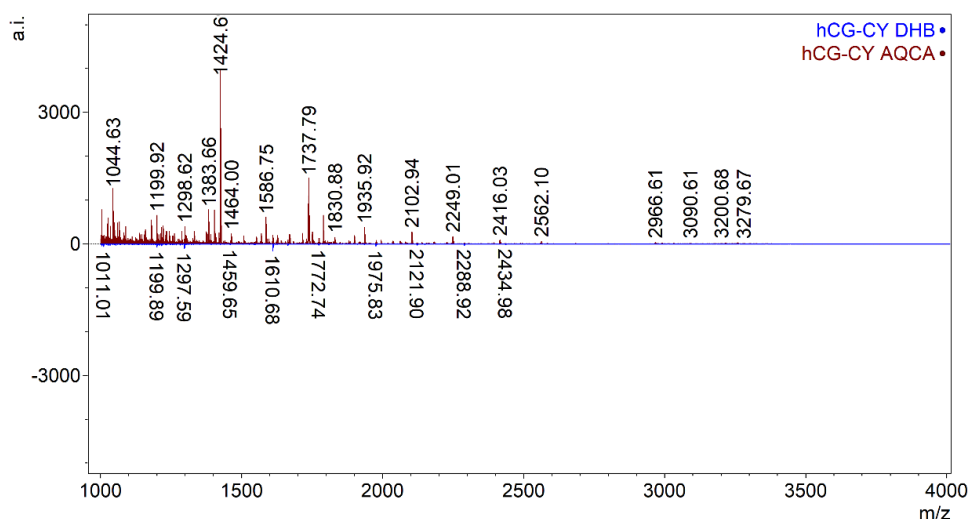
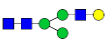

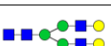





Figure 36 : Comparaison des spectres MALDI-TOF des N-glycanes de la hCG-CY obtenus avec la matrice 3-AQ/p-CA (haut) ou matrice DHB (bas).

L'intensité maximale du signal est 25 fois plus élevée avec la matrice 3-AQ/p-CA, permettant de détecter et d'identifier plus de structures glycosylées, en particulier à hautes valeurs de m/z.

Tableau 3 : N-glycanes représentatifs de la diversité des structures identifiées par MALDI-TOF sur le spectre précédent (Figure 36).

Les structures ont été identifiées par MS/MS et/ou MS et ont été validées par recherche des banques de structures polysaccharidiques. Les colonnes « Intensité » et « S/N » correspondent à la hauteur et le rapport signal/bruit du pic monoisotopique, respectivement.

N-glycane	Matrice	Intensité	S/N	m/z exp.
 A1G1	DHB	95	56,3	1297,59
	3-AQ/p-CA	3725	316,1	1424,69
 A1G1S1 (Na)	DHB	149	115,3	1610,68
	3-AQ/p-CA	1453	201,7	1737,79
 A2G2	DHB	29	22,4	1662,74
	3-AQ/p-CA	628	93,8	1789,85
 M5A1G1	DHB	12	9,0	1621,72
	3-AQ/p-CA	196	27,7	1748,82
 A3G2NAcG1S3F	DHB	<i>Non détecté</i>		
	3-AQ/p-CA	25	19,4	3257,69
 A3G2NAcG1S2F	DHB	<i>Non détecté</i>		
	3-AQ/p-CA	27	19,6	2966,61

Les glycanes détectés dans les deux conditions expérimentales permettent de montrer que la matrice AQ/CA donne des résultats au moins équivalents à la matrice DHB, et bien souvent au moins deux fois meilleurs en termes de sensibilité. Cette nouvelle méthodologie permet notamment la détection de N-glycanes triantennes, A3G2NAcG1S2F et A3G2NAcG1S3F indiqués dans la table ci-dessus (Tableau 3).

1.3 N-glycomique pour l'analyse de glycohormones urinaires

La comparaison des deux échantillons de hCG (hCG-BBT et hCG-CY) a été effectuée en utilisant le protocole développé et optimisé précédemment. La quantité de glycoprotéine a été normalisée à 20µg par échantillon en amont de la déglycosylation par 2U de PNGase F.

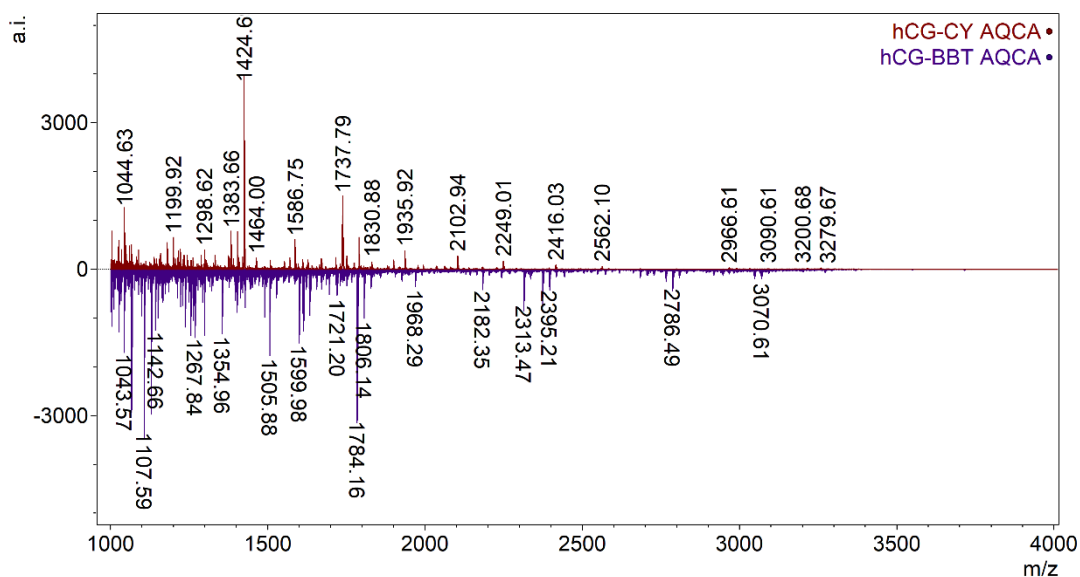


Figure 37 : Spectres MALDI-TOF MS des échantillons hCG-CY (haut) et hCG-BBT (bas), matrice 3-AQ/p-CA.

Les deux spectres ci-dessus (Figure 37) mettent en évidence des compositions très différentes des deux échantillons. Si l'échantillon hCG-CY contient des structures glycosidiques identifiables, l'échantillon hCG-BBT présente un grand nombre de signaux attribués à des espèces peptidiques ou non identifiées en MS/MS. Leur présence affectant l'efficacité d'ionisation des glycanes, ils compromettent la robustesse des approches semi-quantitatives des sucres, et montrent l'importance du processus de purification des glycanes. Plusieurs espèces osidiques sont toutefois communes aux deux échantillons à des intensités relatives variables (Figure 38).

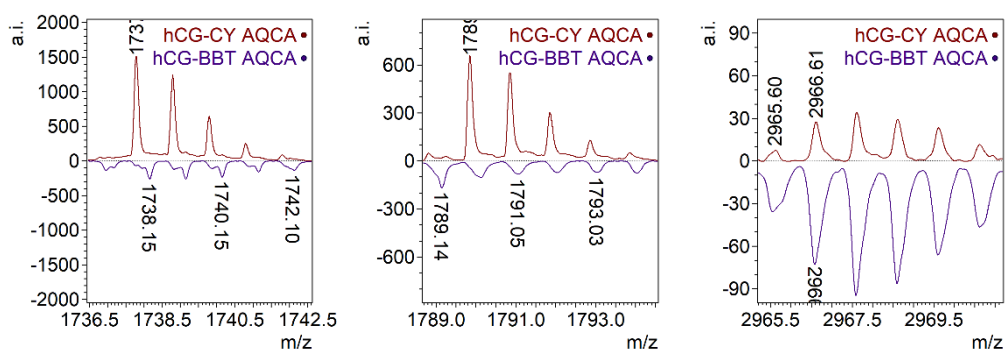


Figure 38 : Spectres MALDI-TOF MS des N-glycanes A1G1S1 (gauche), A2G2 (milieu) et A3G2NAcG1S2F (droite).

Echantillons hCG-CY (haut) et hCG-BBT (bas) analysés avec la matrice 3-AQ/p-CA.

Parmi les N-glycanes communs aux échantillons hCG-BBT et hCG-CY, on peut noter des différences d'intensité relatives entre les deux échantillons, notamment pour les N-glycanes identifiés précédemment. Les formes A1G1S1 et A2G2 produisent un signal très intense dans l'échantillon hCG-CY et sont des formes de N-glycanes relativement communes chez l'Homme, tandis que les signaux correspondants à ces glycanes ne sont que minoritaires dans l'échantillon hCG-BBT. Au contraire, le signal du glycanes A3G2NAcG1S2F est plus intense dans cet échantillon que dans l'échantillon hCG-CY. Les signaux de plusieurs N-glycanes peuvent ainsi être comparés, mais puisque les échantillons sont sensiblement différents dans leur composition (présence de peptides et autres molécules dans l'échantillon hCG-BBT), il est prudent de ne pas surinterpréter ces résultats.

1.4 Conclusions

En modifiant et optimisant la préparation de la matrice MALDI 3-AQ/*p*-CA et des échantillons, la sensibilité et la répétabilité de l'analyse MALDI-TOF a été sensiblement améliorée par rapport au protocole existant (utilisation de la matrice DHB). L'optimisation de ce protocole, d'abord effectuée sur des glycanes issus de protéines commerciales, a été transposée à l'analyse des N-glycanes de deux hCG d'origines géographiques distinctes. Pour cette application, le rapport S/N a été augmenté d'un facteur allant de 2 à 5 en fonction des N-glycanes, et a notamment permis la détection de N-glycanes à des masses plus élevées qu'avec la matrice DHB. Les méthodes développées ici ont montré leur efficacité pour l'étude de différentes glycoprotéines, et peuvent être transposées à des travaux de glycoprotéomique à plus large échelle. D'autres pistes d'amélioration sont étudiées, notamment pour réduire le temps d'incubation de la matrice et de l'échantillon (une heure à 60°C). L'utilisation d'ondes acoustiques (sonication) ou électromagnétiques (micro-ondes) permet une agitation moléculaire locale et peut augmenter sensiblement la vitesse des réactions, comme le peut une activation thermique « classique ». L'utilisation de ces méthodes directement sur la cible MALDI doit cependant être évaluée.

Une approche semi-quantitative a été également abordée mais les probables effets matrice sur l'ionisation des glycanes mettent en évidence l'importance des méthodes de purification des sucres ainsi que l'intérêt d'une séparation en amont de l'analyse, par exemple par couplage HPLC-MALDI. Il semble également nécessaire d'introduire une méthode de normalisation plus efficace que la simple normalisation de la masse initiale de glycoprotéines. En effet, les différences de *site-occupancy* (proportion des sites de glycosylation effectivement occupés par des glycanes) ou de masse des sucres introduisent un biais important en termes de quantité de glycanes analysés. La normalisation de la quantité de glycoprotéines (dosage spécifique de la portion protéique par des anticorps) permettrait une analyse plus correcte et quantitative de la population des N-glycanes.

Une possible dérive de la précision de masse est détectable entre les deux échantillons. L'épaisseur du film de matrice liquide pouvant impacter de façon non négligeable la distribution des ions dans l'espace, et donc la résolution et la précision en masse, la diminution de la quantité d'échantillon déposée pourrait permettre de réduire cette erreur. L'introduction d'un standard à l'échantillon (*spiking*), de structure et de masse connues, permettrait également de corriger d'éventuelles dérives. Cette influence de l'épaisseur du dépôt et donc de sa planéité pour l'analyse MALDI n'est sensible que dans le cas des analyseurs TOF utilisés seuls. Ceci signifie donc que la forme du dépôt n'aurait pas la même influence pour les appareils hybrides de type quadripôle-TOF par exemple.

L'étude comparée des hCG caucasienne (hCG-CY) et asiatiques (hCG-BBT) a toutefois permis de mettre en évidence la présence de différents N-glycanes. Ils sont indicateurs de différences entre les voies de biosynthèse des sucres dans ces populations, qui doivent être confirmées sur des cohortes plus importantes afin de mieux évaluer la diversité au sein des deux groupes.

L'effet biologique de la glycosylation est très variable selon les glycoprotéines. En effet, certaines structures peuvent être reconnues spécifiquement par des

lectines, tandis que l'effet global de la glycosylation est également notable dans de nombreux processus physico-chimiques ou pour adopter une conformation reconnue par certains récepteurs membranaires. L'analyse glycomique permet d'évaluer globalement la population des N-glycanes sur une famille de protéines, et peut s'avérer plus informative que des approches de protéomique *bottom-up*. L'utilisation du MALDI-TOF est également très avantageuse en termes de temps d'analyse puisque de nombreux échantillons peuvent être analysés séquentiellement à très haut débit.

Certains glycanes détectés dans cette étude (notamment A3G2NAcG1S2F et A3G2NAcG1S3F) sont notamment annotés dans les banques de données comme identifiées sur l'uromoduline, la glycoprotéine urinaire la plus abondante. Cette protéine ayant été identifiée dans les échantillons, il est possible que les glycanes ci-dessus en soient issus. L'approche glycomique ne permettant pas la localisation des glycanes identifiés, elle peut être complémentée par la glycoprotéomique *bottom-up*.

D. Troisième Partie :

*Caractérisation structurale de la
glycosylation des gonadotrophines
chorioniques par approche bottom-
up*

1. Protéolyse aspécifique et fragmentations *stepped-HCD* pour la comparaison de deux hCG d'origines géographiques distinctes

1.1 Introduction et problématique

Plusieurs travaux de recherche ont montré et tenté d'expliquer les mécanismes par lesquels la glycosylation de la hCG est impliquée dans son activité, et en particulier sa reconnaissance par certains récepteurs membranaires et leur activation (179). Certaines études ont cherché à identifier les différentes glycoformes de la hCG, afin de montrer en particulier les différences entre les hCG thérapeutiques d'origine urinaire et celle produite par expression d'ADN recombinant en culture cellulaire (183), tandis que certaines études ne se sont attardées qu'à la localisation des sites de O-glycosylation (184). Il semble donc nécessaire de disposer d'outils toujours plus performants pour caractériser précisément l'hétérogénéité structurale des différentes protéoformes de la hCG.

L'approche la plus couramment employée pour caractériser les glycoprotéines, et donc la choriogonadotropine humaine, est d'analyser les glycopeptides produits après protéolyse enzymatique, soit l'approche dite « *bottom-up* ». Comme décrit précédemment (2.3.1.2 Protéolyse enzymatique), l'enzyme la plus utilisée dans ce contexte est la trypsine puisqu'elle permet de générer de façon efficace et robuste des peptides de taille adaptée aux spectromètres de masse et porteurs d'une charge positive à leur extrémité C-terminale.

Cependant, dans le cas de la hCG, l'utilisation de la trypsine ne permet pas de résoudre les sites de O-glycosylation Ser-127 et Ser-132 (Ser-147 et Ser-152 de la hCG β immature)(Figure 39). L'utilisation de protéase avec une distribution des sites de protéolyse plus dense doit permettre de produire des glycopeptides porteurs d'un unique site de glycosylation. Ainsi, la pronase E est composée d'un mélange de protéases ayant globalement la capacité de digérer intégralement le squelette peptidique d'une protéine, à l'exception des domaines « protégés » par l'encombrement stérique de certaines modifications massives, telles les glycanes. Outre la protéolyse quasi-intégrale des régions non glycosylées de la protéine, permettant d'améliorer la détection des

glycopeptides, l'utilisation de telles protéases doit permettre d'isoler chaque site de glycosylation sur un peptide différent.

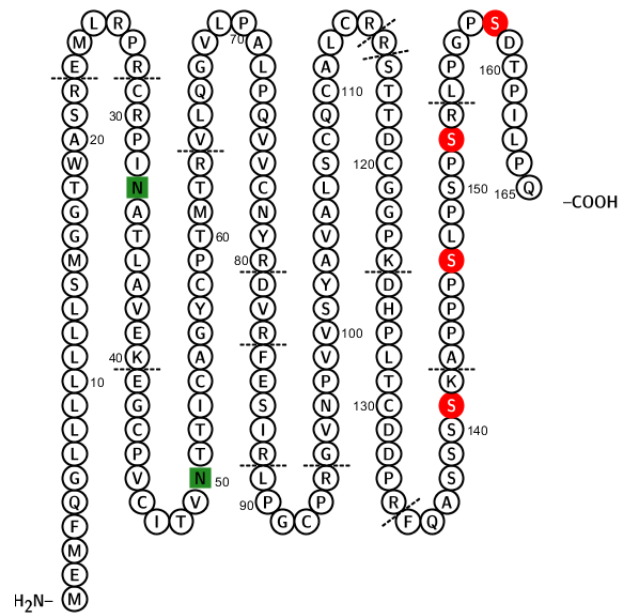


Figure 39 : Séquence en acides aminés de la hCG β (UniProt P0DN86 - CGB3_HUMAN). Les sites de N-glycosylation sont mis en évidence en vert, et les sites de O-glycosylation en rouge. Les sites d'action de la trypsine sont également indiqués par un pointillé. Le peptide signal MEMFQGLLLLLLSMGGTWA est clivé dans la forme mature de la sous-unité. Figure réalisée grâce au logiciel Protter (185).

L'étude de la forme équine de la choriogonadotropine (eCG) est encore plus complexe, puisque son domaine C-terminal est potentiellement porteur de 12 O-glycanes, et bien que tous les sites de O-glycosylation ne soient pas nécessairement occupés simultanément, un peptide tryptique peut porter jusqu'à 8 sites de O-glycosylation avec 5 sites présents sur 5 résidus successifs (sur un peptide constitué de 19 acides aminés). Avec les outils à disposition aujourd'hui il est très compliqué de chercher à identifier de tels peptides. En effet, l'analyse de ces derniers s'apparente davantage à une analyse top-down, encore souvent limitée en glycoprotéomique à l'analyse de protéines monoglycosylées ou à l'analyse sans fragmentation. Dans le cadre de l'étude de la eCG, l'intérêt de la pronase se limitera à diminuer au maximum le nombre de résidus glycosylés sur chaque peptide.

Cependant les peptides protéolytiques issus de telles digestions aspécifiques ne présentent pas nécessairement de résidu basique à leur extrémité, ou même au sein de leur séquence primaire de façon plus générale, et s'ionisent moins efficacement. Ainsi l'état de charge maximum de ces peptides est souvent réduit. Surtout, la fragmentation des peptides devient très dépendante de leur séquence, aboutissant à des fragmentations irrégulières et moins informatives que lors d'une digestion classique à la trypsine qui génère des peptides portant un résidu basique (Arginine ou Lysine) à leur extrémité C-terminale. Certaines études ont montré l'intérêt d'adapter les énergies de fragmentation pour ces peptides non-tryptiques afin de provoquer des fragmentations spécifiques (153). L'utilisation de fragmentations *stepped-energy HCD* pourrait ainsi permettre la détection d'un plus grand nombre de fragments informatifs, y compris les fragments oxoniums, les fragments peptidiques b et y, ou les fragments glycopeptidiques Y.

Plusieurs hCG ont été utilisées dans le cadre de cette étude. Deux échantillons de hCG nous ont été gracieusement fournis par le laboratoire INRA « Physiologie de la Reproduction et des Comportements », purifiées à partir d'urines de femmes soit principalement d'origine asiatique (hCG-BBT) ou principalement d'origine caucasienne (hCG-CY), dans le but de comparer leur état de glycosylation. Un échantillon de choriogonadotropine équine (eCG) purifiée a également été analysé.

Le protocole de protéolyse entrepris a eu pour but de comparer, pour chaque échantillon, l'intérêt de la digestion aspécifique utilisant la pronase avec l'approche classique (digestion tryptique). De précédentes études ont montré que la pronase ne permettait pas toujours une digestion complète de l'échantillon. Cette particularité a d'abord été attribuée à la forte imbrication des deux sous-unités des différentes CG et à l'importance de l'encombrement stérique des différents groupements polysaccharidiques. Afin de tenter d'augmenter l'efficacité de la pronase, une double digestion a été entreprise, en débutant par une digestion tryptique suivie d'une digestion à la pronase.

1.2 Protocole de glycoprotéomique *bottom-up* et identification

1.2.1 Préparation de l'échantillon

Les échantillons de protéine ont été dilués à un rapport volumique 1 : 1 dans une solution de Bicarbonate d'Ammonium (ABC) concentrée à 50mM (pH 7,8). La dénaturation des protéines a été effectuée en suivant un protocole classiquement employé au laboratoire. La réduction des ponts disulfures a été effectuée en solution par l'ajout de dithiothréitol (DTT) à une concentration finale de 10 mM (à partir d'une solution stock concentrée 10X). La solution a été incubée à 56°C pendant 35 minutes sous agitation modérée. Après retour à température ambiante, les échantillons ont été traités avec de l'iodoacétamide (IAM) à une concentration finale de 50 mM (à partir d'une solution stock 10X) afin d'alkyler les cystéines libres et bloquer leur réactivité. Cette réaction s'est déroulée à température ambiante pendant 30 minutes, à l'obscurité pour éviter la destruction du réactif IAM.

Les protéines sont digérées soit par un mélange LysC-Trypsine, ou par la pronase, ou par les deux. Dans les faits, la trypsine hydrolyse la liaison peptidique en position C-terminale des lysines, comme la LysC, le mélange des deux enzymes ne modifie pas les sites de protéolyse mais permet de diminuer le nombre de *missed-cleavages*. Une masse de 0,2 µg de LysC-Trypsine a été utilisée pour chaque digestion, avec une incubation d'environ 18 heures à 37°C sous agitation modérée. Les digestions à la pronase ont été faites dans les mêmes conditions avec une quantité d'enzyme de 10 µg. La double digestion LysC-Trypsine/Pronase a consisté à digérer les échantillons d'abord par la LysC-Trypsine puis par la pronase, dans les conditions énoncées précédemment.

1.2.2 Analyse LC-MS/MS

L'analyse des (glyco)peptides produits par les différents protocoles de protéolyse a été effectuée par une approche classique de protéomique *bottom-up* en LC-MS/MS.

Le système de chromatographie liquide utilisé est un appareil de nanoLC RSLCnano UltiMate 3000 de la société ThermoFisher Scientific. La miniaturisation permet l'utilisation de source nano-électrospray et un gain important en sensibilité par rapport à la LC-ESI conventionnelle. Une pré-colonne de type C18 a été utilisée (Acclaim PepMap100, diamètre interne 300 μm \times 5 mm, taille de particule 5 μm , ThermoFisher Scientific). La colonne analytique utilisée est une colonne capillaire de phase C18 (Acclaim PepMap100 C18, diamètre interne 75 μm \times 50 cm, taille de particule 3 μm , porosité 100 Å, ThermoFisher Scientific). Les différentes phases mobiles sont les suivantes : phase A (98% H₂O, 2% ACN, 0.05% FA), phase A'' (98% H₂O, 2% ACN, 0.1% FA), phase B (90% ACN, 10% H₂O, 0.1% FA). Un montage pré-colonne/colonne a été utilisé, permettant la préconcentration des échantillons dans un volume inférieur au volume total de la colonne analytique (2,2 μL environ). Les échantillons ont été injectés en triplicats dans un volume de 2 μL , dans une boucle de 20 μL . Ils sont ensuite chargés sur la pré-colonne C18 dans une phase mobile composée à 100% de A'', à un débit de 15 $\mu\text{L}/\text{min}$. La rotation de la valve portant la pré-colonne, 5 minutes après l'injection, permet la mise de cette dernière « en ligne » avec la colonne analytique (dans le même flux de phase mobile). Initialement composée d'un mélange de phase A (99%) et de phase B (1%), la composition de la phase mobile augmente de façon linéaire à 40% de phase B en 60 minutes. La proportion de phase B est ensuite rapidement portée à 90% et maintenue pendant 15 minutes afin d'éluer les molécules fortement retenues (*flush*), puis la composition de la phase mobile est ramenée à son état initial (1%B) et équilibrée pendant une durée de 30 minutes. Ce gradient permet une séparation dans le temps des peptides en fonction de leur hydrophobie, principalement. La séparation a été faite à un débit constant de 300 nL/min.

Le système de chromatographie liquide a été couplé avec un appareil de type QqOrbitrap (Q Exactive HF, ThermoFisher Scientific), à l'aide d'une source nano-ESI. Le démarrage de l'acquisition coïncide avec la mise en ligne de la pré-colonne et de la colonne analytique ($T_{\text{injection}} + 5$ minutes), pour une durée de

60 minutes. Le mode de fonctionnement utilisé dans cette étude est couramment dénommé Top10 : un scan MS (résolution 60k, *automatic gain control (AGC) target 3e6*, temps d'injection maximum 60 ms, gamme de masse 375-1500 m/z) suivi de jusqu'à 20 scans MS/MS en mode DDA (top 20, exclusion dynamique des précurseurs fragmentés pendant 20 s, exclusion systématique des ions à $z = 1$ et $z > 8$, résolution 15k, *AGC target 1e5*, temps d'injection maximum 60 ms).

L'énergie de fragmentation NCE (*Normalized Collisional Energy*) a été fixée à une valeur de 28 pour l'approche classique. Dans le cadre des fragmentations *stepped-HCD*, les énergies ont été fixées à 10, 18 et 28, après optimisation. L'utilisation de HCD à basse énergie permet notamment de réduire la migration des monosaccharides en phase gazeuse et d'obtenir des fragments glycopeptidiques plus informatifs, tout en conservant l'information apportée par les ions oxoniums à bas m/z.

1.2.3 Identification des glycoprotéines

Les fichiers bruts obtenus par spectrométrie de masse (chromatogrammes et spectres) ont été traités avec le logiciel d'identification dédié à la protéomique Proteome Discoverer 2.2 (PD 2.2) du fabricant ThermoScientific, et le logiciel PMi-Byonic v3.3.3 de ProteinMetrics, dédié à l'analyse des PTMs et notamment des glycanes.

Les spectres MS/MS ont été filtrés (*node PD 2.2 Spectrum Selector*) avant l'interrogation, notamment pour éliminer les spectres issus de la fragmentation de précurseurs de plus de 10 kDa ou les spectres MS/MS dépourvus de pics à un niveau de S/N suffisant. La banque de séquences protéiques utilisée contient les séquences, annotées et validées expérimentalement dans la base SwissProt, de la hCG (une séquence pour la sous-unité α et 4 séquences pour la sous-unité β). Les sites de coupures recherchés sont réglés sur « non-spécifique » dans le cas des échantillons digérés à la pronase (ou LysC-Trypsine/Pronase), et sur « *full-trypsin* » (sites d'hydrolyse trypsique aux deux extrémités du peptide)

avec une tolérance pour un maximum de deux sites tryptiques manqués, afin de tenir compte de la faible accessibilité de certains sites. Une tolérance de 10 ppm d'erreur sur la mesure de m/z du précurseur et de 20 ppm d'erreur sur les m/z des ions fragments est permise. Les modifications communes des peptides sont recherchées : Carbamidométhylation des cystéines (réglage de fréquence de la modification, *common2*) ; Oxydation des méthionines (*common2*) ; Déamidation des asparagines ou des glutamines (*common1*). Les banques de données de glycanes les plus complètes ont été utilisées pour tenir compte de toutes les glycoformes possibles et pour optimiser le processus de validation : banque de compositions de N-glycanes contenant 309 structures (*rare1*), ou de O-glycanes (contenant 78 structures, *rare1*). Le nombre maximum de modifications autorisées sur chaque peptide est fixé à une modification rare et deux modifications communes.

1.3 Résultats

1.3.1 Traitement des résultats d'interrogation

Les résultats issus de l'identification des glycopeptides par le logiciel Pmi-Byonic ont été filtrés sur la base de données statistiques, spectrales, et structurales.

Le Score d'identification est une donnée calculée sur la base du spectre et des fragments identifiés dans ce dernier, selon un système de bonus/malus donnant plus ou moins de poids aux ions selon qu'ils soient significatifs d'une espèce glycosylée ou au contraire redondants. La valeur statistique PEP (*Posterior Error Probability*) quant à elle est globalement équivalente à une *p-value*, représentative de la probabilité que l'identification soit erronée, et calculée à partir de plusieurs données statistiques (Score, Delta de m/z du précurseur, *etc.*).

Cependant, l'utilisation de protéolyse non spécifique pour la O-glycoprotéomique augmente considérablement le nombre potentiel de produits de digestion. La digestion *in silico* et l'interrogation des banques de

données requièrent donc une puissance de calcul important et demandent un temps de recherche important. L'analyse d'un nombre important d'échantillons conduisant à des temps totaux de recherche irréalistes (pouvant excéder plusieurs jours par échantillon), il a été décidé de cibler les banques de séquences protéiques au maximum et de ne conserver que les séquences des protéines possibles. Par conséquent, l'utilisation d'une banque randomisée (*decoy*) pour le calculer du taux de faux-positifs (*False Discovery Rate*, FDR) n'est pas possible. Ceci conduit également à ce que l'utilisation exclusive de la PEP pour filtrer les PSMs (*Peptide Sequence Match*) incorrects conduise à des identifications aberrantes. D'autre part, le fait de baser le filtre uniquement sur la valeur de Score entraîne également un nombre élevé d'identifications de qualité médiocre, du fait du poids important de certains ions non discriminants dans le système de scoring par bonus/malus. Par conséquent, nous avons défini une stratégie permettant une meilleure automatisation de la validation des données. Les espèces identifiées avec une valeur de Score inférieure à 100 ou une PEP supérieure à 0,05 (5%) ont été supprimées. Les peptides non glycosylés ont également été supprimés.

La combinaison de différents monosaccharides conduit parfois à la production d'espèces isobares, compliquant la validation des identifications. L'analyse manuelle des spectres a montré que les espèces identifiées comme contenant plus d'un fucose sont en réalité des faux-positifs, pour la plupart résultant de la confusion entre un NeuAc ($M = 291,0954$ Da) et deux fucoses ($M = 292,1158$ Da, i.e. $146,0579$ Da par fucose), combinée à la mauvaise attribution de la masse monoisotopique du précurseur (Figure 40). L'erreur théorique de cette confusion est de 42 ppm (10,2 mmu), mais la présence d'ions diagnostics permet souvent de trancher en faveur des glycopeptides non-fucosylés ou mono-fucosylés. Par précaution, toutes les identifications de glycopeptides porteurs de plus d'un fucose ont été purgées en première approche.

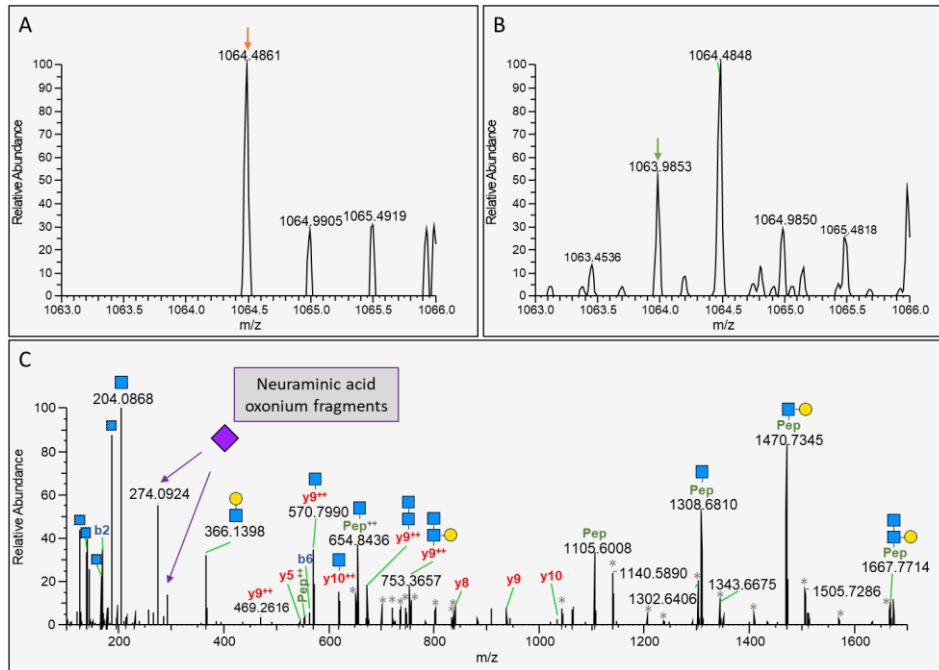


Figure 40 : Spectres illustrant la mauvaise identification de peptides multi-fucosylés. (A) Spectre MS FullScan précédant le spectre MS2 incorrectement identifié. (B) Spectre MS moyenne de 0,2 min avant à 0,2 min après le spectre A. La flèche orange indique le précurseur incorrectement identifié, tandis que la flèche verte indique le précurseur correct sur le spectre moyenné. (C) Spectre MS/MS séquencé comme APPPSLPSPSR – HexNAc(2)Hex(2)Fuc(2) montrant des fragments oxoniums caractéristiques du NeuAc aux m/z 274,0924 et 292,1034.

De façon similaire, l'analyse manuelle des spectres des espèces identifiées comme porteuse d'un résidu NeuGc (masse du résidu $M = 307,0903$ Da) ont montré qu'il s'agissait plus probablement de la combinaison d'un résidu NeuAc ($M = 291,0954$ Da) et d'une oxydation ($M = 15,99491$ Da). Des études ont montré l'absence du gène permettant la production de NeuGc chez l'Homme (186), rendant la présence de ce monosaccharide anecdotique dans les glycoprotéines humaines et due à l'intégration de monosaccharides issus de l'alimentation (187). Les identifications de glycopeptides contenant des résidus NeuGc ont également été supprimées.

Les glycopeptides ont ensuite été triés en fonction de leur séquence et du/des sites de glycosylation qu'ils couvrent (sites de N-glycosylation $\alpha 1$, $\alpha 2$, $\beta 1$, ou $\beta 2$, et sites de O-glycosylation O1, O2, O3, et/ou O4).

Pour les échantillons d'hCG digérés par la pronase, le processus de validation a éliminé en moyenne 90,4 % des paires « glycosite – glycane » identifiées et

pouvant apporter un doute quant à leur véracité. La majorité de ces identifications ont été éliminées sur la base de leur Score ou de leur PEP, le « filtre multi-Fuc » éliminant 24,6% des paires identifiées en moyenne, et le « filtre NeuGc » en éliminant en moyenne 5,7%.

1.3.2 Comparaison des méthodes de protéolyse

Les différentes stratégies de protéolyse ont été évaluées et comparées sur la base du nombre de glycopeptides distincts identifiés pour chaque site de glycosylation (après filtres, voir partie précédente).

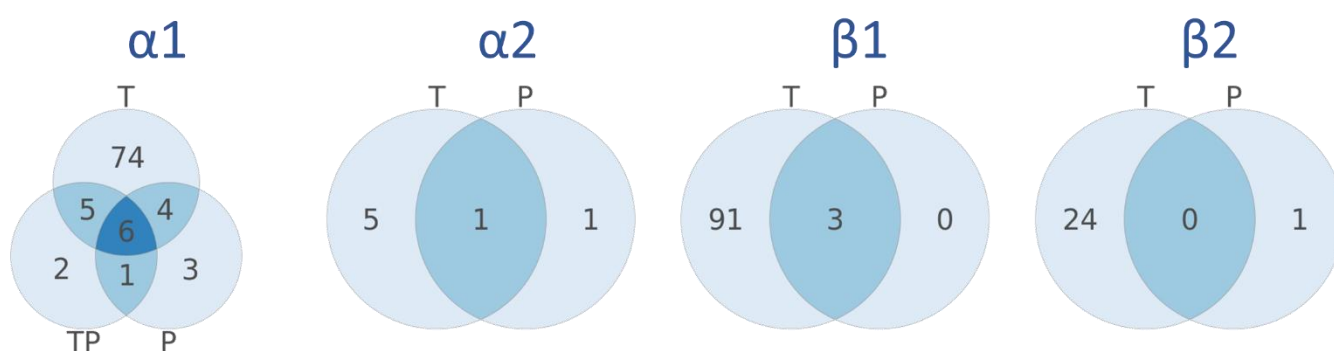


Figure 41 : Comparaison des glycopeptides identifiés pour chaque site de N-glycosylation avec différentes protéases, pour l'échantillon hCG-CY avec une NCE fixe.
T : LysC/Trypsine. P : Pronase. TP : LysC/Trypsin-Pronase.

Pour la caractérisation de l'hétérogénéité des N-glycanes de la hCG, la trypsine est sans conteste l'enzyme de choix. L'utilisation de la pronase, avec ou sans prédigestion tryptique, ne permet qu'un nombre marginal d'identification différentes, au prix d'une baisse significative du nombre total d'identifications (Figure 41).

Cette particularité pourrait s'expliquer par la bonne répartition des sites de N-glycosylation sur la protéine et leur relatif bon accès pour les protéases. Il est possible que la pronase produise des peptides de très faible longueur (pouvant se limiter à un ou deux acides aminés comme reporté dans la littérature), et que ceux-ci ne soient donc pas retenus sur la phase C18. De tels peptides seraient ainsi éliminés lors de l'étape de préconcentration. Cette hypothèse est soutenue par l'absence complète de peptides identifiés pour les sites α2, β1 et β2 dans les échantillons doublement digérés. En effet, la double digestion LysC/Trypsin-

Pronase doit produire des peptides de taille inférieure à la pronase seule, mais pour ces trois sites de N-glycosylation la totalité des glycopeptides identifiés est perdue. Les glycopeptides identifiés sont également tous glycosylés par des formes de masses importantes et majoritairement sialylées, plus susceptibles d'être retenus en RPLC.

Alternativement, la faible spécificité de la pronase pouvant conduire à des peptides de tailles variables, il est également possible que le grand nombre de peptides différents ainsi produits et leur faible efficacité d'ionisation soient responsable du faible nombre d'identification passant les différents filtres.

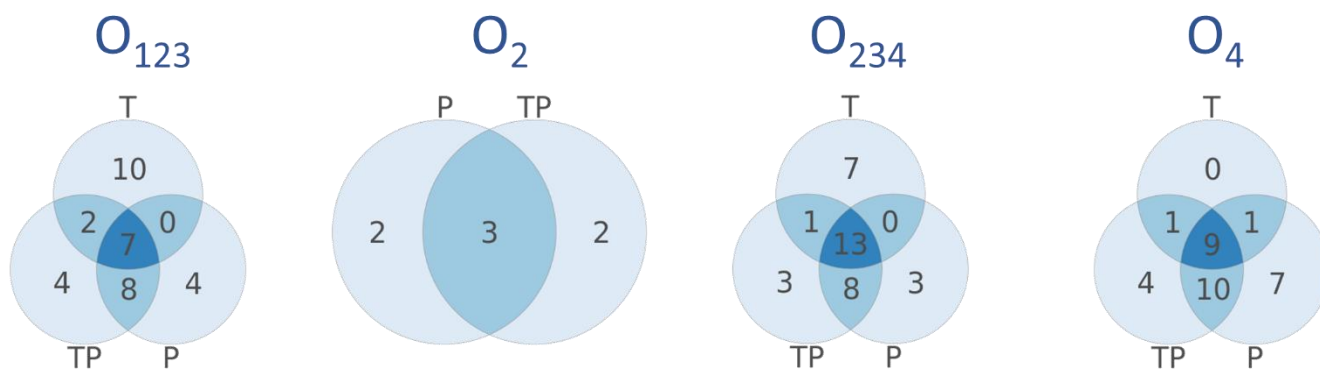


Figure 42 : Comparaison des glycopeptides identifiés pour chaque site de O-glycosylation avec différentes protéases, pour l'échantillon hCG-CY avec une NCE fixe.

T : LysC/Trypsine. P : Pronase. TP : LysC/Trypsine-Pronase. Les sites O1 et O3 n'ayant pas été isolés dans ces expériences, les peptides comprenant entre autres ces sites ont été représentés.

Contrairement aux N-glycopeptides et comme attendu, les sites de O-glycosylation sont mal caractérisés par la trypsine, mais de façon plus surprenante ne le sont pas non plus par la pronase ou la digestion trypsine-pronase (Figure 42). En détails, la trypsine seule ne permet l'identification d'aucun glycopeptide porteur d'un unique site de O-glycosylation de façon exclusive (les O-glycopeptides O2 et O4 identifiés le sont également avec la pronase).

Il est probable que les peptides porteurs d'un unique site de O-glycosylation, s'ils sont produits, soient faiblement ionisés ou mal fragmentés, d'où l'introduction de la fragmentation *stepped-HCD*.

1.3.3 Fragmentation *stepped-HCD*

Le mode de fragmentation *stepped-HCD* consiste à utiliser plusieurs énergies de fragmentation afin d'augmenter la couverture des fragments produits dans la cellule HCD. Tous les fragments sont injectés simultanément dans l'analyseur (Orbitrap) et un seul spectre MS/MS composite est généré (Figure 43).

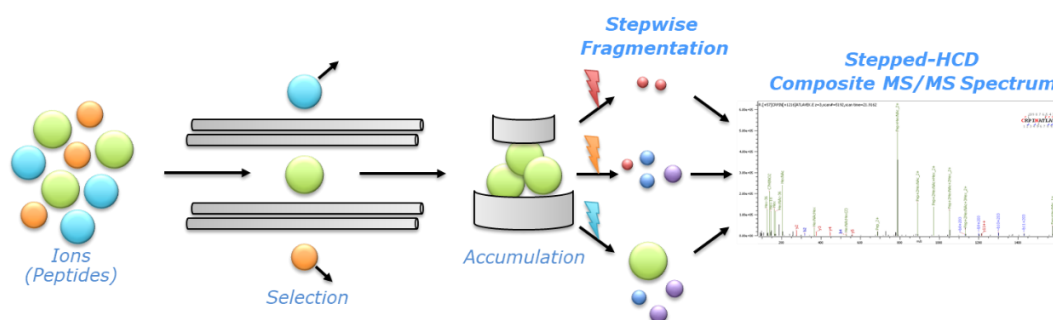


Figure 43 : Principe de la fragmentation *stepped-HCD*.

Comme dans un protocole classique, un m/z (donc dans l'idéal un précurseur) est isolé par le quadrupôle et accumulé dans la C-trap. Les précurseurs sont ensuite séparés en 2 ou 3 fractions et fragmentés à différents niveaux d'énergie. L'ensemble des fragments produits est injecté dans l'Orbitrap et conduit à la production d'un unique spectre MS/MS.

1.3.3.1 Isolation des sites de O-glycosylation

Le but de ce mode de fragmentation est d'identifier de nouveaux O-glycopeptides, en particulier sur les sites O1 et O3, non isolés jusqu'à présent.

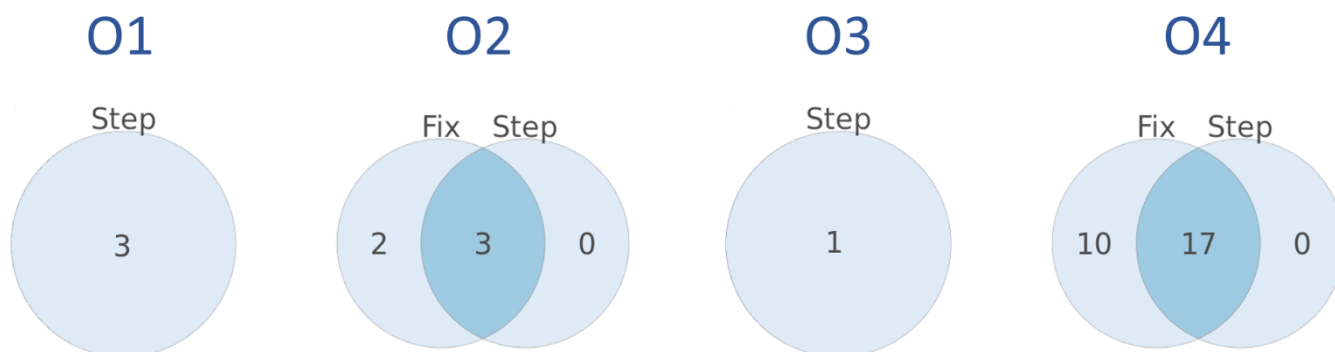


Figure 44 : Comparaison des fragmentations à énergie unique ou *stepped-HCD* pour l'identification des O-glycopeptides produits par une digestion par la pronase.

L'échantillon hCG-CY est utilisé pour cette étude. Le nombre de structure identifiées est indiqué pour chaque glycosite, avec la fragmentation *stepped-HCD* (Step) ou classique (Fix).

Pour la première fois dans ce projet, des glycopeptides porteurs uniquement des sites de glycosylation O1 ou O3 ont été identifiés avec une qualité suffisante

pour passer les filtres appliqués aux données (Figure 44). La digestion par la pronase associée à la fragmentation *stepped-HCD* a permis d'identifier sur le site O1 les glycanes HexNAc(2)Hex(2)NeuAc(1), HexNAc(1)Hex(1)NeuAc(1), et HexNAc(1)Hex(1)NeuAc(2), ainsi que le glycanes HexNAc(1)Hex(1)NeuAc(1) sur le site O3 (Figure 46).

D'autres glycopeptides ont été identifiés pour ces sites mais les données statistiques ne permettent pas leur validation. Afin d'augmenter la qualité des spectres, l'augmentation de la concentration en espèces glycosylées peut être envisagée, par utilisation de méthodes de purification préalables à la LC. Cependant, les glycopeptides représentent 11264 spectres sur 11897 spectres MS/MS identifiés dans l'échantillon hCG-CY digéré par la pronase et fragmenté à énergie constante, soit 94,7% des PSMs sur les trois réplicats. La très forte proportion de glycopeptides produits par une digestion par la pronase laisse donc peu de chances d'augmenter encore la proportion de telles espèces après digestion, mais peut toutefois permettre leur concentration dans un volume plus faible.

S'agissant parfois de glycopeptides de taille relativement faible et donc d'hydrophobie limitée, l'utilisation de méthodes de rétention plus adaptées aux composés hydrophiles (par exemple HILIC ou PGC) doit être envisagée.

Enfin, une nouvelle méthode d'identification des glycopeptides est présentée dans ce chapitre (Partie 2, Analyse approfondie des données spectrales pour l'identification de glycoformes minoritaires) et a montré des résultats intéressants pour l'identification de N-glycopeptides peu abondants. Cette approche pourrait s'avérer utile pour l'identification des O-glycopeptides peu abondants ou faiblement ionisés.

1.3.3.2 Réduction du taux de faux-positifs

L'analyse des O-glycopeptides identifiés sur les sites O2 et O4 semble montrer que la fragmentation classique est plus efficace car elle permettrait d'identifier un plus grand nombre d'espèces. Cependant, cette observation a mis en

évidence l'imperfection des algorithmes d'identification des O-glycopeptides mais également notre processus de validation dans ce contexte.

Tableau 4 : Structures O-glycanes identifiées pour le site O4 de la hCG-CY lors d'une digestion par la pronase, comparaison des méthodes de fragmentation à énergie constante ou stepped-HCD. Les structures identifiées uniquement dans le cas de fragmentation NCE classique sont mis en évidence.

hCG-CY – Site O4 – Pronase	NCE constante	Stepped-HCD
HexNAc(1)	+	+
HexNAc(1)Hex(1)	+	+
HexNAc(1)Hex(1)Fuc(1)	+	+
HexNAc(1)Hex(1)NeuAc(1)	+	+
HexNAc(1)Hex(1)NeuAc(2)	+	+
HexNAc(1)Hex(1)NeuAc(3)	+	-
HexNAc(1)Hex(2)Fuc(1)	+	+
HexNAc(2)Hex(1)	+	+
HexNAc(2)Hex(1)NeuAc(1)	+	+
HexNAc(2)Hex(1)NeuAc(2)	+	-
HexNAc(2)Hex(2)	+	+
HexNAc(2)Hex(2)Fuc(1)NeuAc(1)	+	-
HexNAc(2)Hex(2)Fuc(1)NeuAc(2)	+	+
HexNAc(2)Hex(2)NeuAc(1)	+	+
HexNAc(2)Hex(2)NeuAc(2)	+	+
HexNAc(2)Hex(3)	+	-
HexNAc(2)Hex(3)Fuc(1)	+	+
HexNAc(3)Hex(1)NeuAc(1)	+	-
HexNAc(3)Hex(2)Fuc(1)NeuAc(1)	+	-
HexNAc(3)Hex(3)	+	+
HexNAc(3)Hex(3)Fuc(1)NeuAc(2)	+	-
HexNAc(3)Hex(3)NeuAc(1)	+	+
HexNAc(3)Hex(3)NeuAc(2)	+	+
HexNAc(5)Hex(3)	+	-
HexNAc(5)Hex(3)Fuc(1)NeuAc(1)	+	+
HexNAc(6)Hex(3)Fuc(1)	+	-
HexNAc(6)Hex(4)	+	-

En effet, au premier regard on dénombre pour le site O4 de cet échantillon (hCG-CY) un total de 27 O-glycopeptides identifiés lors d'une digestion par la pronase, dont seulement 17 avec la fragmentation *stepped-HCD*, tandis que la fragmentation classique aboutit à l'identification de la totalité de ces espèces (Tableau 4). Ici les glycanes ne sont représentés que par leur composition car, contrairement aux N-glycanes qui suivent une biosynthèse bien connue, la formation des O-glycanes n'est pas aussi bien caractérisée et leur structure doit être déterminée uniquement sur la base des données spectrales, qui sont parfois insuffisamment informatives.

L'analyse manuelle des spectres exclusivement annotés dans des conditions de fragmentation classiques montre cependant un nombre important d'incohérences. Ces identifications, souvent issues d'un unique PSM, présentent notamment des propriétés spectrales faisant planer le doute sur leur robustesse, comme illustré en Figure 45. D'abord, la présence récurrente de fragments oxoniums liés aux NeuAc dans des glycopeptides identifiés comme n'en contenant pas montre qu'il est nécessaire de donner plus de poids à la présence ou l'absence anormale de ces fragments diagnostic. Il est cependant plus compliqué d'automatiser cette observation, notamment du fait que la composition du glycanes ne présente rien d'anormal ici, contrairement aux espèces multi-fucosylées déjà éliminées. De plus, l'absence de signaux associés aux fragments Y (peptide portant des fragments du glycanes) ne permet pas de confirmer l'identification comme c'est généralement le cas pour les glycopeptides correctement annotés. Ici, on observe notamment deux fragments à 1163,63 m/z et 1366,72 m/z, correspondant très probablement à la perte de la HexNAc directement liée au peptide ($\Delta m = 203,09$ Da). Le signal du peptide déglycosylé serait donc à une valeur de m/z de 1163,63, et non pas de 1292 m/z comme ici calculé ($[M+H]^+_{\text{précurseur}} = 2184,0$ Da et $[M]_{\text{glycanes}} = 892$ Da).

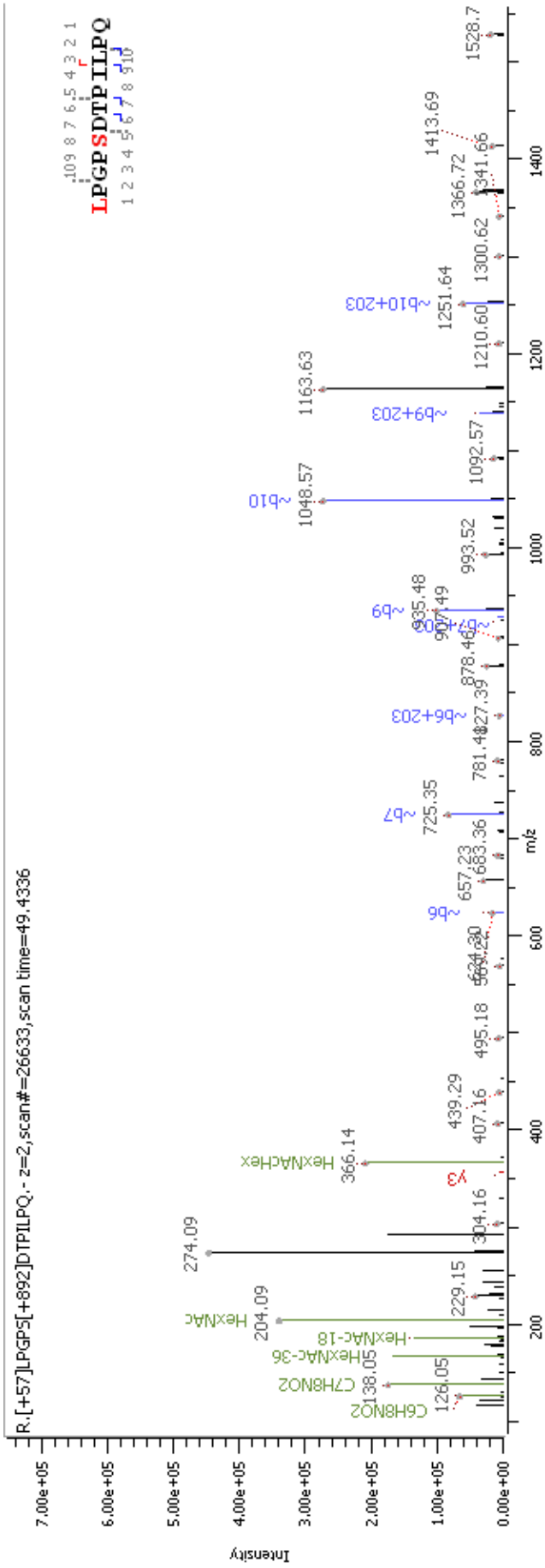
Ces erreurs étant retrouvées de façon récurrente dans les identifications exclusives à la méthode de fragmentation à NCE constante, mais non détectées lors de fragmentation *stepped-HCD*, il semble que la couverture des fragments

produits par cette dernière permette de s'affranchir de ces identifications erronées. Ainsi, cette approche est retenue puisqu'elle permet, sans vérification manuelle des spectres, de s'affranchir de ces faux-positifs.

Bien qu'elle ne soit pas indispensable, la fragmentation *stepped-HCD* a été conservée pour l'identification des N-glycopeptides tryptiques, mais les résultats des deux méthodes de fragmentation sont conservés puisqu'elles permettent chacune un nombre non négligeable d'identification exclusives.

Figure 45 : Spectre MS/MS du glycopeptide LPGPSDTPILPQ porteur du O-glycane HexNAc(2)Hex(3).

(page suivante) La présence d'ions oxonioms aux m/z 274,09 et 292,10, issus du NeuAc, invalide cette identification. De plus, la relative faible abondance des ions glycopeptidiques Y ne permet pas de confirmer l'identification proposée.



1.4 Glycosylation de la hCG et origine géographique

Les développements méthodologiques précédemment introduits (digestion aspécifique, fragmentation *stepped-HCD*, traitement de données) sont utilisées pour identifier avec le maximum de confiance les structures des glycanes présents sur les 8 sites de glycosylation de deux hCG urinaires différentes de par l'origine géographique dont sont issues les femmes donneuses. Ainsi, comme expliqué précédemment (paragraphe 1.3 de ce chapitre), les N-glycopeptides ont été identifiés en utilisant le protocole *bottom-up* avec une digestion tryptique et une fragmentation à NCE constante ou en *stepped-HCD*. Les O-glycopeptides quant à eux sont issus des digestions par la pronase et la fragmentation *stepped-HCD* uniquement. Seuls les peptides ayant passé les différents filtres (détaillés en 1.3.1) sont conservés. Seules les compositions des glycanes identifiés sont fournies par PMi-Byonic, les structures SNFG indiquées dans la table ci-dessous (Figure 46) sont déterminées comme les plus probables au vu des informations spectrales, des structures annotées dans les banques de structures glycosidiques, et des connaissances sur la biosynthèse des glycanes dont nous disposons. Elles pourraient néanmoins être erronées, notamment dans l'enchaînement des monosaccharides ou leurs ramifications, et pourraient être confirmées par d'autres techniques de fragmentation, l'introduction de spectrométrie de mobilité ionique, ou encore par des approches de biologie moléculaire.

Figure 46 : Comparaison des glycopeptides identifiés pour chaque site de N- et de O-glycosylation entre deux échantillons d'origines géographiques distinctes.

(page suivante) Structures SNFG des glycanes identifiés pour chaque site de glycosylation, obtenues par approche bottom-up (trypsine ou pronase, NCE constante ou stepped-HCD).

hCG-BBT	hCG-CY	Both	
			O1
			O2
			O3
		14 O-glycans	O4
		70 N-glycans	A1
			A2
		61 N-glycans	B1
			B2

Cette table met d'abord en évidence une certaine homologie de structures glycosidiques entre les deux échantillons, puisque 159 structures sont communes entre les deux échantillons sur tous les sites, tandis que 30 ou 95 formes sont uniquement présentes dans l'échantillon hCG-BBT ou hCG-CY, respectivement.

Cependant, on peut également remarquer un nombre non négligeable de structures identifiées uniquement dans l'un ou l'autre des échantillons. Certaines caractéristiques des glycanes exclusivement identifiés dans l'échantillon hCG-BBT ou l'échantillon hCG-CY sont regroupées dans la table suivante (Tableau 5).

Tableau 5 : Comparaison des caractéristiques des glycanes identifiés uniquement dans un des deux échantillons.

Les proportions sont exprimées relativement au total des glycanes identifiés uniquement dans l'échantillon considéré. Les caractéristiques mises en évidences sont discutées ci-après.

	hCG-BBT	hCG-CY
Formes exclusives	30	95
Fucoses	12	42
	40.00%	44.21%
Mono-sialylations	4	23
	13.33%	24.21%
Di-sialylations	5	10
	16.67%	10.53%
Tri-sialylations	2	7
	6.67%	7.37%
>3 NeuAc	0	1
	0.00%	1.05%
High-mannoses	3	7
	10.00%	7.37%
Hybrides	3	6
	10.00%	6.32%
Complexes biantennaires	5	21
	16.67%	22.11%
Complexes triantennaires	2	24
	6.67%	25.26%
Complexes >3 antennes	7	21
	23.33%	22.11%
Antennes di-galactosylées	6	5
	20.00%	5.26%

La fucosylation des glycanes ne semble pas survenir préférentiellement dans l'un ou l'autre des échantillons, avec 40% et 44% des glycanes exclusifs étant fucosylés dans l'échantillon BBT et CY, respectivement. Si on observe un léger défaut de mono-sialylation dans l'échantillon BBT (13% vs. 24%), on observe également plus de formes di-sialylées dans cet échantillon que dans l'échantillon CY (17% vs. 11%). Globalement, la sialylation des glycanes de la hCG a l'air de relativement peu varier entre ces deux populations.

Une des différences les plus notables est la proportion de N-glycanes complexes triantennes, quatre fois plus grande dans l'échantillon hCG-CY que dans l'échantillon hCG-BBT. On retrouve dans ce dernier une plus forte proportion d'antennes multi-galactosylées, en particulier les N-glycanes complexes A2G4 et FA4G8, identifiés respectivement sur les sites $\beta 2$ et $\alpha 1$ (20% vs. 5%).

1.5 Conclusions

Les développements méthodologiques introduits dans cette partie ont montré un intérêt certain pour la résolution des sites de O-glycosylation de la hCG, mais également pour la meilleure qualité des identifications. Néanmoins, il s'est également avéré qu'un travail de validation manuel était requis afin d'éviter les erreurs d'identification, bien que la combinaison de la digestion par la pronase et la fragmentation *stepped-HCD* ait permis, dans cette étude, d'augmenter la confiance dans les identifications glycopeptidiques.

L'efficacité de la détection de tels glycopeptides est cependant également limitée par la capacité de rétention de ces analytes en chromatographie liquide. L'introduction d'autres méthodes de séparation, notamment la LC-HILIC, la LC-PGC, ou la LC multidimensionnelle (HILIC-C18, C18-PGC, *etc.*) pourrait permettre d'augmenter le nombre de petits glycopeptides détectables.

Des différences structurales ont été mises en évidence entre les glycopeptides issus de la hCG d'origine caucasienne (CY) et la hCG d'origine asiatique (BBT), mais ces résultats pourraient être approfondis à différents niveaux. D'abord,

une approche quantitative pourrait permettre d'affiner la comparaison à plus de finesse qu'une simple dichotomie « absence / présence ». La quantification des glycoformes permettrait également d'éviter l'effet « seuil » des techniques de filtration, pouvant aboutir à des conclusions erronées lors de l'identification d'espèces passant de justesse certains critères de sélection. Une façon relativement simple à mettre en place serait le marquage isobare TMT.

L'approche top-down a seulement été effleurée au cours de ce projet, et principalement pour la sous-unité α puisqu'elle n'est modifiée que par un maximum de deux glycanes. La complexité des combinaisons possibles des glycanes de la sous-unité β et leur superposition avec un nombre important d'états d'oxydation de la glycoprotéine rendent ce travail extrêmement complexe, mais nécessaire. En effet, un des rôles des glycanes de la hCG est de former une enveloppe hydrophile et chargée, modifiant ses propriétés physico-chimiques et sa reconnaissance par les récepteurs membranaires. L'effet combiné des glycanes est donc au moins aussi important que leurs structures individuelles.

Il faut également insister sur le fait que les échantillons analysés sont issus d'urines de plusieurs femmes regroupées en un seul « *pool* ». Cela permet une normalisation des glycoprotéoformes présentes par effet de moyenne. Cependant, l'analyse de cohortes d'échantillons individuels pourrait apporter des réponses quant à l'impact sur la glycosylation de la hCG de certaines conditions physiologiques, comme l'âge de la donneuse, ou l'évolution de la glycosylation au cours de la grossesse.

Enfin, la significativité biologique de ces structures pourrait être étudiée par des tests de biologie moléculaire afin de déterminer si l'activité des hCG ainsi glycosylées est altérée. Ces conclusions pourraient être une piste pour expliquer les différences de réponse aux traitements de la fécondité chez la femme.

2. Analyse approfondie des données spectrales pour l'identification de glycoformes minoritaires

2.1 Introduction et problématique

Le déroulement général d'une expérience de glycoprotéomique *bottom-up* peut se résumer en trois étapes. La préparation consiste à produire des glycopeptides à partir de l'échantillon initial par protéolyse, et peut comprendre purifications, préfractionnements, enrichissements et séparations chromatographiques ou électrophorétiques, au niveau des protéines ou des peptides. L'analyse par spectrométrie de masse en tandem permet de mesurer la masse des ions formés et, après leur fragmentation (MS/MS), de mesurer la masse de leurs fragments. Les informations ainsi collectées permettent de séquencer et d'identifier les peptides et leurs glycanes, plusieurs approches différentes mais interdépendantes sont possibles, dont le séquençage peptidique *de novo*, la comparaison à des banques de séquences protéiques, ou à des banques de spectres.

Ces différentes méthodes se basent sur la masse des fragments peptidiques et glycopeptidiques pour identifier le précurseur fragmenté. Cependant l'abondance desdits fragments est très dépendante de la nature du glycopeptide et de sa concentration initiale. Si la majorité des protéines sont glycosylées, la forte hétérogénéité de ces modifications conduit à des concentrations parfois très faibles pour une glycoforme donnée. De plus, et comme explicité précédemment (*cf.* 2.3 Glycoprotéomique *bottom-up*), la faible efficacité d'ionisation des glycopeptides diminue d'autant plus l'abondance des fragments d'intérêt. Une solution envisagée au laboratoire afin de confirmer ou de proposer une identification précise des spectres MS/MS faiblement peuplés est d'utiliser les proportions relatives des ions oxoniums (fragments de sucres). En effet, chaque monosaccharide produit des fragments spécifiques (188), en particulier : les fragments mono-déshydratés ($\Delta\text{H}_2\text{O}$, -18 Da), les fragments di-déshydratés ($\Delta 2\text{H}_2\text{O}$, -36 Da), et les fragments résultants de la perte de deux molécules d'eau et d'un formaldéhyde ($\Delta 2\text{H}_2\text{O}$ -

CH₂O, -66 Da), dont il est possible d'extraire des informations structurales sur les glycanes. Cette possibilité a récemment été confirmée par les travaux menés par Marcus Hoffmann *et al.* (153) sur différents standards commerciaux (immunoglobuline G, lactotransferrine et fibrinogène humains, ribonucléase B bovine). La différenciation entre les O-glycopeptides et des N-glycopeptides sur la base des ions oxoniums produits en MS/MS a également été démontrée (189). Des travaux ont également permis d'observer une différence d'abondance des ions oxoniums produits par les glycopeptides à O-GalNAc et ceux à O-GlcNAc (190).

Les travaux développés dans ce chapitre ont pour objectif de vérifier certaines propriétés spectrales proposées par M. Hoffmann *et al.* ainsi que d'évaluer leur applicabilité à une protéine biopharmaceutique, la hCG.

2.2 Protocole de glycoprotéomique *bottom-up* et identification

Les trois familles de N-glycanes sont, par ordre de structure mannosylée croissante, les complexes, les hybrides, et les oligomannoses. Les ions oxoniums produits par ces N-glycanes seront des dérivés des GlcNAc, des NeuAc, et des hexoses (galactose et mannose principalement). L'analyse des spectres de fragmentation de glycopeptides différant uniquement par la composition de leur groupement glycosylé (séquence peptidique et charge identiques, notamment) montre que la présence et l'intensité des ions oxoniums varie en fonction de cette variable. En particulier, les ions issus de la fragmentation des hexoses (mannose ou galactose) ne sont détectés qu'à des intensités relatives très faibles dans les spectres MS/MS attribués à des glycopeptides complexes. Le processus de fragmentation privilégié pour les antennes de ces N-glycanes (formes non sialylées) ne semble donc pas passer par la perte des galactoses terminaux. Au contraire, la perte simultanée d'au moins deux monosaccharides sous la forme de l'ion oxonium GalGlcNAc (m/z 366,14) est favorable.

Les ions $[\text{Hex} + \text{H}]^+$, $[\text{Hex} \Delta\text{H}_2\text{O} + \text{H}]^+$, $[\text{Hex} \Delta 2\text{H}_2\text{O} + \text{H}]^+$, et $[\text{Hex} \Delta 2\text{H}_2\text{OCH}_2\text{O} + \text{H}]^+$ (m/z 163,06 ; 145,05 ; 127,04 et 97,03, respectivement) sont détectables dans les spectres MS/MS de N-glycopeptides *high-mannose*. De

plus, comme montré en Figure 47, l'intensité relative de ces ions est corrélée au nombre de résidus mannose dans le N-glycopeptide précurseur.

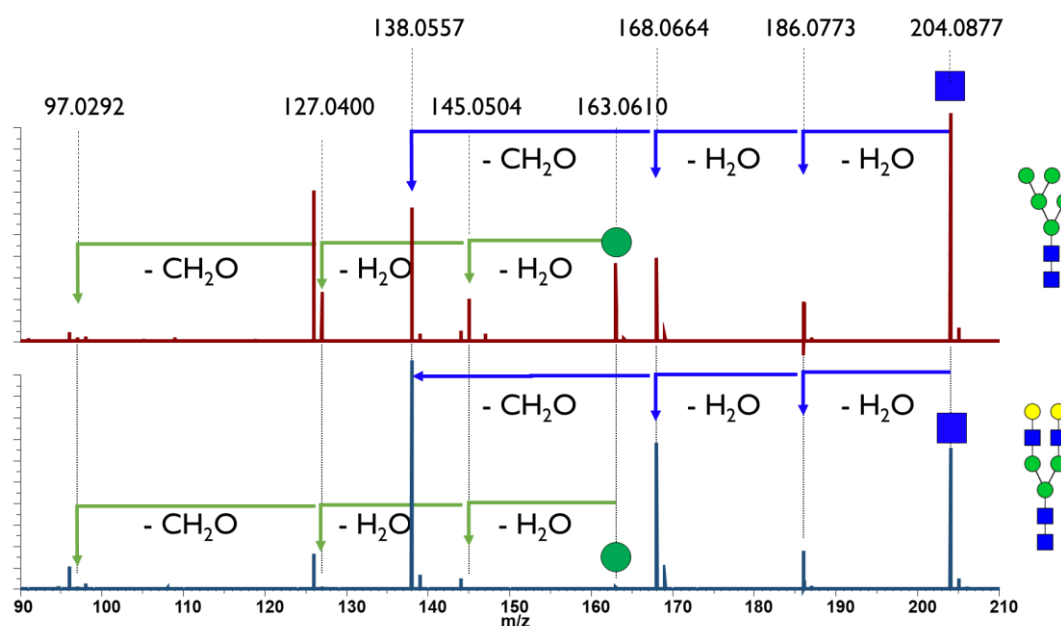


Figure 47 : Spectres de fragmentation HCD des glycopeptides NVTSESTCCVAK portant un N-glycane complexe ou high-mannose.

La structure du N-glycane identifié est indiquée en vis-à-vis de chaque spectre. L'intensité relative des ions oxoniums liés au mannose augmente avec la proportion de mannose dans le N-glycopeptide fragmenté.

L'impact des différentes caractéristiques structurales des N-glycopeptides fragmentés sur le profil des ions oxoniums a été évalué. Nous avons ici choisi délibérément de reprendre le protocole retenu afin de l'explicitier et d'argumenter les différentes étapes.

2.2.1 Préparation de l'échantillon

Le protocole suivi, de la préparation de l'échantillon jusqu'à la séparation en chromatographie liquide et la mesure de masse, est celui-ci mis en place et publié dans le cadre d'un projet connexe d'analyse de la glycosylation de la hCG après digestion sur réacteur enzymatique (147). En effet, ce protocole a permis la détection et l'identification de glycopeptides tryptiques de la hCG. Les travaux présentés à ce chapitre ayant pour objectif d'utiliser la « signature oxonium » des spectres MS/MS de N-glycopeptides, les paramètres

susceptibles d'impacter la formation d'ions et leur fragmentation doivent être précisément définis.

En résumé, l'échantillon de départ est une solution injectable d'hCG recombinante (Ovitrelle®, Organon) concentrée à 0,5µg/µL. Cette protéine est produite par transfection de l'ADN codant pour les deux sous-unités de la hCG chez les cellules CHO (*Chinese Hamster Ovary*). S'il est évident que les glycoformes synthétisées dans ces cellules ne sont pas identiques ou en concentrations égales à celles produites chez la femme (191,192), le modèle CHO est couramment utilisé pour la production de (glyco)protéines recombinantes à visée thérapeutique du fait de la similarité des processus de biosynthèse. La solution a été filtrée sur membrane (Microcon 10kDa) permettant de retenir les deux sous-unités de la hCG (masses moyennes de 14 kDa et 23 kDa pour α et β , respectivement), d'éliminer les excipients et de remplacer le solvant original par un tampon Tris (Tris 50 mM, CaCl₂ 10 mM, pH 8.0), compatible avec les réactions subséquentes.

La réduction et l'alkylation des cystéines ont été effectuées sur la membrane avec, respectivement, une solution de DTT à 5 mM (incubation de 40 minutes à 56°C), puis avec une solution d'iodoacétamide à 25 mM (incubation à température ambiante dans l'obscurité pendant 20 minutes). Ces traitements permettent de rompre les ponts disulfures présents entre résidus cystéine, au nombre de 5 pour la sous-unité α et de 6 pour la sous-unité β . Ces liaisons covalentes ayant un rôle structural important, la dénaturation de la hCG est nécessaire afin d'augmenter l'accessibilité des résidus visés par la trypsine et permettre la digestion de la protéine.

Après lavage des agents réducteurs et alkylants puis reprise en tampon Tris, l'échantillon a été digéré par la trypsine (*sequencing grade* bovine, Sigma Aldrich) à un rapport massique enzyme / protéine de 1 : 100, pendant 15h à 37°C sous agitation modérée. L'incubation longue permet la digestion complète de la hCG par la trypsine et permet l'isolation des sites de N-glycosylation sur des peptides distincts (*cf.* Figure 27). Les échantillons ont été dilués 20 fois avant

analyse par spectrométrie de masse afin d'éviter la saturation des détecteurs en MS et par soucis de comparabilité avec la digestion sur IMER entreprise en parallèle.

2.2.2 Analyse LC-MS/MS

La séparation en chromatographie liquide a été effectuée sur un système RSLCnano UltiMate 3000 (ThermoFisher Scientific). Les peptides ont été chargés en triplicats dans un tampon aqueux (98% H₂O, 2% ACN, 0,05% TFA) sur une cartouche de préconcentration (PepMap 100 C18, diamètre 300 µm) à un débit de 15 µL/min, dans le but de remplacer la solution saline dans laquelle a été faite la digestion par un solvant adapté à la séparation chromatographique. En particulier, cette étape a pour but de retirer les sels résiduels de l'échantillon, car ils ont un effet néfaste sur l'efficacité d'ionisation en mode ESI. L'échantillon est ensuite transféré sur une colonne analytique au format capillaire (PepMap 100 C18, diamètre 75 µm, longueur 50 cm, taille de particule 3 µm, porosité 100 Å). La phase mobile est constituée d'un mélange de solvant A (98% H₂O, 2% ACN, 0,1% FA) et B (10% H₂O, 90% ACN, 0,1% FA) dans des proportions initiales de 98% A et 2% B, à un débit de 220nL/min. Le gradient de la séparation chromatographique consiste en une augmentation de la part de tampon B, de 2% à 40% en 60 minutes, puis à 90% en une minute.

La chaîne de chromatographie est couplée à un appareil de type QqOrbitrap (Q Exactive, ThermoFisher Scientific) par l'intermédiaire d'une source nano-ESI. Les volumes manipulés permettent l'utilisation d'une source nano-ESI, et l'augmentation importante de la sensibilité en LC-MS par rapport aux sources ESI conventionnelles. Le spectromètre de masse est configuré pour fonctionner en mode *data-dependent* top 10, soit l'alternance d'un spectre MS (gamme de masse de 400 à 2000 m/z, résolution de 70 000 à 200 m/z, temps maximum d'injection de 100 ms, remplissage cible de la C-trap AGC 3e6) et de la fragmentation des espèces associées aux 10 pics les plus intenses (fenêtre d'isolation du précurseur de 2 m/z, gamme de masse variable avec minimum fixé à 90 m/z, résolution de 15 000 à 200 m/z, temps maximum d'injection

120 ms, remplissage cible de la C-trap AGC 1e5, énergie de collision normalisée NCE 25, exclusion systématique des précurseurs monochargés et exclusion dynamique des précurseurs fragmentés de 10s).

2.2.3 Identification glycoprotéique

L'identification des glycopeptides et l'annotation des spectres a été effectuée grâce au logiciel PMI-Byonic™ (version 2.15.7), spécialisé dans l'identification de glycoprotéines. La banque de données protéiques utilisées pour la digestion *in silico* est constituée de 5 séquences annotées dans Swiss-Prot, correspondant aux séquences confirmées des différentes isoformes de la hCG α (P01215) et β (P0DN86 ; P0DN87 ; Q6NT52 ; A6NKQ9). L'utilisation d'une banque de données réduite permet de réduire de façon importante la charge de calcul et la durée de l'interrogation, au prix de l'augmentation du taux de faux-positifs. Sur un échantillon purifié, cette pratique est néanmoins préférable du fait du fonctionnement des algorithmes d'interrogation et du système de *scoring* des résultats, conduisant à une baisse du nombre d'identification lors de l'utilisation d'une grande base de données pour un échantillon simple.

La banque de données utilisée pour les N-glycanes est une banque de structures mammaliennes « large », contenant les masses et compositions de 309 N-glycanes de mammifères (banque Byonic « 309 mammalian no sodium »). Elle permet de maximiser le nombre d'annotations, en particulier puisque les glycoprotéines ici étudiées sont produites dans un organisme hétérologue non humain. Les identifications des N-glycanes sont limités à une identification par peptide.

Les autres paramètres d'interrogation sont réglés comme décrit ci-après. Les sites de protéolyse sont ceux de la trypsine, c'est-à-dire du côté carboxylique des résidus Lysine et Arginine, avec une tolérance pour deux *missed cleavages* au maximum (afin de tenir compte de l'accessibilité limitée de certains sites). Les modifications communes (2 modifications par peptide maximum) recherchées sont : la carbamidométhylation des cystéines, lysines et résidus N-

terminaux (alkylation par l'IAM des thiols principalement, et moins fréquemment des fonctions amines); l'oxydation des méthionines (modification spontanée en solution aqueuse); la déamidation des asparagines et glutamines (réaction spontanée mais également produit de déglycosylation de l'asparagine notamment en conditions neutres ou basiques). La tolérance d'erreur sur la mesure de m/z est fixée à 10 ppm en MS et 20mDa en MS/MS.

2.2.4 Extraction des signaux d'intérêt

Les fichiers « résultats » produits par le logiciel Byonic contiennent notamment, pour chaque spectre dont l'identification a été possible, les métadonnées (échantillon, numéro de spectre, l'état de charge du précurseur, *etc.*), la séquence du peptide, ses modifications et, le cas échéant, la composition du sucre identifié, ainsi que la protéine dont est issue ce peptide.

Dans le but de prendre en compte tous les spectres, y compris ceux dont la qualité de l'identification (le score) est faible et ne présente que peu de fragments glycopeptidiques, aucun filtre particulier n'a été appliqué quant au score des résultats. Ainsi, les spectres peu intenses seront également pris en compte pour l'analyse de la fragmentation et de la production d'ions oxoniums. Afin de ne pas prendre en compte les biais introduits par la possible migration de fucose ou d'acide sialique en phase gazeuse, seuls les formes ni fucosylés ni sialylés des glycopeptides *high-mannose* (M5 à M10), hybrides (M4A1G1 à M6A1G1) ou complexes (A1G1 et A2G2) ont été conservés. De ce fait, les spectres MS/MS montrant une trace évidente de la présence d'au moins un NeuAc (fragments à m/z 274, 292, *etc.*), mais étant attribués à des formes non sialylées, ont également été retirés du jeu de données. Ces données représentent, après filtre, 246 spectres MS/MS sur un total de 51 339 spectres MS/MS (et 2 219 N-glycopeptides identifiés).

Le logiciel Byonic ne permet pas d'extraire en un fichier les données spectrales (intensités et m/z) ainsi que les métadonnées (identification, état de charge, *etc.*). Il a donc été nécessaire de compiler ces données au moyen d'une macro écrite

spécialement dans ce but. Les intensités de chaque ion oxonium sont extraites avec une tolérance de 10mDa d'erreur sur la masse et associées aux métadonnées « nom du fichier brut », « Nom de la protéine identifiée », « Séquence du peptide », « Modifications (hors glycosylation) », « Famille du N-glycane », « Composition du N-glycane », « Identifiant du spectre MS/MS », et « Etat de charge du précurseur ». Afin de normaliser les intensités des ions étudiés, des ratios sont calculés entre les fragments issus de la GlcNAc ou de l'hexose et ayant subi la même voie de fragmentation (monosaccharide entier, mono-déshydraté, di-déshydraté, *etc.*).

Les rapports calculés sont les suivants :

$$\begin{aligned} & I_{\text{GlcNAc}} / I_{\text{Hexose}} \\ & I_{\text{GlcNAc } \Delta\text{H}_2\text{O}} / I_{\text{Hexose } \Delta\text{H}_2\text{O}} \\ & I_{\text{GlcNAc } \Delta 2\text{H}_2\text{O}} / I_{\text{Hexose } \Delta 2\text{H}_2\text{O}} \\ & I_{\text{GlcNAc } \Delta 2\text{H}_2\text{O}-\text{CH}_2\text{O}} / I_{\text{Hexose } \Delta 2\text{H}_2\text{O}-\text{CH}_2\text{O}} \end{aligned}$$

Parmi les 246 spectres considérés, la présence de certains spectres MS/MS ne contenant pas certains fragments induit que certains de ces ratios ne pourront pas toujours être calculés. Les N-glycanes complexes en particulier ne produisent que relativement peu de fragments issus du mannose. Dans le cadre de ce développement de méthode, seuls les spectres ayant permis de calculer ces 4 ratios sont conservés pour l'analyse statistique. Ce filtre permet la conservation de 98 spectres.

Ces données permettront d'appliquer cette approche à un échantillon, la hCG, et non pas à standard, pour la première fois. De plus, l'impact de plusieurs facteurs sur la population des ions oxoniums sera évalué.

2.3 Résultats : Analyse des données spectrales

Tout comme l'avaient déjà remarqué d'autres groupes (153), l'intensité des ions oxoniums liés au mannose augmente drastiquement lorsque des N-glycopeptides de la famille des *high-mannose* sont sélectionnés et fragmentés. Les données expérimentales et les ratios calculés sont intégralement répertoriés dans la table suivante (Tableau 6).

Tableau 6 : Identification de glycopeptides par analyse des ions oxoniums.

98 spectres MS/MS ont été filtrés. Les données issues de l'identification par Byonic ainsi que la quantification des fragments oxoniums sont récapitulés dans cette table. Les ions issus de la GlcNAc sont représentés en bleu, ceux issus du mannose ou d'un autre hexose sont représentés en vert.

Métadonnées								Intensités								Ratios			
Réplikat #	Protéine	Seq. Peptide	Mod. Peptide	Glyc. (Fam.)	Glycane	ID Spectre	z	m/z 97.0	m/z 127.0	m/z 138.1	m/z 45.1	m/z 163.1	m/z 168.1	m/z 186.1	m/z 204.1	Int. 204/163	Int. 186/145	Int. 168/127	Int. 138/97
r3	CGB3	CRPINATLAVEK	Cam1	Complex	A1G1	4956	3	1.77E+03	3.18E+03	2.38E+05	2.00E+03	3.61E+03	1.12E+05	4.53E+04	1.27E+05	35.322	22.610	35.317	134.576
r1	CGB3	CRPINATLAVEK	Cam1	Complex	A2G2	5252	3	1.68E+04	1.35E+04	1.81E+06	4.04E+03	3.80E+03	1.00E+06	2.93E+05	8.55E+05	225.147	72.585	74.114	107.563
r1	CGB3	CRPINATLAVEK	Cam1	Complex	A2G2	5186	3	5.52E+04	1.74E+04	6.03E+06	1.75E+04	1.43E+04	3.20E+06	9.71E+05	2.67E+06	187.082	55.462	183.991	109.357
r1	CGB3	CRPINATLAVEK	Cam1	Complex	A2G2	5116	3	1.27E+05	7.87E+04	1.54E+07	3.01E+04	2.76E+04	8.69E+06	2.64E+06	7.07E+06	256.103	87.687	110.474	121.029
r1	CGB3	CRPINATLAVEK	Cam1	Complex	A2G2	5047	3	2.73E+04	1.61E+04	2.53E+06	3.44E+03	3.48E+03	1.39E+06	4.31E+05	1.19E+06	342.328	125.286	86.411	92.719
r2	CGB3	CRPINATLAVEK	Cam1	Complex	A2G2	4983	3	2.46E+04	1.92E+04	3.18E+06	5.71E+03	4.21E+03	1.75E+06	5.13E+05	1.48E+06	352.071	89.760	91.063	129.384
r2	CGB3	CRPINATLAVEK	Cam1	Complex	A2G2	4917	3	9.63E+04	5.77E+04	9.19E+06	2.09E+04	1.65E+04	5.29E+06	1.50E+06	4.32E+06	262.079	71.922	91.703	95.446
r2	CGB3	CRPINATLAVEK	Cam1	Complex	A2G2	4846	3	5.29E+04	4.18E+04	6.91E+06	2.52E+04	2.11E+04	3.89E+06	1.15E+06	3.26E+06	154.353	45.671	92.913	130.575
r3	CGB3	CRPINATLAVEK	Cam1	Complex	A2G2	4917	3	7.69E+04	4.50E+04	8.60E+06	2.33E+04	1.37E+04	4.74E+06	1.42E+06	4.08E+06	298.219	60.938	105.377	111.910
r3	CGB3	CRPINATLAVEK	Cam1	Complex	A2G2	4845	3	1.31E+05	8.20E+04	1.22E+07	2.01E+04	2.38E+04	6.69E+06	2.09E+06	5.84E+06	245.684	104.155	81.562	93.060
r1	CGB3	CRPINATLAVEK	Cam1	HM	M10	5090	3	3.22E+04	8.12E+04	7.60E+04	1.20E+05	1.43E+05	5.01E+04	2.15E+04	9.34E+04	0.654	0.180	0.617	2.362
r1	CGB3	CRPINATLAVEK	Cam1	HM	M10	5024	3	5.63E+04	1.27E+05	8.93E+04	2.19E+05	2.54E+05	5.09E+04	2.21E+04	1.24E+05	0.488	0.101	0.400	1.586
r2	CGB3	CRPINATLAVEK	Cam1	HM	M10	4813	3	4.01E+04	9.95E+04	6.26E+04	1.58E+05	1.68E+05	3.63E+04	1.64E+04	9.42E+04	0.562	0.104	0.365	1.563
r2	CGB3	CRPINATLAVEK	Cam1	HM	M10	4747	3	3.38E+04	9.03E+04	4.32E+04	1.19E+05	1.39E+05	2.92E+04	1.27E+04	6.67E+04	0.481	0.107	0.323	1.277
r3	CGB3	CRPINATLAVEK	Cam1	HM	M10	4796	3	4.40E+04	1.17E+05	6.92E+04	1.76E+05	1.98E+05	4.58E+04	1.90E+04	1.17E+05	0.589	0.108	0.390	1.571
r3	CGB3	CRPINATLAVEK	Cam1	HM	M10	4730	3	4.68E+04	1.09E+05	6.33E+04	1.57E+05	1.78E+05	4.13E+04	1.93E+04	8.44E+04	0.475	0.122	0.378	1.353
r1	CGB3	CRPINATLAVEK	Cam1	Hybrid	M4A1G1	5169	3	2.03E+03	6.24E+03	2.85E+05	8.28E+03	9.24E+03	1.48E+05	5.16E+04	1.69E+05	18.327	6.230	23.772	140.108
r2	CGB3	CRPINATLAVEK	Cam1	Hybrid	M4A1G1	4996	3	2.25E+03	3.55E+03	1.33E+05	1.71E+03	2.78E+03	6.57E+04	2.39E+04	8.43E+04	30.304	13.976	18.498	59.296
r2	CGB3	CRPINATLAVEK	Cam1	Hybrid	M4A1G1	4934	3	2.73E+03	3.43E+03	2.26E+05	3.21E+03	7.02E+03	1.09E+05	3.81E+04	1.09E+05	15.585	11.876	31.714	82.807
r3	CGB3	CRPINATLAVEK	Cam1	Hybrid	M4A1G1	4908	3	3.68E+03	8.21E+03	2.93E+05	7.83E+03	7.11E+03	1.34E+05	4.51E+04	1.72E+05	24.192	5.760	16.334	79.664
r1	CGB3	CRPINATLAVEK	Cam1	HM	M5	5399	3	1.77E+03	1.10E+04	1.96E+04	9.31E+03	1.23E+04	1.03E+04	7.69E+03	2.52E+04	2.055	0.826	0.935	11.116
r1	CGB3	CRPINATLAVEK	Cam1	HM	M5	5328	3	7.21E+03	3.38E+04	4.78E+04	3.25E+04	3.09E+04	2.75E+04	1.18E+04	6.13E+04	1.986	0.363	0.813	6.621

r1	CGB3	CRPINATLAVEK	Cam1	HM	M5	5253	3	2.46E+04	1.09E+05	1.71E+05	8.90E+04	1.07E+05	9.93E+04	4.62E+04	2.10E+05	1.964	0.519	0.911	6.954
r1	CGB3	CRPINATLAVEK	Cam1	HM	M5	5187	3	7.36E+04	3.59E+05	6.28E+05	3.51E+05	3.86E+05	3.26E+05	1.43E+05	6.84E+05	1.775	0.408	0.908	8.533
r1	CGB3	CRPINATLAVEK	Cam1	HM	M5	5156	2	2.20E+03	1.17E+04	8.28E+04	2.15E+04	2.66E+04	4.16E+04	1.67E+04	1.30E+05	4.887	0.780	3.559	37.542
r1	CGB3	CRPINATLAVEK	Cam1	HM	M5	5123	3	7.23E+04	3.48E+05	5.49E+05	3.35E+05	3.44E+05	3.00E+05	1.57E+05	6.73E+05	1.958	0.468	0.862	7.587
r2	CGB3	CRPINATLAVEK	Cam1	HM	M5	5147	3	3.35E+03	1.38E+04	2.32E+04	1.09E+04	1.30E+04	1.36E+04	7.25E+03	2.33E+04	1.796	0.665	0.991	6.932
r2	CGB3	CRPINATLAVEK	Cam1	HM	M5	5081	3	8.34E+03	3.74E+04	6.00E+04	3.52E+04	3.37E+04	3.01E+04	1.35E+04	8.29E+04	2.459	0.382	0.806	7.193
r2	CGB3	CRPINATLAVEK	Cam1	HM	M5	5016	3	2.13E+04	1.21E+05	1.91E+05	1.10E+05	1.21E+05	1.15E+05	5.11E+04	2.44E+05	2.008	0.463	0.947	8.965
r2	CGB3	CRPINATLAVEK	Cam1	HM	M5	4956	3	7.20E+04	3.35E+05	5.65E+05	3.34E+05	3.45E+05	3.01E+05	1.38E+05	6.56E+05	1.905	0.413	0.898	7.843
r2	CGB3	CRPINATLAVEK	Cam1	HM	M5	4890	3	3.62E+04	1.84E+05	2.90E+05	1.72E+05	1.72E+05	1.70E+05	8.30E+04	3.50E+05	2.037	0.482	0.925	8.009
r3	CGB3	CRPINATLAVEK	Cam1	HM	M5	5136	3	1.75E+03	1.31E+04	3.27E+04	9.71E+03	1.67E+04	1.13E+04	2.58E+03	2.85E+04	1.707	0.266	0.863	18.653
r3	CGB3	CRPINATLAVEK	Cam1	HM	M5	5070	3	3.35E+03	3.85E+04	5.92E+04	3.54E+04	3.51E+04	3.09E+04	1.59E+04	6.92E+04	1.971	0.448	0.804	17.649
r3	CGB3	CRPINATLAVEK	Cam1	HM	M5	5000	3	1.65E+04	1.22E+05	1.93E+05	1.09E+05	1.15E+05	1.02E+05	5.30E+04	2.18E+05	1.895	0.486	0.841	11.723
r3	CGB3	CRPINATLAVEK	Cam1	HM	M5	4939	3	8.15E+04	3.64E+05	5.72E+05	3.45E+05	3.57E+05	3.20E+05	1.70E+05	7.15E+05	2.004	0.493	0.880	7.019
r3	CGB3	CRPINATLAVEK	Cam1	HM	M5	4873	3	5.79E+04	3.17E+05	4.67E+05	2.92E+05	2.91E+05	2.67E+05	1.30E+05	5.61E+05	1.930	0.445	0.841	8.069
r1	CGB3	CRPINATLAVEK	Cam1	Hybrid	M5A1G1	5124	3	2.06E+03	9.52E+03	2.91E+05	8.51E+03	9.17E+03	1.37E+05	4.99E+04	1.69E+05	18.477	5.861	14.434	141.170
r2	CGB3	CRPINATLAVEK	Cam1	Hybrid	M5A1G1	4945	3	2.17E+03	2.37E+03	1.05E+05	2.99E+03	3.53E+03	5.64E+04	1.92E+04	6.87E+04	19.470	6.428	23.803	48.149
r2	CGB3	CRPINATLAVEK	Cam1	Hybrid	M5A1G1	4879	3	1.66E+03	3.87E+03	2.09E+05	3.12E+03	6.84E+03	1.14E+05	3.63E+04	1.21E+05	17.724	11.609	29.470	126.216
r3	CGB3	CRPINATLAVEK	Cam1	Hybrid	M5A1G1	4852	3	2.12E+03	7.86E+03	2.46E+05	9.66E+03	7.72E+03	1.32E+05	3.87E+04	1.53E+05	19.826	4.007	16.781	115.715
r1	CGB3	CRPINATLAVEK	Cam1	HM	M6	5235	3	2.45E+03	2.21E+04	2.55E+04	2.02E+04	2.67E+04	1.05E+04	7.07E+03	3.29E+04	1.228	0.350	0.478	10.399
r1	CGB3	CRPINATLAVEK	Cam1	HM	M6	5167	3	1.47E+04	9.16E+04	9.43E+04	9.00E+04	1.15E+05	5.86E+04	2.73E+04	1.25E+05	1.082	0.303	0.640	6.425
r1	CGB3	CRPINATLAVEK	Cam1	HM	M6	5101	3	2.69E+04	1.55E+05	1.63E+05	1.59E+05	1.65E+05	9.12E+04	4.50E+04	1.99E+05	1.203	0.283	0.589	6.052
r2	CGB3	CRPINATLAVEK	Cam1	HM	M6	4968	3	5.99E+03	3.35E+04	3.62E+04	3.54E+04	4.23E+04	2.36E+04	1.29E+04	4.51E+04	1.065	0.365	0.705	6.043
r2	CGB3	CRPINATLAVEK	Cam1	HM	M6	4901	3	2.04E+04	1.16E+05	1.19E+05	1.23E+05	1.37E+05	6.36E+04	3.33E+04	1.55E+05	1.134	0.272	0.549	5.855
r2	CGB3	CRPINATLAVEK	Cam1	HM	M6	4836	3	6.51E+03	5.30E+04	5.85E+04	5.82E+04	6.13E+04	3.07E+04	1.63E+04	7.07E+04	1.154	0.281	0.579	8.988
r3	CGB3	CRPINATLAVEK	Cam1	HM	M6	4971	3	1.61E+03	1.79E+04	2.77E+04	2.71E+04	2.97E+04	1.65E+04	3.15E+03	3.47E+04	1.167	0.116	0.921	17.198
r3	CGB3	CRPINATLAVEK	Cam1	HM	M6	4906	3	1.71E+04	1.02E+05	1.04E+05	1.04E+05	1.22E+05	6.69E+04	3.06E+04	1.45E+05	1.189	0.295	0.653	6.059
r3	CGB3	CRPINATLAVEK	Cam1	HM	M6	4842	3	2.14E+04	1.24E+05	1.32E+05	1.48E+05	1.53E+05	6.94E+04	3.72E+04	1.86E+05	1.217	0.250	0.561	6.167

r1	CGB3	CRPINATLAVEK	Cam1	HM	M7	5138	3	6.28E+03	3.11E+04	2.64E+04	3.96E+04	4.39E+04	1.24E+04	8.25E+03	3.89E+04	0.887	0.208	0.397	4.205
r1	CGB3	CRPINATLAVEK	Cam1	HM	M7	5068	3	6.21E+03	4.36E+04	4.15E+04	6.28E+04	7.39E+04	2.83E+04	1.26E+04	6.12E+04	0.829	0.200	0.648	6.681
r2	CGB3	CRPINATLAVEK	Cam1	HM	M7	4892	3	5.95E+03	2.99E+04	2.88E+04	3.05E+04	3.95E+04	1.54E+04	8.15E+03	3.21E+04	0.812	0.267	0.516	4.838
r2	CGB3	CRPINATLAVEK	Cam1	HM	M7	4826	3	2.54E+03	3.65E+04	3.62E+04	4.33E+04	5.18E+04	1.45E+04	1.01E+04	4.84E+04	0.935	0.234	0.397	14.289
r3	CGB3	CRPINATLAVEK	Cam1	HM	M7	4864	3	3.82E+03	3.62E+04	3.96E+04	4.15E+04	4.86E+04	2.04E+04	8.87E+03	4.61E+04	0.948	0.214	0.562	10.371
r3	CGB3	CRPINATLAVEK	Cam1	HM	M7	4797	3	3.70E+03	3.67E+04	3.48E+04	4.69E+04	5.92E+04	2.52E+04	9.68E+03	5.22E+04	0.881	0.207	0.688	9.399
r1	CGB3	CRPINATLAVEK	Cam1	HM	M8	5129	3	5.86E+03	2.48E+04	3.24E+04	2.80E+04	3.28E+04	1.78E+04	2.47E+03	3.01E+04	0.917	0.088	0.720	5.538
r1	CGB3	CRPINATLAVEK	Cam1	HM	M8	5057	3	8.22E+03	5.37E+04	4.19E+04	6.48E+04	1.14E+05	3.04E+04	8.66E+03	6.00E+04	0.528	0.133	0.566	5.105
r2	CGB3	CRPINATLAVEK	Cam1	HM	M8	4858	3	2.62E+03	2.20E+04	2.57E+04	3.07E+04	4.13E+04	1.25E+04	7.23E+03	2.67E+04	0.646	0.236	0.568	9.806
r2	CGB3	CRPINATLAVEK	Cam1	HM	M8	4792	3	6.69E+03	3.53E+04	2.38E+04	4.67E+04	6.07E+04	1.77E+04	7.66E+03	3.42E+04	0.563	0.164	0.502	3.553
r3	CGB3	CRPINATLAVEK	Cam1	HM	M8	4843	3	7.45E+03	3.46E+04	3.36E+04	4.02E+04	5.36E+04	1.99E+04	8.30E+03	4.10E+04	0.765	0.207	0.575	4.505
r3	CGB3	CRPINATLAVEK	Cam1	HM	M8	4774	3	6.96E+03	4.66E+04	4.14E+04	5.53E+04	8.32E+04	1.83E+04	1.05E+04	5.11E+04	0.614	0.190	0.393	5.944
r3	CGB3	PINATLAVEK	NA	HM	M5	6138	2	2.02E+03	3.33E+03	4.19E+04	5.77E+03	8.25E+03	2.09E+04	1.03E+04	4.29E+04	5.206	1.776	6.257	20.785
r1	GLHA	NVTSESTCCVAK	Cam2	Complex	A1G1	3556	3	2.98E+03	3.16E+03	3.13E+05	4.85E+03	6.45E+03	1.80E+05	5.85E+04	2.63E+05	40.773	12.071	56.822	104.740
r1	GLHA	NVTSESTCCVAK	Cam2	Complex	A1G1	3490	3	2.72E+03	2.51E+03	5.01E+05	3.61E+03	1.07E+04	2.52E+05	9.34E+04	4.00E+05	37.509	25.878	100.622	184.107
r3	GLHA	NVTSESTCCVAK	Cam2	Complex	A1G1	3298	3	2.40E+03	2.35E+03	2.48E+05	1.84E+03	6.18E+03	1.34E+05	5.42E+04	2.10E+05	33.964	29.440	56.779	103.429
r1	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3593	3	1.82E+04	1.20E+04	1.89E+06	4.46E+03	1.56E+04	1.16E+06	3.37E+05	1.16E+06	74.657	75.490	96.832	103.869
r1	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3531	3	5.03E+04	3.00E+04	5.86E+06	2.65E+04	2.93E+04	3.52E+06	1.05E+06	3.56E+06	121.668	39.708	117.503	116.531
r1	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3462	3	1.92E+05	7.91E+04	1.94E+07	6.93E+04	1.12E+05	1.20E+07	3.26E+06	1.15E+07	103.061	46.999	151.910	100.858
r1	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3412	2	5.74E+03	3.17E+03	1.10E+06	2.60E+03	5.29E+03	8.19E+05	1.81E+05	6.52E+05	123.118	69.534	258.183	191.951
r1	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3401	3	1.39E+05	8.21E+04	1.57E+07	4.56E+04	1.07E+05	9.51E+06	2.80E+06	9.50E+06	89.074	61.369	115.855	113.071
r2	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3442	3	6.27E+03	2.33E+03	4.82E+05	1.74E+03	2.72E+03	2.75E+05	7.91E+04	2.75E+05	101.096	45.550	118.223	76.808
r2	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3380	3	8.70E+03	6.61E+03	1.04E+06	4.26E+03	3.86E+03	6.39E+05	1.78E+05	5.82E+05	150.805	41.824	96.695	119.013
r2	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3311	3	3.90E+04	1.78E+04	3.41E+06	1.39E+04	2.21E+04	2.03E+06	5.94E+05	2.03E+06	92.013	42.693	114.097	87.604
r2	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3249	3	9.73E+04	7.15E+04	1.07E+07	6.00E+04	3.83E+04	6.38E+06	2.01E+06	6.29E+06	164.009	33.492	89.195	110.469
r2	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3195	3	3.73E+04	2.31E+04	4.07E+06	1.50E+04	2.22E+04	2.31E+06	6.74E+05	2.25E+06	101.283	44.836	100.057	109.381
r2	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3181	3	8.49E+04	6.58E+04	1.14E+07	3.42E+04	7.52E+04	6.84E+06	2.05E+06	6.82E+06	90.663	60.091	104.042	134.412

r3	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3379	3	1.44E+04	8.24E+03	1.65E+06	4.14E+03	5.62E+03	9.89E+05	2.86E+05	9.87E+05	175.605	69.166	119.992	114.711
r3	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3318	3	4.40E+04	2.94E+04	4.93E+06	1.39E+04	2.62E+04	2.97E+06	8.81E+05	2.95E+06	112.313	63.283	101.021	111.973
r3	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3253	3	1.24E+05	9.94E+04	1.58E+07	6.34E+04	8.30E+04	9.72E+06	2.81E+06	9.37E+06	112.896	44.364	97.784	127.572
r3	GLHA	NVTSESTCCVAK	Cam2	Complex	A2G2	3190	3	4.30E+04	3.79E+04	5.41E+06	1.91E+04	4.06E+04	3.37E+06	9.49E+05	3.29E+06	81.091	49.801	89.001	125.792
r1	GLHA	NVTSESTCCVAK	Cam2	HM	M10	3235	3	3.63E+03	1.69E+04	2.55E+04	3.22E+04	6.37E+04	2.20E+04	9.75E+03	5.48E+04	0.860	0.303	1.300	7.037
r1	GLHA	NVTSESTCCVAK	Cam2	Hybrid	M4A1G1	3489	3	4.11E+03	1.08E+04	6.21E+05	1.16E+04	2.67E+04	3.38E+05	1.28E+05	5.48E+05	20.554	11.052	31.431	151.206
r1	GLHA	NVTSESTCCVAK	Cam2	Hybrid	M4A1G1	3419	3	1.31E+04	1.90E+04	1.06E+06	1.97E+04	4.07E+04	6.10E+05	2.17E+05	9.84E+05	24.201	10.987	32.052	80.955
r2	GLHA	NVTSESTCCVAK	Cam2	Hybrid	M4A1G1	3262	3	1.55E+03	3.17E+03	2.86E+05	5.00E+03	1.03E+04	1.65E+05	5.72E+04	2.52E+05	24.541	11.452	51.981	185.030
r2	GLHA	NVTSESTCCVAK	Cam2	Hybrid	M4A1G1	3194	3	3.60E+03	6.90E+03	4.54E+05	3.99E+03	1.46E+04	2.61E+05	9.62E+04	4.07E+05	27.860	24.111	37.832	125.979
r3	GLHA	NVTSESTCCVAK	Cam2	Hybrid	M4A1G1	3273	3	3.49E+03	2.66E+03	3.79E+05	8.59E+03	1.55E+04	2.18E+05	7.70E+04	3.50E+05	22.615	8.960	81.874	108.736
r3	GLHA	NVTSESTCCVAK	Cam2	Hybrid	M4A1G1	3211	3	2.13E+03	1.01E+04	3.56E+05	9.64E+03	1.88E+04	1.86E+05	7.15E+04	2.95E+05	15.738	7.420	18.427	167.218
r1	GLHA	NVTSESTCCVAK	Cam2	HM	M5	3554	3	2.18E+03	1.02E+04	3.44E+04	9.98E+03	1.79E+04	1.98E+04	1.12E+04	5.76E+04	3.215	1.125	1.944	15.797
r1	GLHA	NVTSESTCCVAK	Cam2	HM	M5	3481	3	8.30E+03	4.08E+04	1.19E+05	5.13E+04	6.95E+04	7.51E+04	4.07E+04	2.21E+05	3.180	0.794	1.842	14.314
r1	GLHA	NVTSESTCCVAK	Cam2	HM	M5	3414	3	1.32E+04	6.97E+04	2.44E+05	8.06E+04	1.25E+05	1.32E+05	7.91E+04	4.08E+05	3.252	0.982	1.899	18.533
r1	GLHA	NVTSESTCCVAK	Cam2	HM	M5	3403	2	2.45E+03	1.47E+04	2.48E+05	3.18E+04	8.28E+04	1.89E+05	9.08E+04	6.27E+05	7.576	2.852	12.896	101.302
r2	GLHA	NVTSESTCCVAK	Cam2	HM	M5	3272	3	2.15E+03	1.27E+04	4.48E+04	1.57E+04	2.32E+04	2.42E+04	1.47E+04	7.71E+04	3.323	0.938	1.909	20.869
r2	GLHA	NVTSESTCCVAK	Cam2	HM	M5	3204	3	7.31E+03	2.94E+04	1.03E+05	3.46E+04	4.89E+04	6.25E+04	3.39E+04	1.84E+05	3.769	0.979	2.125	14.144
r3	GLHA	NVTSESTCCVAK	Cam2	HM	M5	3289	3	1.55E+03	1.20E+04	4.64E+04	1.55E+04	2.40E+04	2.44E+04	1.58E+04	7.37E+04	3.069	1.016	2.035	29.915
r3	GLHA	NVTSESTCCVAK	Cam2	HM	M5	3222	3	2.92E+03	2.96E+04	1.10E+05	3.57E+04	5.25E+04	5.67E+04	3.24E+04	1.77E+05	3.363	0.909	1.919	37.564
r1	GLHA	NVTSESTCCVAK	Cam2	Hybrid	M5A1G1	3380	3	2.98E+03	9.94E+03	3.99E+05	9.46E+03	2.84E+04	2.15E+05	7.86E+04	3.83E+05	13.479	8.308	21.646	133.946
r1	GLHA	NVTSESTCCVAK	Cam2	Hybrid	M5A1G1	3308	3	2.74E+03	2.75E+03	1.85E+05	1.44E+03	8.73E+03	1.06E+05	3.32E+04	1.58E+05	18.112	23.103	38.354	67.654
r1	GLHA	NVTSESTCCVAK	Cam2	HM	M6	3338	3	2.14E+03	1.36E+04	3.90E+04	2.00E+04	3.42E+04	2.14E+04	1.07E+04	7.86E+04	2.297	0.537	1.573	18.196

Afin de comparer l'importance de chacun des ratios calculés et d'évaluer l'importance des caractéristiques des précurseurs sur la fragmentation, une analyse en composantes principales a été effectuée. Elle permet de réduire le nombre de variables par changement de l'espace des données et de redistribuer les variables pour trouver celles qui ont le plus d'importance dans la séparation des différents spectres. Les ratios sont définis comme les variables numériques, et les métadonnées comme les groupes (variables nominales). Une linéarisation des données est nécessaire (ici un logarithme en base 2), car le calcul de ratios d'intensités implique par nature des données très densément regroupées entre 0 et 1 pour les spectres dont les ions oxoniums de mannose sont majoritaires. La significativité des nouvelles variables (composantes principales, ou PC) est déterminée par leur importance dans la variance totale des données (Figure 48).

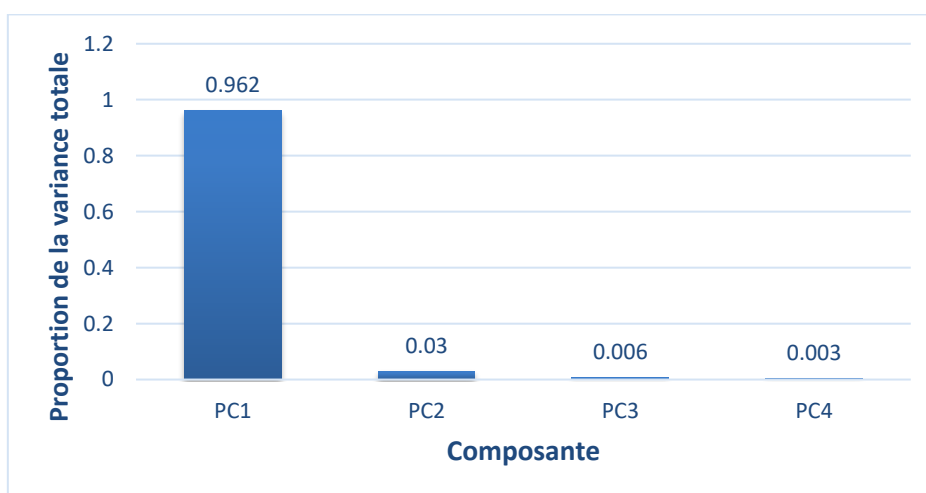


Figure 48 : Histogramme des contributions à la variance totale de chaque composante principale. La composante PC1 permet à elle seule d'expliquer 96,2 % de la variance.

Tableau 7 : Valeurs propres des composantes principales.

Elles représentent la significativité de chacune des composantes principales, car elles représentent les axes de plus grande variance où les données sont susceptibles de changer.

Composante Principale # Valeur propre

1	3.84604167176114
2	0.118015839128387
3	0.0239786253093598
4	0.0119638638011166

La composante principale 1 (PC1) permet d'expliquer la grande majorité de la variance des différents points (96,2 %), et la PC2 permet d'en expliquer 3 %. La composante PC1 a une valeur propre élevée et supérieure à 1 ($\lambda_1 = 3,84$) montrant une grande séparation des différents ratios dans cette dimension.

L'analyse détaillée des vecteurs propres de chaque composante principale (Tableau 8) montre en particulier que dans l'espace PC1/PC2, les variables 204/163 et 138/97 sont quasi-orthogonales, indiquant leur indépendance. Au contraire, les ratios 204/163, 186/145 et 168/127 ont des vecteurs propres très similaires dans l'espace PC1/PC2, indiquant une forte redondance de l'information qu'ils apportent. S'agissant des ratios des formes intactes, mono-déshydratée et di-déshydratée des monosaccharides, respectivement, il ne semble pas aberrant que leur formation en MS/MS soit fortement corrélée. La perte de deux H₂O et d'un groupement CH₂O formant les fragments à 138 m/z (GlcNAc $\Delta 2\text{H}_2\text{O}-\text{CH}_2\text{O}$) ou à 97 m/z (Hex $\Delta 2\text{H}_2\text{O}-\text{CH}_2\text{O}$) permet, dans la dimension PC2, un bon degré de séparation.

Tableau 8 : Vecteurs propres des composantes principales.

La PC1 est affectée de manière quasi-identique par toutes les variables, tandis que la composante PC2, bien que moins significative, permet une résolution plus fine des différents ratios.

VARIABLES	PC1	PC2	PC3	PC4
204/163	-0.50202	-0.40750	-0.58635	0.48796
186/145	-0.50508	-0.30609	-0.02604	-0.80655
168/127	-0.50538	-0.11185	0.78795	0.33349
138/97	-0.48730	0.85308	-0.18612	-0.01258

La représentation des valeurs dans l'espace PC1/PC2 (Figure 49) vise à mettre en évidence une séparation des classes de N-glycanes (complexes, hybrides, et *high-mannose*) dans ces deux dimensions.

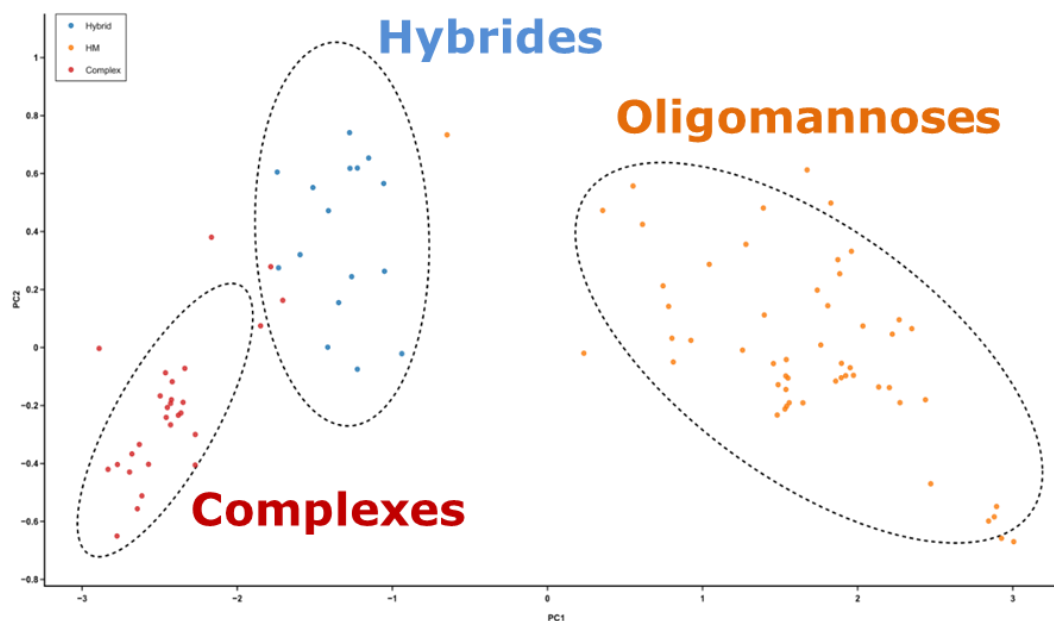


Figure 49 : Analyse de composante principale, ions oxoniums, clusterisation par famille.

Le regroupement par famille de N-glycane montre une séparation évidente des trois classes, sans distinction faite de la nature du peptide ou d'aucune autre variable. L'espace est constitué des PC1 et PC2, chaque point correspond à un glycopeptide dans cet espace (rouge pour les complexes, bleus pour les hybrides, jaunes pour les high-mannose). Les régions entourées correspondent à l'ellipse comprenant à 95 % de confiance les valeurs issues de chaque groupe pour une distribution de type Student.

Ce changement de variable permet un évident regroupement des caractéristiques spectrales en fonction de la famille des N-glycopeptides. Cette représentation montre effectivement une forte séparation dans la dimension PC1, comme le laissait supposer le vecteur propre de cette variable. La PC2 (axe vertical), quant à elle, permet d'augmenter la séparation entre les points des classes de glycopeptides complexes et hybrides, notamment. On observe une disparité des valeurs variant en fonction des classes, pouvant être due à différents facteurs physico-chimiques, en particulier la nature exacte du sucre porté par le glycopeptide.

En utilisant non pas les familles de N-glycanes (complexes, etc.), mais la composition exacte du glycane (par exemple, A2G2, M5, etc.) pour la clusterisation des données, on cherche à expliquer la variabilité intra-classe de la précédente visualisation.

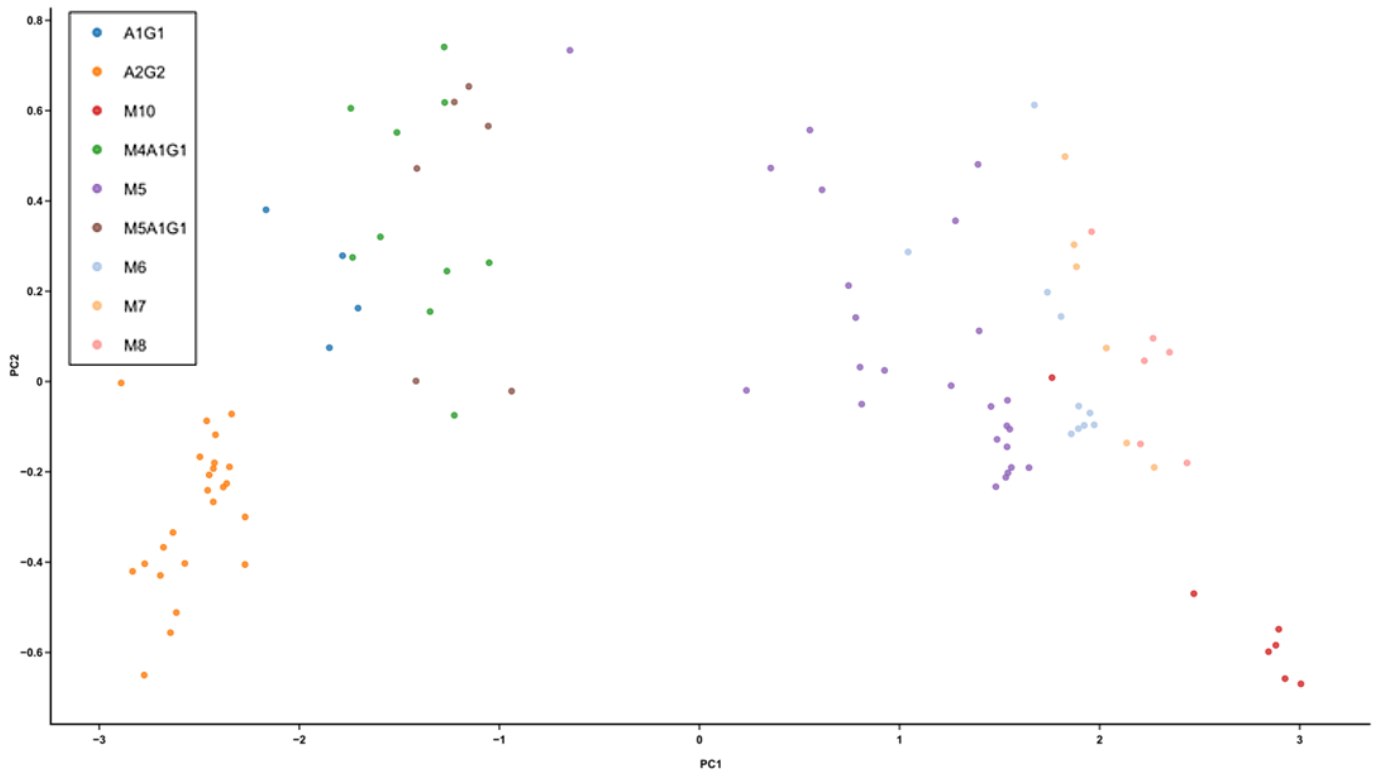


Figure 50 : Analyse de composante principale, ions oxoniums, clusterisation par structure.

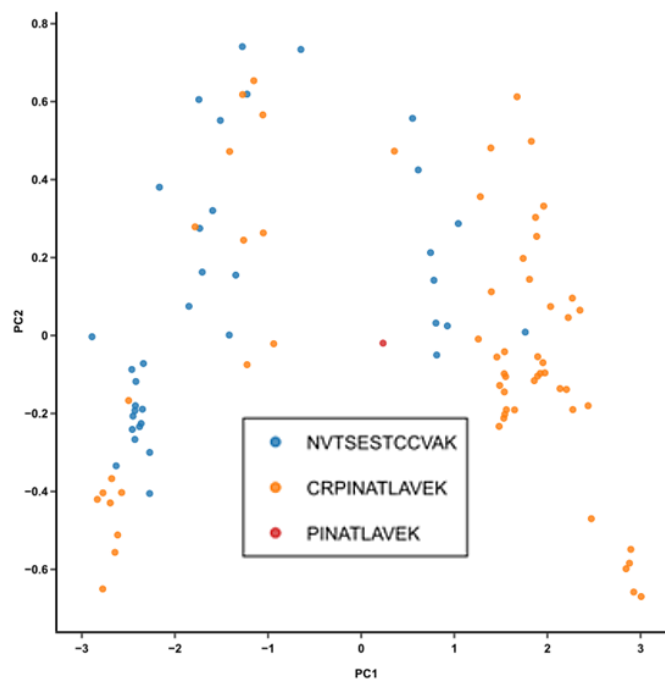


Figure 51 : Analyse de composante principale, ions oxoniums, clusterisation par séquence peptidique.

Bien qu'elle ne soit pas aussi franche entre chaque classe (composition du glycane) avec cette clusterisation (Figure 50), il apparaît clairement une séparation entre les différentes compositions au sein des N-glycanes *high-*

mannose, ainsi qu'entre les complexes A1G1 et A2G2. Les hybrides semblent présenter une variabilité plus importante, rendant les formes A1G1, M4A1G1 et M5A1G1 peu distinguables dans ces dimensions.

Au sein de la classe des N-glycanes *high-mannose*, la valeur de la PC1 augmente avec l'augmentation du nombre de mannoses. Cela signifie, en termes de variables originales, que les ratios calculés diminuent et donc que l'intensité relative des oxoniums liés au mannose augmente. La classe « M10 » présente une valeur jugée aberrante vis-à-vis du reste du groupe (*outlier*), dont les coordonnées sont (1,762 ; 0,009) quand la moyenne des points du reste du groupe est de (2,837 ; 0,588). Un examen plus approfondi de cette valeur montre qu'il s'agit en réalité d'un glycopeptide dont la séquence est NVTSESTCCVAK, tandis que le reste des glycopeptides « M10 » a pour séquence CRPINATLAVEK, avec le même état de charge ($z=3$). Il est prématuré de définir une tendance à partir d'une valeur unique, et cela nécessiterait une meilleure répétition des mesures pour un tel glycopeptide. Cette observation semble néanmoins en contradiction avec les conclusions de Hoffmann *et al.*(153) qui ne voyaient aucune corrélation entre les abondances des ions oxoniums issus d'une N-glycoforme spécifique et la séquence de son squelette peptidique. La famille des N-glycopeptides complexes A2G2 montre la même hétérogénéité intra-classe, majoritairement dans la dimension PC2, qui peut majoritairement être attribuée à la séquence du peptide porteur de ce sucre (Figure 51).

Afin de vérifier s'il s'agit de variations dues à des facteurs non-considérés ou si la formation des ions oxoniums est dépendante du peptide porteur d'un N-glycane donné, il serait également intéressant de prendre en compte les informations de liaison et de stéréochimie en plus de la composition des sucres, qui n'ont pas non plus été prises en compte dans les articles précédemment cités, par spectrométrie de masse couplée à la mobilité ionique notamment.

L'impact de l'état de charge du précurseur sur la formation des ions oxoniums a été jugée peu significative dans les publications antérieures à nos travaux.

Nos données ne contiennent que peu de fragments issus d'ions doublement chargés, l'état de charge majoritaire étant $z=3$. Sans être assuré de la significativité statistique de cette observation, il est néanmoins intéressant de noter que la classe $z=2$ a systématiquement des valeurs de PC1 inférieures d'environ 0,5-1 aux valeurs correspondant à des ions trichargés ($z=3$).

2.4 Conclusions

L'application de la quantification des ions oxoniums pour l'identification de glycopeptides tryptiques de la hCG a été effectuée pour la première fois. Cette approche a permis de confirmer la possibilité d'utiliser les ions oxoniums pour séparer les N-glycopeptides complexes, hybrides, et *high-mannose*. En allant plus loin dans la démarche, la séparation des différents états de mannosylation des N-glycopeptides *high-mannose* et des formes complexes A1G1 et A2G2 a également été mise en évidence.

L'impact non négligeable de la séquence du peptide porteur d'un N-glycane donné sur l'abondance des ions oxoniums présents en MS/MS a été prouvée. Néanmoins, puisque l'échantillon est d'origine biologique, il est possible que cette variation soit due à des différences stéréochimiques que les appareils de spectrométrie de masse ne permettent pas d'élucider sans l'apport de la mobilité ionique.

E. Synthèse et Discussion

Les gonadotrophines chorioniques sont des protéines d'intérêt majeur, dans les domaines de la reproduction (assistée ou non) ou du dopage notamment. La structure de ces protéines, fortement glycosylées, n'est pas suffisamment caractérisée du fait du manque d'outils performants pour l'étude de glycoprotéines très hétérogènes. Pourtant, la connaissance de ces modifications post-traductionnelles pourrait permettre d'expliquer les différences d'efficacité de traitements de la fécondité chez certaines femmes. De tels outils pourraient également être utilisés dans la lutte anti-dopage pour identifier l'utilisation d'hCG d'origine exogène à des fins d'augmentation de performance physique.

Des travaux préliminaires (147) ont permis de mettre en évidence l'intérêt d'utiliser des méthodes de protéolyse alternative pour augmenter la résolution des différents sites de glycosylation, et ainsi perfectionner la caractérisation de la N-glycosylation de la hCG. Ces travaux ont évolué vers une réflexion quant à la possibilité de caractériser la O-glycosylation de cette hormone par des approches de protéomique *bottom-up* originales.

L'utilisation de la pronase a d'abord été discutée, puisque certains travaux antérieurs ont montré qu'elle permet la déplétion des peptides non glycosylés, réduisant ainsi la compétition de ces derniers avec les glycopeptides d'intérêt lors du processus d'ionisation. L'utilisation de ces protéases était notamment d'un grand intérêt pour l'étude de la hCG du fait du rapprochement de ses sites de O-glycosylation. La pronase a immédiatement permis la production de nouveaux glycopeptides d'intérêt, mais ces derniers étaient parfois difficilement identifiables du fait de la faible qualité de leurs spectres de fragmentation. Cette limitation a en partie été surmontée par l'utilisation d'une méthode de fragmentation alternative (*stepped-HCD*). Les fragmentations *stepped-HCD* ont apporté des améliorations sur trois points, l'identification de nouveaux N-glycopeptides tryptiques, la détection et l'identification de O-glycopeptides portant un unique site de glycosylation, et l'augmentation de la qualité des spectres de fragmentation des O-glycopeptides et des identifications (Troisième Partie, Protéolyse aspécifique et fragmentations

stepped-HCD pour la comparaison de deux hCG d'origines géographiques distinctes).

Ces approches ont été appliquées à deux échantillons d'hCG urinaires purifiés à partir d'urines de femmes caucasiennes ou asiatiques. Il existe peu de travaux sur la caractérisation de la glycosylation entre différentes populations géographiquement distinctes. Les résultats présentés dans ce manuscrit sont donc très novateurs, dans le sens où ils ont montré pour la première fois que la glycosylation de la hCG semblait varier entre les femmes de différentes origines. Ces travaux pourraient être une base de travail pour la recherche de glycoprotéofomes ayant une activité spécifique (rapidité d'action, déclenchement de la cascade de signalisation, durée de vie dans la circulation, *etc.*) selon les patientes. Il convient également de considérer des approches quantitatives et statistiques sur de plus grandes cohortes d'échantillons, afin notamment d'évaluer l'impact d'autres facteurs physiologiques sur l'état de glycosylation de la hCG.

Les améliorations ont ici porté principalement sur les méthodes de protéolyse et de spectrométrie de masse. D'autres améliorations peuvent être envisagées, notamment l'utilisation de méthodes de fragmentation alternatives proposées sur les spectromètres les plus récents, en particulier la fragmentation EThcD (175,193). Cette dernière permet la production d'ions très informatifs sur la localisation des sites de glycosylation et sur la structure de ces modifications. La fragmentation EThcD consistant en une fragmentation à base de transfert d'électron (ETD) suivie d'une fragmentation collisionnelle HCD, la variation des énergies de collision utilisées (« *stepped-EThcD* ») pourrait permettre d'améliorer d'avantage la qualité des identifications. Une limitation de cette technique de fragmentation reste cependant la relative lenteur des modes de fragmentation impliquant des interactions avec des électrons. Pour des applications de glycoprotéomique, il a été entrepris de ne déclencher les fragmentations EThcD que lorsqu'une première fragmentation HCD produit des ions oxoniums, rapporteurs de la glycosylation du précurseur (194). Le

temps de cycle ainsi libéré pourrait permettre de réduire le temps global d'une analyse LC-MS, ou d'améliorer la sensibilité de la détection en augmentant le temps d'accumulation des ions.

La digestion par des enzymes peu ou non spécifiques produit des glycopeptides dont la partie peptidique peut être de petite taille, jusqu'à un unique acide aminé. L'utilisation de méthodes d'enrichissement et de séparation plus adaptées à ces glycopeptides, parfois trop hydrophiles pour être séparés en RPLC sur phase inverse C18, peut également être une piste d'amélioration des stratégies de glycoprotéomique *bottom-up*. Le couplage de plusieurs méthodes séparatives, déjà souvent utilisé pour le préfractionnement d'échantillons destinés à des analyses à l'état intact (*top-down*), est tout à fait préconisé pour l'analyse *bottom-up* de glycoprotéines multi-glycosylées. En effet, la très forte hétérogénéité des différentes glycoprotéoformes nécessite des outils de séparation performants pour la caractérisation de molécules ne variant parfois que par la position d'un monosaccharide sur le glycane. Notamment, le couplage de la chromatographie PGC à la RPLC-C18 semble très prometteuse, puisque les glycopeptides non retenus sur cette dernière sont ensuite injectés sur une phase graphitique permettant leur séparation. Ce couplage est également plus simple à mettre en place que d'intégrer la chromatographie HILIC du fait de la composition des phases mobiles utilisées.

Une des dimensions de séparation potentiellement la plus prometteuse pour le futur de la glycoprotéomique est la spectrométrie de mobilité ionique. Cette technologie n'est pas récente mais de récents développements ont conduit à son intégration dans les appareils commerciaux de dernière génération (par exemple dans le Synapt G2S de Waters, le FAIMS pour le Fusion Eclipse de ThermoScientific, le TIMS pour le timsTOF fleX de Bruker Daltonics). La mobilité ionique permet de discerner les glycoformes isomères et de caractériser l'orientation des liaisons osidiques, déterminante pour la connaissance de l'encombrement stérique des glycanes.

Cette séparation s'avèrera sûrement utile pour améliorer l'identification des glycopeptides en utilisant la signature des ions oxoniums produits en MS/MS. Cette approche, développée dans ce manuscrit (Troisième Partie, Analyse approfondie des données spectrales pour l'identification de glycoformes minoritaires), a mis en évidence la très forte corrélation entre l'intensité de certains fragments oxoniums et l'identité du glycopeptide fragmenté. Les différentes classes de N-glycopeptides (complexes, hybrides, et *high-mannose*) ont d'abord été très clairement clustérisées, en n'utilisant uniquement les informations spectrales relatives aux ions oxoniums et sans utiliser les fragments glycopeptidiques Y produits en HCD. De plus, l'impact de plusieurs propriétés du glycopeptide sur l'abondance des oxoniums (séquence du peptide, état de charge) a été mise en évidence. Une analyse plus approfondie des données a également permis de montrer qu'une séparation fine des glycopeptides était possible en fonction de la composition exacte de ces derniers.

Cette approche a pour but notamment d'identifier les glycopeptides dont la fragmentation n'est pas suffisamment informative, souvent du fait de la faible concentration du glycopeptide considéré. S'il convient de procéder à des mesures de la sensibilité de l'identification par cette méthode, elle semble être une piste intéressante pour la caractérisation des glycoprotéofomes minoritaires des glycoprotéines hétérogènes.

L'extension de la méthode à des glycopeptides plus variés doit également être envisagée. En particulier, les O-glycopeptides et les glycanes portant des résidus NeuAc ou Fuc mériteraient d'être intégrés à cette approche. Dans le futur, la fragmentation des glycoprotéines végétales ou de mammifères (non-humains) pourraient également être étudiée, et l'impact de certains monosaccharides spécifiques évalué (xylose, acide N-glycolylneuraminique, *etc.*).

Les méthodes développées dans la troisième partie de ce manuscrit visent à identifier des glycopeptides, et donc d'envisager le rôle biologique de la

glycosylation comme étant dépendant de la structure de glycanes ponctuels. Il faut cependant également considérer la glycosylation comme un phénomène global ayant un effet à l'échelle des différentes glycoprotéofomes d'une protéine. Les approches de glycoprotéomique *top-down* et de glycomique pourraient permettre d'apporter des réponses à ces problématiques. Si la première approche n'a pas été développée au cours de ce projet, le développement d'appareils à très haute résolution, rapides et provoquant une fragmentation efficace comme ceux par exemple issus de la technologie Orbitrap, doit permettre le développement de l'étude de glycoprotéines hétérogènes à l'échelle intacte. Ces techniques sont toutefois encore trop immatures à ce jour pour pouvoir caractériser toutes les glycoprotéofomes d'un échantillon biologique complexe. En effet le traitement des données demeure fastidieux et totalement manuel dans bien des cas.

L'analyse des glycanes indépendamment des protéines peut être une approche alternative efficace pour caractériser l'ensemble des (N-)glycanes d'un échantillon. La méthodologie développée dans la deuxième partie de cet ouvrage a permis d'augmenter de façon importante la sensibilité d'une méthode de N-glycomique, utilisant notamment une méthode de dérivation des glycanes sur cible MALDI. Si dans le cadre de cette étude il s'est simplement agi de réaliser des cartographies MS (en s'aidant de la MS/MS pour les identifications), il est possible de réaliser des études de fragmentation assez spécifiques (195).

Par ailleurs, la perte de sensibilité associée à la manipulation de l'échantillon peut être réduite grâce à l'utilisation de puce microfluidique. La fabrication de réacteurs miniaturisés permettant d'intégrer totalement le traitement de l'échantillon à un *workflow* ne nécessitant pas de pipetage ou précipitation peut grandement augmenter la quantité de matériel analysé. L'intégration de méthodes de séparation à un tel *workflow* (LC-MALDI) permet également d'augmenter de façon importante la sensibilité de la détection, notamment en réduisant la compétition au niveau du processus d'ionisation et en limitant la

saturation des détecteurs par les espèces majoritaires (196). Ces futures approches sont notamment envisagées sur les solutions technologiques récentes telles que le timsTOF fleX qui propose une source MALDI couplée à un tube de mobilité ionique et un analyseur Q-TOF de dernière génération.


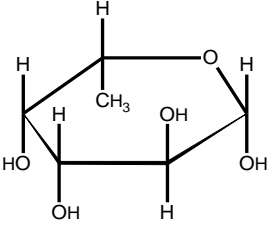

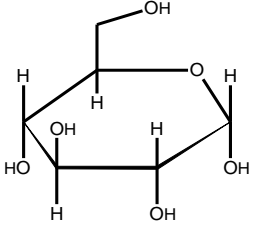

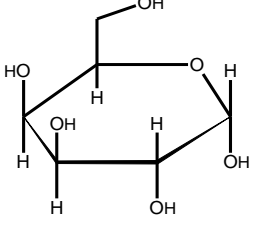

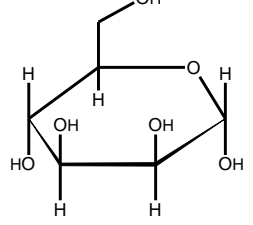
Les approches de glycoprotéomique au sens large que sont la glycomique et la glycoprotéomique *bottom-up* se sont montrées également informatives pour la caractérisation de la glycosylation des protéines, mettant en lumière leur forte complémentarité. Elles ouvrent des perspectives majeures pour la compréhension des systèmes biologiques dans l'ère post-protéomique, c'est à-dire l'ère de l'étude fonctionnelle des protéoformes.


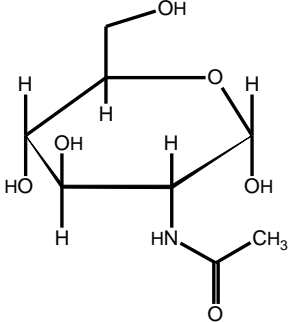

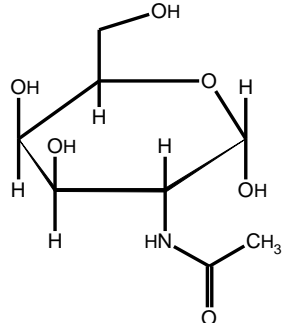

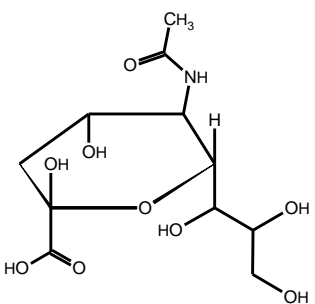
F. Annexes

1. Monosaccharides et glycanes

Les monosaccharides qui constituent les glycanes les plus courants chez les mammifères sont représentés dans la table ci-dessous. Leur nom complet et abrégé y est indiqué, ainsi que leur structure SNFG, leur structure chimique détaillée, et leur masse lorsqu'ils sont intégrés à un glycane (pour calculer la masse du monosaccharide libre en solution, il faut ajouter la masse de la molécule d'eau perdue lors de la formation de la liaison osidique).

Tableau 9 : Table récapitulative des structures monosaccharidiques des glycoprotéines humaines.

Monosaccharide	Nom abrégé	Masse de l'adduit (Da)	Structure SNFG	Structure détaillée
Fucose	Fuc	146,0579		
Glucose	Glc (Hex)	162,0528		
Galactose	Gal (Hex)	162,0528		
Mannose	Man (Hex)	162,0528		

<p style="text-align: center;">N-Acétylglucosamine</p>	<p style="text-align: center;">GlcNAc</p>	<p style="text-align: center;">203,0794</p>		
<p style="text-align: center;">N-Acétylgalactosamine</p>	<p style="text-align: center;">GalNAc</p>	<p style="text-align: center;">203,0794</p>		
<p style="text-align: center;">Acide N-acétylneuraminique</p>	<p style="text-align: center;">NeuAc</p>	<p style="text-align: center;">291,0954</p>		

2. Spectrométrie de masse

2.1 Notions en MS

Etat de charge :

L'état de charge, ou « z », indique le nombre total de charges stabilisées sur une molécule (protéine, (glyco)peptide, sucre, *etc.*). Il peut être positif ou négatif selon les conditions expérimentales, et il convient d'indiquer le signe de « z » dans le cas de travaux utilisant les deux polarités. Le signe positif est le plus souvent omis dans le cadre de la protéomique car, sauf cas particuliers, ces ions sont souvent analysés en mode positif.

Isotopes, massif isotopique et masse monoisotopique :

Les atomes existent sous différentes formes, appelées « isotopes », ayant en commun un même nombre de protons, mais qui diffèrent par le nombre de neutron contenus dans leur noyau. La plupart des isotopes stables (non-radioactifs) d'un atome donné ont des propriétés physico-chimiques quasi-identiques. Ces isotopes sont présents à différentes proportions à l'état naturel (par exemple, le ^{12}C représente environ 98,9% du Carbone total, le ^{13}C environ 1,1%, et le ^{14}C est présent à l'état de traces), et constituent toutes les molécules.

En spectrométrie de masse, l'ensemble des signaux (« pics ») d'une molécule et des molécules ne variant que par les isotopes de ses atomes est appelé **massif isotopique**. L'écart entre chaque pic du massif isotopique correspond, dans le cas de molécules organiques, à la masse d'un neutron supplémentaire soit environ 1 uma, où 1 uma est défini comme 1/12 de la masse de l'isotope ^{12}C du carbone. La masse d'un neutron est 1,00866491595(49) uma, la masse du proton (1Da) étant légèrement inférieure à 1,007276466583(32) uma avec masse du proton = $1,6726 \cdot 10^{-27}$ kg, masse du neutron = $1,6749 \cdot 10^{-27}$ kg, masse de l'électron = $9,1094 \cdot 10^{-31}$ kg.

La masse d'une molécule constituée uniquement des isotopes les plus légers de ses constituants (^{12}C , ^1H , ^{16}O , ^{14}N , *etc.*) est sa **masse monoisotopique**. Il s'agit

donc, en spectrométrie de masse, de la masse du premier pic (isotope le plus léger) d'un massif isotopique. Certaines atomes (métaux) ont des distributions isotopiques particulières, avec des écarts de deux neutrons entre leurs isotopes stables. L'isotope majoritaire de ces atomes n'est pas toujours l'isotope le plus léger.

Résolution :

La résolution est définie comme l'écart minimum entre deux signaux pour que ceux-ci soient résolus en spectrométrie de masse. Dans la pratique, on utilise plutôt la notion de pouvoir résolutif à mi-hauteur, qui correspond au rapport $\frac{m}{\Delta m}$, avec « m » la masse de l'ion considéré et « Δm » la largeur de ce pic à mi-hauteur.

Précision de masse :

La précision d'un spectromètre de masse, ou plus généralement d'une mesure, qualifie sa faculté à fournir une valeur expérimentale proche de la valeur théorique vraie. La précision est une combinaison de la **fidélité** et de la **justesse**.

La **fidélité** est définie comme « l'étroitesse d'accord entre des résultats d'essai indépendants obtenus sous des conditions stipulées. » (ISO 5725), soit la capacité à fournir des valeurs expérimentales proches pour des mesures répétées, et donc avec une erreur aléatoire faible. La notion de fidélité ne peut en aucun cas décrire que la valeur mesurée soit ou ne soit pas proche de la valeur vraie.

La **justesse** est quant à elle définie par la norme ISO 5725 comme « l'étroitesse de l'accord entre la valeur moyenne obtenue à partir d'une large série de résultats d'essais et une valeur de référence acceptée », soit la faible différence entre la moyenne des mesures et la valeur vraie. L'erreur aléatoire n'entre donc pas en compte dans le calcul de la justesse. C'est souvent la justesse qui est calculée en spectrométrie de masse, sous la forme d'une erreur absolue donnée

en Daltons, en millième d'unité de masse (mmu), *etc.* ou d'une erreur relative donnée en pourcentage, en partie par million (ppm), *etc.*

Sensibilité :

La **sensibilité** en spectrométrie de masse correspond à la quantité minimale détectable par un appareil. Cette valeur est dépendante de tous les éléments du spectromètre de masse et des méthodes séparatives qui lui sont couplées, mais également de la molécule considérée, notamment du fait des différences d'efficacité d'ionisation entre des molécules distinctes.

Gamme de masse :

La **gamme de masse** (de m/z) est l'étendue des valeurs de m/z qu'il est possible de détecter en une unique acquisition. Elle dépend de l'analyseur et du détecteur du spectromètre de masse. La nécessité de pouvoir disposer d'une grande gamme de masse est essentielle en MALDI puisque la plupart des ions sont monochargés, tandis que l'ionisation ESI permet la formation d'ions multichargés et avec des valeurs de m/z plus faibles.

Gamme dynamique :

La **gamme dynamique** correspond à l'étendue de concentrations relatives qu'il est possible de détecter simultanément. Elle représente la capacité d'un appareil à détecter une espèce faiblement abondante malgré la présence concomitante d'une autre molécule plus abondante. C'est une caractéristique importante car les échantillons biologiques sont souvent constitués de mélanges de protéoformes d'abondances variées. Les concentrations des protéines humaines plasmatiques ont par exemple une étendue d'un facteur parfois supérieur à 10^{10} , l'albumine représentant plus de la moitié de la concentration massique du sérum (>30 mg/mL) tandis que certaines interleukines peuvent être présentées à de très faibles quantités (<10 pg/mL)(197). De plus, les glycopeptides produits par protéolyse sont produits en quantité souvent bien inférieure aux peptides non glycosylés issus de la même protéine. Ces derniers

ayant également une efficacité d'ionisation supérieure, la gamme dynamique des appareils est une caractéristique critique en glycoprotéomique.

Il est également possible de diminuer le besoin d'une gamme dynamique conséquente en couplant la spectrométrie de masse à des méthodes séparatives et de purification (chromatographie liquide, mobilité ionique, *etc.*).

Durée de cycle, d'accumulation, de transient :

Un cycle d'analyse en spectrométrie de masse correspond à tout ou partie des étapes : de sélection (quadrupôle, piège à ions), d'accumulation (piège à ions, *C-trap*), de fragmentation (piège à ions, cellule de collision), de transfert (quadrupôle, multipôles) et d'analyse de masse des ions (tube à temps de vol, *orbitrap*, cellule FTICR, *etc.*). En mode DDA TopN (acquisition MS et fragmentation des N ions les plus intenses), **le temps de cycle** correspond donc au temps nécessaire à la production d'un spectre MS et de N spectres MS/MS. Dans les appareil hybrides de type QqOrbitrap, les ions sont accumulés dans une *C-trap* pendant le temps nécessaire à atteindre un nombre suffisant (**temps d'accumulation**, ou *fill time*). Ils sont ensuite injectés dans l'analyseur principal (Orbitrap) dans lequel ils orbitent autour d'une électrode centrale, le signal électronique étant acquis pendant une période appelée « **transient** ».

2.2 Source d'ionisation

(nano-)ESI

L'ionisation par électro-nébulisation est une technique d'ionisation douce qui permet de rendre des macromolécules chargées en évitant toutefois leur fragmentation. Cette méthode consiste à appliquer une tension à un liquide pour le nébuliser et former un spray. Elle peut donc être directement couplée à des méthodes de séparation et d'injection en phase liquide (chromatographie liquide, électrophorèse capillaire, *etc.*). Les ions produits peuvent être formés de plusieurs charges.

Le liquide contenant les molécules ionisables est souvent au moins partiellement constitué d'un solvant volatil (acétonitrile, méthanol, *etc.*). Pour augmenter la densité de charge, il est souvent enrichi en ions H⁺ (dans la plupart des études de protéomique) par l'ajout d'acides organiques volatils. La désolvatation des molécules peut être favorisée par l'application d'un flux de gaz neutre et le chauffage du flux de liquide. Ce dernier est soumis à une tension électrique (généralement de 1 kV à 5 kV), afin de former un spray d'ions.

Les nano-gouttelettes formées sont constituées d'ions, dont la densité augmente au fur et à mesure de l'évaporation des molécules de solvant. L'accumulation des charges à la surface des gouttelettes augmente jusqu'à ce qu'elles ne puissent plus supporter l'augmentation de la tension superficielle (limite de Rayleigh). Les dislocations successives des gouttelettes ayant accumulé un nombre important de charges du fait de la répulsion coulombienne entre celles-ci (explosion coulombienne) peuvent se succéder jusqu'à ce qu'une unique molécule ionisée soit isolée en phase gazeuse (**modèle de la charge résiduelle**). Un second modèle consiste en l'éjection d'ions à partir des nano-gouttelettes en cours d'évaporation, ou **modèle de l'évaporation ionique**. Ces modèles semblent co-exister et surviennent probablement à différentes proportions en fonction de la nature et de la taille de la molécule captant les charges.

La miniaturisation de la source ESI (nano-ESI ou NSI) permet, en diminuant les volumes et les débits manipulés, et donc en formant un spray plus petit et en le rapprochant de l'entrée des appareils, d'augmenter fortement la quantité d'ions transférés en phase gazeuse. Ces modifications permettent d'augmenter la sensibilité des méthodes de spectrométrie de masse.

MALDI

L'ionisation et la désorption Laser Assistée par Matrice (ou *Matrix-Assisted Laser Desorption/Ionization*) est une méthode d'ionisation également très utilisée pour la spectrométrie de masse, en particulier pour l'analyse de biomolécules.

Les échantillons sont analysés à l'état solide, co-cristallisés sur une surface plane avec des molécules aromatiques, la matrice. La désorption et l'ionisation s'effectue au moyen d'un Laser pulsé (émettant généralement des radiations UV) qui excite la matrice, dont la longueur d'onde d'absorption doit correspondre à celle du faisceau d'ions. L'excitation photonique de la matrice conduit à l'augmentation de la pression dans le dépôt solide qui conduit à l'ablation de la région irradiée et la production d'un plumeau de matrice et d'analytes, à l'état gazeux, neutres ou protoné(e)s. La dislocation de ces agrégats a lieu progressivement, conduisant à la « désolvatation » complète des analytes, ainsi dépourvus de molécules de matrice. Une désolvatation incomplète conduit quant à elle à des adduits de matrice sur les molécules d'intérêt.

Des échanges de protons peuvent avoir lieu entre la matrice et les biomolécules d'intérêt, mais ceux-ci sont généralement limités, conduisant à des états de charge bien moindres en MALDI qu'en ESI par exemple, avec la formation d'ions monochargés dans le cas de peptides protéolytiques (de masse inférieure à 5 kDa) et des états de charge typiquement inférieurs à 10 pour des protéines de 60 kDa.

2.3 Analyseur

TOF

Un analyseur à Temps de Vol (ou *Time Of Flight*, TOF) est constitué d'une région d'accélération des ions (source) et d'une région libre de champ (tube de vol). Les ions sont d'abord accélérés dans un champ électrique uniforme et acquièrent une énergie cinétique identique et calibrée. Les ions entrent dans la région libre de champ du tube de vol (sous vide) à une vitesse constante et inversement proportionnelle à la racine carrée de leur rapport m/z . La mesure

précise de la distance parcourue par les ions (longueur du tube du vol) et du temps nécessaire à la détection du signal permet le calcul de leur vitesse et donc de leur rapport m/z .

L'énergie acquise par les ions est théoriquement uniforme, mais dans les faits les ions ont initialement une énergie variable (du fait de l'excitation variable des molécules de matrices par le Laser selon leur profondeur ou leur position excentrée, par exemple). Pour compenser la distribution d'énergie au sein des ions de mêmes m/z , l'utilisation d'un miroir électrostatique (ou « réflectron ») permet, en appliquant un champ électrique de sens opposé au mouvement des ions, de compenser les disparités d'énergie cinétique initiales et d'augmenter la distance parcourue par les ions afin d'augmenter la résolution de l'appareil.

Cet analyseur est fréquemment couplé au mode d'ionisation MALDI car sa vitesse d'acquisition est faible et compatible avec la rapidité de ce mode d'ionisation, et parce que l'absence de limite maximale de m/z détectable est élevée et permet de travailler avec les ions à faible état de charge produits en MALDI.

QqOrbitrap

L'analyseur hybride QqOrbitrap est constitué de deux analyseurs distincts que sont le quadropôle (Q) et l'Orbitrap. Il est de plus en plus répandu dans les laboratoires de spectrométrie de masse et de protéomique pour sa capacité à analyser les peptides avec de bonnes performances (vitesse, sensibilité, précision de masse, résolution).

L'analyseur quadripolaire est constitué de quatre électrodes parallèles et disposées autour de l'axe de déplacement des ions « z ». Les barreaux, de section circulaire ou hyperbolique, sont parcourus par des courants électriques constants, de signe identique deux à deux. A ce champ électrique constant est ajoutée une composante variable qui provoque des variations du champ dans un axe perpendiculaire au mouvement des ions. Certains ions sont alors déviés et éjectés du quadropôle en fonction de leur m/z et des valeurs de tension

appliquées. Pour un champ donné, seuls les ions ayant une valeur de m/z permettant leur entrée en résonance sont transmis et détectés. En balayant différentes tensions, le quadripôle peut analyser une importante gamme de m/z . En appliquant une tension de composante constante nulle, le quadripôle agit comme un guide d'ions et permet la transmission de tous les ions (de m/z supérieur au seuil minimal imposé par le champ variable).

Cet analyseur primaire est couplé à un Orbitrap par l'intermédiaire d'une cellule d'accumulation des ions (*C-trap*). Une fois la quantité d'ions désirée atteinte dans la *C-trap*, ces derniers sont injectés dans l'Orbitrap. Il est constitué d'une électrode oblongue centrale et d'une électrode « gaine » entre lesquelles les ions orbitent. La fréquence d'oscillation sur l'axe longitudinal est directement reliée à la valeur de m/z des ions, tandis que les oscillations radiales et la vitesse de rotation autour de l'électrode centrale en sont indépendantes. Cette fréquence est mesurée par le courant image induit que les ions produisent sur deux électrodes disposées autour de l'Orbitrap. La transformée de Fourier du signal ainsi enregistré permet de le décomposer en signaux périodiques, leur fréquence étant convertie en mesure de m/z .

2.4 Modes de fragmentation (MS/MS)

CID

La fragmentation induite par collision (*Collision-Induced Dissociation*, CID) est une méthode de fragmentation couramment utilisée en protéomique. Elle consiste à accélérer les ions au sein d'un vide modéré afin de favoriser les collisions entre ces derniers et les molécules de gaz neutre (air, N₂, He...). L'énergie cinétique transférée aux ions peut aboutir à une énergie vibrationnelle suffisante pour provoquer la dissociation des ions. Les peptides se fragmentent alors généralement au niveau des liaisons peptidiques, mais peut également aboutir à des pertes de neutres (H₂O, NH₃) ou de groupements labiles (glycanes, *etc.*).

HCD

Le mode de fragmentation HCD (pour *Higher-energy Collisional Dissociation*) est proche de la fragmentation CID, à la différence qu'elle a lieu dans une cellule de fragmentation multipolaire distincte et dédiée. Elle permet la production et le transfert d'ions sans limite basse de m/z , contrairement aux fragmentations CID effectuées dans des pièges ioniques, permettant notamment la détection de tags isobares (TMT) ou d'ions oxoniums.

Article 1

La mise en place des stratégies de bottom-up pour l'étude hCG a été réalisée sur des standards. Cette étude pour la caractérisation de la glycosylation de médicaments à base d'hormone hCG d'origines recombinante (r-hCG) ou urinaire (u-hCG) a été menée en collaboration avec le Laboratoire de Sciences Analytiques, Bioanalytiques et Miniaturisation (LSABM). Nos collaborateurs développent des réacteurs enzymatiques miniaturisés (IMER) de trypsine et de pepsine et ont pu ainsi étudier leurs performances.

Ont notamment été comparées les glycopeptides identifiés après une digestion sur IMER ou en solution, en utilisant les mêmes enzymes. Cette comparaison a montré la complémentarité des deux approches et l'identification de glycopeptides exclusifs dans chacune d'elles. La complémentarité de différentes protéases a été mise en évidence pour la caractérisation de la N-glycosylation de la hCG, en solution et sur IMER.

Bien que ne remplaçant pas la digestion en solution « classique », l'utilisation de tels réacteurs permet notamment un gain de temps important, en réduisant le temps d'incubation de l'échantillon et des protéases de 15 heures (digestion sur la nuit) à 30 minutes. De plus, la réutilisabilité des IMER a été mise en évidence, et la possibilité de les intégrer à un système LC-MS/MS pourrait permettre de réduire la perte de matériel et de sensibilité liée à la manipulation des échantillons.

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 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, Elsevier, 2020, 206, pp.120171. 10.1016/j.talanta.2019.120171. hal-02330897

HAL Id: hal-02330897

<https://hal.archives-ouvertes.fr/hal-02330897>

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 Perchepped^{1,*}, Nicolas Eskenazi^{2,*}, 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 posttranslational 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 Nglycan), 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 obtain 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 Nglycopeptides 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 Nglycosylation 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 (PODN86, PODN87, 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 μm , 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 µL of supernatant were introduced in a well of a 96-microwell plate. Then, 100 µL 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 UVspectrophotometer (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 µL 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 µg 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 µg of hCG were introduced in a 5 µL 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 µL min⁻¹. The digestion was carried-out during 30 min at 37°C using a « stop-flow » approach. Finally, 100 µL 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 µL 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 µL min⁻¹ in buffer C: 0.05 % TFA in H₂O/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 Å, ThermoFisher Scientific) at a flow rate of 220 nL min⁻¹ with a gradient composed of A (0.1 % FA in H₂O/ACN 98/2 (v/v)) and B (0.1 % FA in H₂O/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. Byonic™ score threshold used for the validation of the glycopeptide identification.

Enzyme	Glycosylation site	Byonic™ 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 ^βN₁₃ (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].

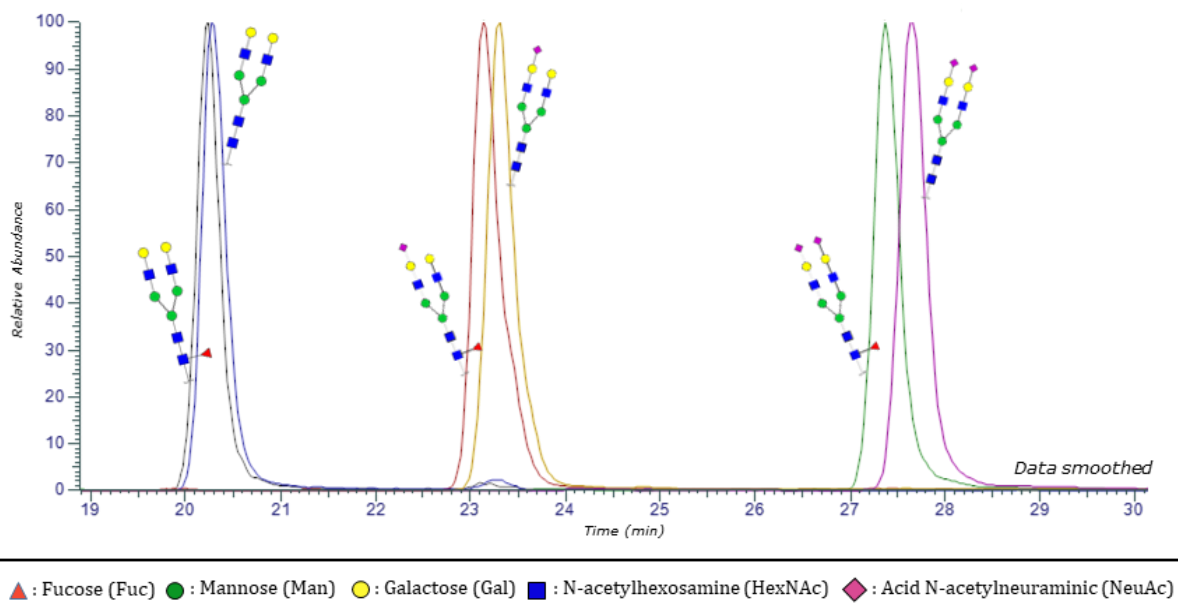


Figure 1. Extracted Ion Chromatograms of glycoforms of CRPIN₁₃ATLAVEK sequence (⁶N₁₃) produced from r-hCG digestion with T-IMER and analysis by LC-MS/MS exhibiting the influence of sialic acids and fucose on the retention time of the glycopeptide (only its glycan moiety was represented).

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.

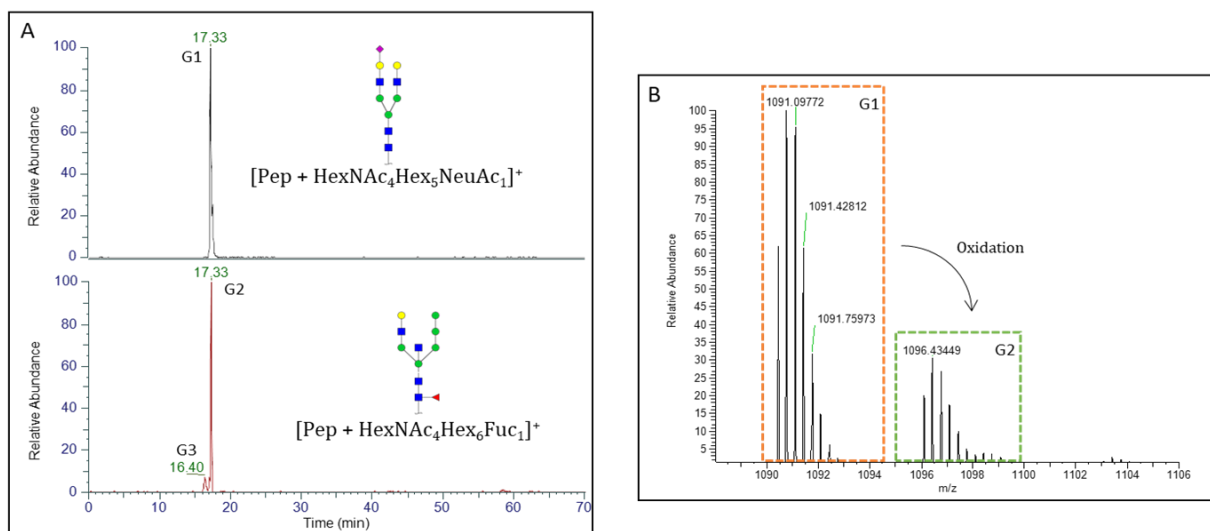


Figure 2. (A) Extracted Ions Chromatograms of two glycopeptides identified on α N52 by ByonicTM software obtained by the nanoLC-MS/MS analysis of the T-IMER digest of r-hCG. (B) Isotopic MS profile combined from 16.08 to 16.6 min.

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 Byonic™. For instance, a glycan composition corresponding to HexNAC₄Hex₇ on the peptide containing ^αN₅₂ 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 Byonic™, 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 ^αN₇₈ glycosylation site due to potential steric hindrance of the carbohydrate moiety. Moreover, the identification of glycopeptides from β N₃₀ 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 ^αN₅₂ was fucosylated while

glycopeptides from ^{13}N 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 PIN₁₃ATLAVEK (^{13}N) were also identified on CRPIN₁₃ATLAVEK (^{13}N). 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 LGEDNINVEGNEQFISASK) 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 $^{\alpha}N_{52}$. It is worthwhile to notice that both digestion modes involving trypsin led to identified glycans only on $^{\alpha}N_{52}$ and $^{\beta}N_{13}$ glycosylation sites and no glycan identified on $^{\alpha}N_{78}$ and $^{\beta}N_{30}$. 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.

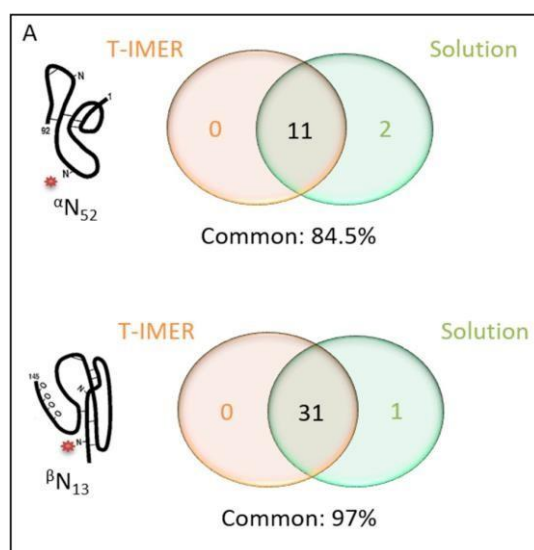


Figure 3. Comparison of the numbers of N-glycans identified on a specific glycosylation site after the digestion of r-hCG with trypsin-based IMER and in solution.

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 ($^{\alpha}N_{78}$ and $^{\beta}N_{30}$). 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, PIMER 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 $^{\alpha}N_{52}$ and $^{\beta}N_{13}$, 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 $^{\alpha}N_{78}$ and $^{\beta}N_{30}$, 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 .

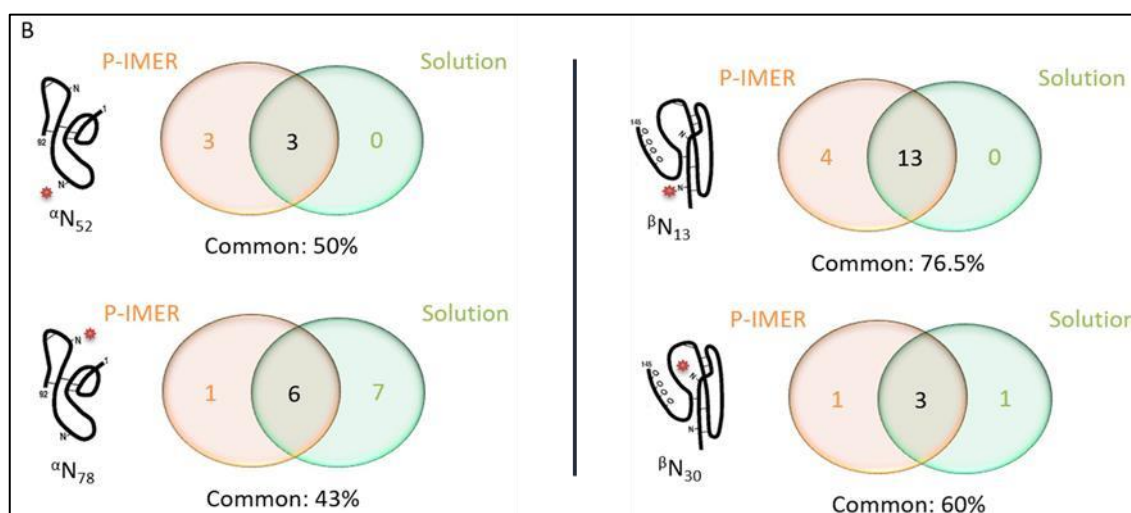


Figure 4. Comparison of the numbers of N-glycans identified on a specific glycosylation site after the digestion of r-hCG with pepsin-based IMER and in solution.

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 $^{\alpha}\text{N}_{52}$, whereas the digestion in solution led to seven additional glycans on $^{\alpha}\text{N}_{78}$. 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.

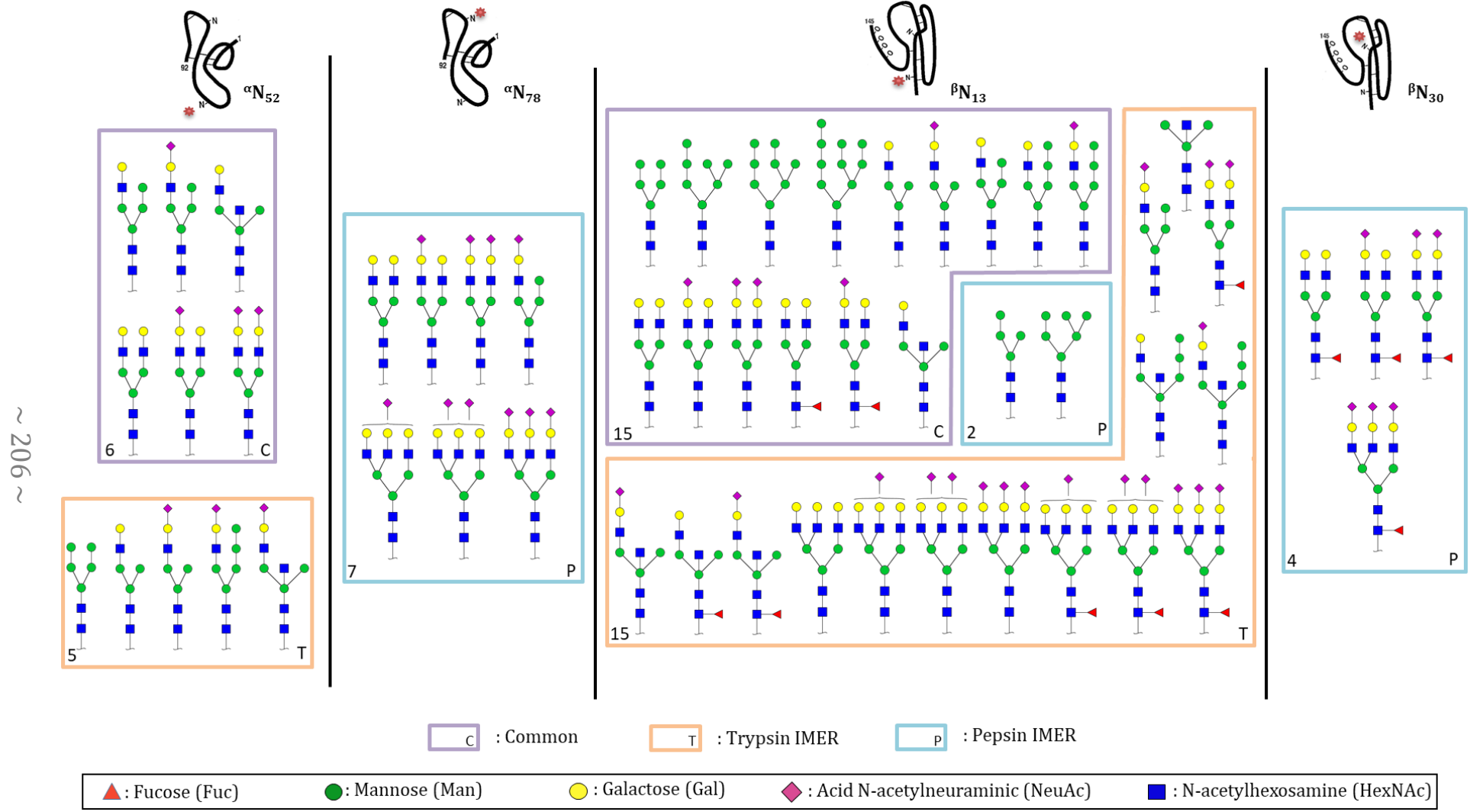


Figure 5. Glycans identified on glycopeptides after digestions on trypsin- and pepsin-based IMERs of r-hCG and analysis by LC-MS/MS. (Common identified Glycans between both enzymes: 39%).

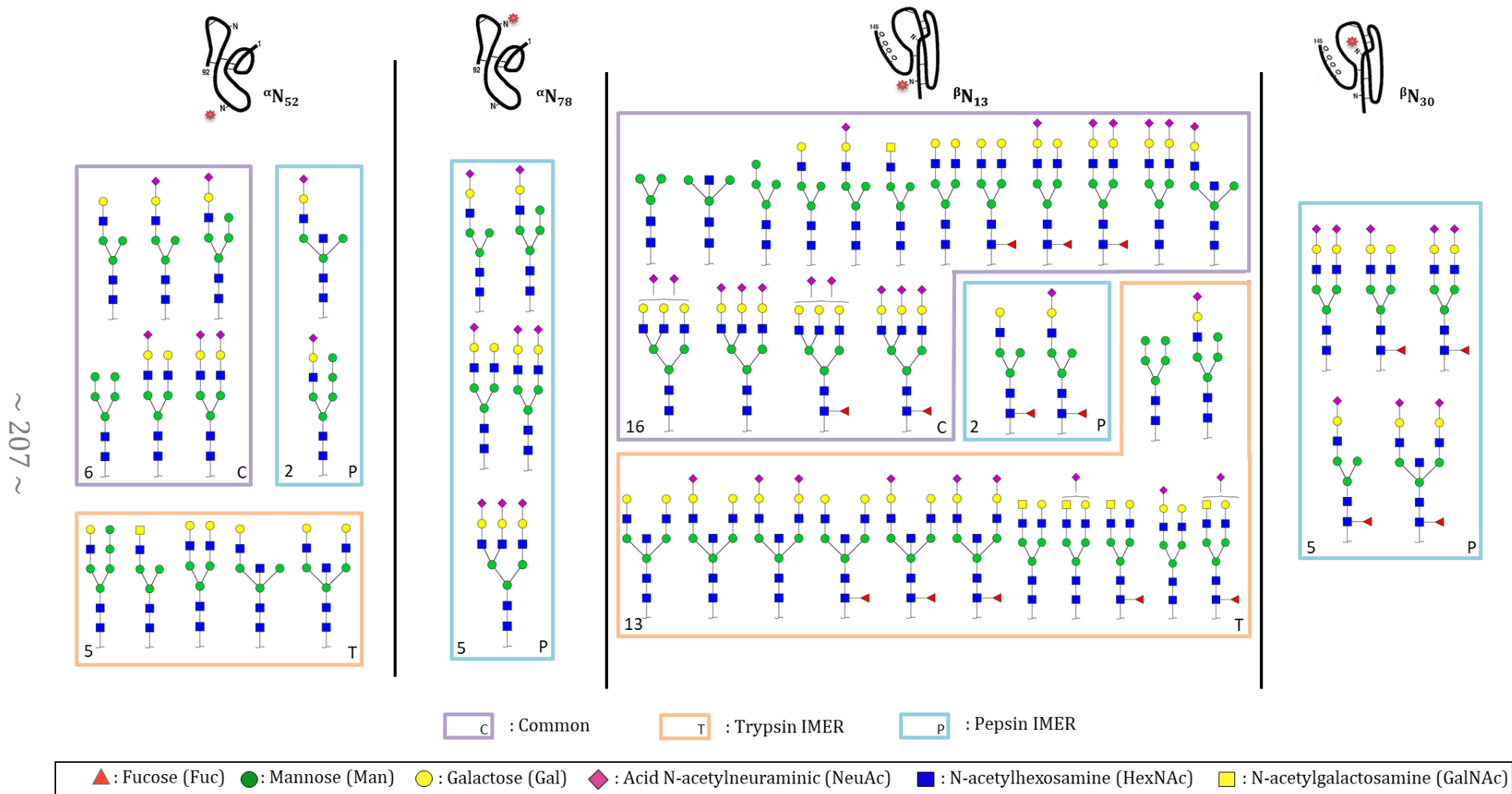


Figure 6. Glycans identified on glycopeptides after digestions on trypsin- and pepsin-based IMERs of u-hCG and analysis by LC-MS/MS. (Common identified N-glycans between both enzymes: 41%).

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 $^{\alpha}\text{N}_{78}$ and $^{\beta}\text{N}_{30}$. 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 $^{\beta}\text{N}_{30}$ 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 $^{\beta}\text{N}_{13}$, 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.

$\alpha N52$	r-hCG	u-hCG	$\beta N13$	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

$\alpha N78$	r-hCG	u-hCG	$\beta N30$	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-10LABX-31), of PSL University, and the Conseil Régional d'Île-de-France.

REFERENCES

- [1] O. Nørregaard Jensen, Modification-specific proteomics: characterization of posttranslational 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 SiteSpecific Location, Identification and Quantification of N-/O- Glycosylation in Human Chorionic Gonadotropin Drug Products, *AAPS J.* 19 (2017) 846–855. doi:10.1208/s12248017-0062-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. Fidler, 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 β -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 Dorselaer, 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.

Article 2

Les cartographies des glycanes sont un outil analytique intéressant de façon très large de nombreux domaines. Dans le cadre d'un projet initié par l'Institut Pasteur de Nouvelle-Calédonie (Unité de Recherche et d'Expertise sur la Leptospirose), la composition d'un biofilm produit par *Leptospira interrogans* a été étudiée. L'utilisation des lectines WGA (*wheat germ agglutinin*) et ConA (*Concanavalin A*) a montré la présence de molécules osidiques dans la matrice extracellulaire, en particulier composées de GlcNAc, NeuAc, α -mannose et/ou α -glucose.

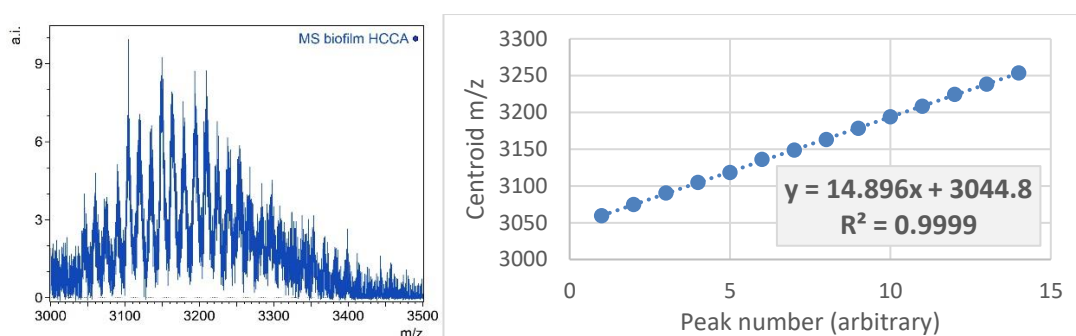


Figure 52 : Massif d'intérêt détecté par MALDI-TOF.

Les écarts de 14,9 m/z entre chaque pic du massif sont constants.

L'identification de ces résidus a été entreprise par MALDI-TOF/TOF. L'utilisation d'une matrice HCCA a mis en évidence la présence d'un signal important entre 3000 et 3500 m/z qui pourrait correspondre, en termes de gamme de masse, aux espèces glycosylées recherchées. Cependant, le signal disparaissant après traitement protéasique (Protéinase K), il doit nécessairement être au moins en partie constitué d'acides aminés. L'identification formelle de ces espèces n'a pas encore été réalisée et validée cependant.

1 **The zoonotic pathogen *Leptospira interrogans* mitigates environmental**
2 **stress through cyclic-di-GMP-controlled biofilm production**

3 Roman Thibeaux¹, Marie Estelle Soupé-Gilbert¹, Malia Kainiu¹, Dominique Girault¹, Emilie
4 Bierque¹, Julien Fernandes², Heike Bähre³, Anthony Douyère⁴, Nicolas Eskenazi⁵, Joëlle
5 Vinh⁵, Mathieu Picardeau⁶ and Cyrille Goarant¹.

6 1 Institut Pasteur de Nouvelle-Calédonie, Unité de Recherche et d'Expertise sur la Leptospirose, Nouméa,
7 Nouvelle-Calédonie

8 2 Institut Pasteur, UTechS PBI, Centre de Ressources et Recherches Technologiques (C2RT), Paris, France.

9 3 Hannover Medical School, Research Core Unit Metabolomics, Hannover, Germany

10 4 Université de la Nouvelle-Calédonie, Institut des Sciences Exactes et Appliquées, Plateau MET/MEB,
11 Nouméa, Nouvelle-Calédonie

12 5 École Supérieure de Physique et de Chimie Industrielles de la ville de Paris, Spectrométrie de Masse
13 Biologique et Protéomique, CNRS, Université Paris-Sciences-et-Lettres, Paris, France

14 6 Institut Pasteur, Unité Biologie des Spirochetes, Paris, France

15 **Abstract**

16 The zoonotic bacterium *Leptospira interrogans* is the aetiological agent of leptospirosis, a re-emerging
17 infectious disease that is a growing public health concern. Most human infections are due to bacterial
18 persistence in the environment. Like many bacterial species, *L. interrogans* forms a biofilm; however,
19 the latter's role as a protective structure in maintaining virulent leptospires in the environment or in the
20 host had not previously been addressed. Here, we examined spatial and time-domain changes in
21 biofilm production by *L. interrogans*. Our observations showed that biofilm formation in *L.*
22 *interrogans* is a highly dynamic process and leads to a polarized architecture. We notably found that
23 the biofilm matrix is composed of extracellular DNA, which enhances the biofilm's cohesiveness. By
24 studying *L. interrogans* mutants with defective diguanylate cyclase and phosphodiesterase activities,

25 we showed that biofilm production is regulated by intracellular levels of bis-(3'-5')-cyclic dimeric
26 guanosine monophosphate (C-di-GMP) and underpins the bacterium's ability to withstand a wide
27 variety of simulated environmental stresses. Our present results (i) show how the C-di-GMP pathway
28 regulates biofilm formation by *L. interrogans*, (ii) provide insights into the environmental persistence
29 of *L. interrogans* and, more generally, (iii) highlight leptospirosis as an environmentally-borne threat
30 to human health.

31 **Introduction**

32 Pathogenic species of bacteria from the genus *Leptospira* cause leptospirosis, a zoonosis that has been
33 observed worldwide but is especially prevalent in tropical low-income countries¹. The disease affects
34 1.03 million humans annually, and kills nearly 60,000². Pathogenic leptospires persist in the renal
35 tubules of chronically infected reservoir animals, and are excreted in the urine³ – thus contaminating
36 the environment. Leptospires can survive in the environment for months^{4, 5}. Contact with a
37 contaminated environment is the predominant cause of transmission to humans⁶. Pathogenic
38 *Leptospira* are able to produce a biofilm *in vitro*,^{7,8} and the presence of pathogenic *Leptospira* within a
39 biofilm in the environment has been reported - suggesting that this structure has an important role in
40 *Leptospira*'s ability to survive *in natura*^{4,9} and in water distribution systems^{10,11}. Studies of reservoir
41 mammals have highlighted the presence of bacterial aggregates within the lumen of proximal renal
42 tubules^{12,13}. These aggregates share some morphological features with biofilms, which might explain
43 why *Leptospira* are so persistent and why antibiotic treatments fail to clear the bacteria from the
44 kidneys¹⁴.

45 Biofilms are multicellular aggregates embedded in a three-dimensional self-produced matrix that
46 confers protection and has tissue-like properties¹⁵. The matrix formed by bacteria is held together by
47 interconnecting compounds, such as self produced polysaccharides, proteins, extracellular DNA
48 (exDNA) and cell lysis products, and material from the surrounding environment¹⁶. In nature, biofilms
49 are frequently established on hydrophilic solid and semi solid surfaces, such as soil, rocks and plants.
50 Under adverse conditions (such as desiccation, osmotic shock, and exposure to toxic compounds, UV
51 radiation and predators), the microbial community as a whole can provide protection¹⁷⁻²¹.

52 *Leptospira*'s ability to remain viable and infectious in various environments (such as the soil, water,
53 and mammalian hosts) suggests that these bacteria have fast-acting, sensitive regulatory systems that
54 enable them to adapt to various environmental challenges. In most biofilm-producing bacterial species
55 studied to date, biofilm formation is regulated by the signalling molecule bis-(3'-5')-cyclic dimeric
56 guanosine monophosphate (C-di-GMP)²²⁻²⁴. C-di-GMP is produced by diguanylate cyclases (DGCs)
57 and degraded by phosphodiesterases (PDEs)^{25, 26}. Production of this intracellular signalling molecule is
58 low in motile cells whereas levels peak during biofilm formation^{27, 28}. Data on C-di-GMP levels in
59 *Leptospira* are scarce^{29, 30}. An analysis of the *L. interrogans* genome found 25 proteins predicted to be
60 involved in C-di-GMP metabolism³¹. Most of these enzymes were found to be active *in vivo*³⁰. C-di-
61 GMP therefore appears to control many facets of *Leptospira*'s lifecycle.

62 Although resistant forms have not been described to date, we hypothesized that *L. interrogans*' biofilm
63 helps virulent leptospires to persist in the environment. We therefore sought to better understand the
64 formation, organization, composition, role and regulation of biofilms in this bacterium, notably using a
65 broad range of microscopy techniques. Our study generated novel, unexpected results that deepen our
66 understanding of how *L. interrogans*' biofilm is formed and organized. In particular, we revealed the
67 critical role of exDNA in biofilm cohesion. By studying mutants with impairments in the C-di-GMP
68 pathway, we evidenced the key role of intracellular C-di-GMP in controlling biofilm formation and
69 motility in *L. interrogans*. We also showed that biofilm production confers resistance to simulated
70 environmental stresses but does not seem to promote virulence or chronic infectivity.

71

72 **Results**

73 **Biofilm formation is a collective and dynamic process**

74 We first sought to describe the formation of *L. interrogans*' biofilm. To this end, we applied a
75 combination of phase-contrast microscopy, time-lapse imaging, confocal laser scanning microscopy
76 (CLSM), and scanning electron microscopy (SEM) (Fig. 1). The phase-contrast microscopy data (Fig.
77 1A) indicated that the *Leptospira* biofilm changes its shape significantly over time – notably by
78 expanding through a branching process. These bacterial aggregates' sugar moieties were stained with

79 Crystal violet (Fig. 1B). Additional live/dead staining of the biofilm aggregates (Fig. 1C) showed that
80 most of the bacteria in the biofilms were alive. The biofilm's shape varied markedly from one
81 *Leptospira* species to another, with dotted, branching and reticulated patterns (Supplemental Fig. 1A).
82 Our time-lapse microscopy and SEM images revealed the whole sequence of biofilm formation (Fig.
83 1E, F). Motile planktonic *L. interrogans* cells were observed shortly after bacteria had been seeded in
84 the glass slide chamber. Small dots of multicellular structures then appeared over the first 72h of
85 culture (Fig. 1E, F, G). Live imaging showed that microcolonies originated from aggregation of
86 planktonic bacteria (which were steadily depleted from the supernatant), rather than by focal clonal
87 growth (Supplemental Fig 1B). These multicellular structures were not anchored to the culture dish
88 but remained motile for up to seven days, and moved together on the substrate surface (Movie 1). As
89 more and more multicellular structures appeared over time, we observed collisions and
90 interconnections – leading to massive aggregates that expanded across the surface of the microscope
91 slide (Fig. 1F, G) and, ultimately, stopped moving due to a lack of space. Supplemental Fig. 1C shows
92 a coalescence event that occurred on day three and led to a decrease in the number of aggregates.
93 Extracellular matrix (ECM) deposits were visible under the SEM as early as day 1. After a week, the
94 mature biofilms had grown perpendicularly to the substrate (Fig. 1F) and were more than 50 μm thick.
95 Our CLSM experiments highlighted the presence of a void within the aggregates (Fig. 1G) formed by
96 the collective, synchronized migration of the bacteria; after spreading out, the aggregate contracted to
97 create the void. The full sequence of events is shown in Supplemental Fig. 1D and Supplemental
98 Movie S1.

99

100 **The biofilm's architecture is highly polarized**

101 A thorough SEM and CLSM analysis of the mature biofilm (Fig. 2) highlighted a number of
102 characteristic structural features. The basal (substrate) side of the biofilm had a rough texture (Fig. 2A,
103 B), and large channels (>five μm in diameter) were visible. The spiral-shaped bacteria could be clearly
104 seen because little or no ECM was present (Fig. 2C). Given the porous nature of the basal side of the
105 biofilm, the entire structure was anchored to the substrate by a small contact area. In contrast, the
106 apical (exposed) side of the biofilm was flat and smooth (Fig. 2D). High-magnification imaging

107 revealed tiny pores (diameter < 1 μm) and bacteria enmeshed within a thick matrix. An analysis of
108 three-dimensional sections of the mature biofilm (Fig. 2F) confirmed the porous, foam-like nature of
109 the structure. The channels that permeated the entire biofilm were wider on the basal side than on the
110 apical side (Fig. 2G). The mature biofilm contained unexpected phenotypic features, such as budding
111 structures, rod-shaped bacteria, and branching, mucus-like filaments (Fig. 2H).

112

113 **Extracellular DNA has a structural role in the *Leptospira* biofilm**

114 We used fluorescent probes to assess the mature biofilm's chemical composition (Fig. 3A).
115 Extracellular DNA was abundant within the mature biofilm. No proteins (other than those constituting
116 the bacterial backbone) were detected in the matrix. Specific wheat germ agglutinin staining of N-
117 acetyl-d-glucosamine and N-acetylneuraminic acid confirmed that these residues were present in the
118 matrix. Additional concanavalin A staining revealed the presence of x-mannopyranosyl and x-
119 glucopyranosyl moieties within the matrix. We further explored the matrix's composition by using
120 enzymatic treatments (Fig. 3B). Exposure to proteinase K or alginate lyase did not dissociate the
121 biofilm structure. Conversely, DNase disrupted the biofilm after 24 h, with a significant loosening of
122 the biofilm's overall structure. Sodium periodate dissolved the biofilm in less than 3 h, releasing
123 planktonic cells into the supernatant. A MALDI-TOF analysis of the matrix (Supplemental Fig. 2)
124 showed mass peaks at 3000-3500 m/z, and a mass difference between the peaks of 14.9 m/z
125 (Supplemental Fig. 2A, B). A correspondence with known adducts could not be established.

126

127 **The intracellular level of C-di-GMP regulates biofilm production**

128 We assessed the ability to form a biofilm in 15 *L. interrogans* transposon mutants with defects in the
129 C-di-GMP pathway or in putative biofilm components. Based on their biofilm production phenotype,
130 four mutants were selected for further study (Fig. 4A-D). Three transposon insertions occurred within
131 the C-di-GMP synthesis pathway, namely in the mutants *dgcA* (LMANV2_v2_150005; GGDEF-only;
132 DGC, C-di-GMP synthesis), *pdeA* (LMANV2_v2_270021; EAL-only; PDE, C-di-GMP degradation),
133 and *pdeB* (LMANV2_v2_90001; EAL-only; PDE, C-di-GMP degradation). The fourth mutant (*akrA*)
134 carried a mutation in an aldoketoreductase-encoding gene (LMANV2_v2_50029) that is reportedly

135 over-expressed and involved in biofilm production in *Staphylococcus aureus* and *Vibrio cholerae*^{32,33}.
136 After 3 weeks of culture, the proportion of biofilm surface coverage was 15.83% for *akrA*, 16.05% for
137 *dgcA*, and 44.6% for the wild-type (WT) strain (Fig. 4A, B). Conversely, *pdeA* and *pdeB* displayed
138 abnormally high levels of biofilm production, with surface coverages of 54.17% and 57.17%,
139 respectively (Fig. 4C, D). The mutants' phenotypes were rescued by complementation with the WT
140 gene. The complemented C-*akrA* and C-*dgcA* strains completely recovered their ability to produce a
141 biofilm, with 2.9- and 2.7-fold increases in the surface coverage, respectively, (Fig. 4A, B). The C-
142 *pdeA* and C-*pdeB* strains also changed their phenotype (Fig. 4C, D) but the proportion of biofilm
143 surface coverage fell to a value below that of the WT parental strain, with a 4.4- and 2.3-fold decrease
144 in biofilm production respectively. Trans-complementation may not restore gene expression to WT
145 levels, even when a low-copy-number plasmid is used for complementation.

146 We assessed changes in *Leptospira's* proliferative and motile capacities. The four mutant strains
147 exhibited much the same growth kinetics as the WT strain, and did not display any impairments in
148 proliferation (Fig. 4E). However, we observed discrepancies in velocity (Fig. 4F and Supplemental
149 Table S2). The velocity was significantly higher in *dgcA* than in the WT. In contrast, *pdeA* and *pdeB*
150 were less motile than the WT. Given that *akrA* was not affected in the C-di-GMP pathway, this strain
151 was used as a control for a C-di-GMP-independent defect in biofilm production; its velocity was
152 similar to that of the WT. We next assessed the impact of mutations in DGCs (using the *dgcA* strain)
153 or PDEs (using *pdeA* and *pdeB*) on intracellular C-di-GMP concentrations (Fig. 4G and Supplemental
154 Table S3). The level was significantly lower in *dgcA* than in the WT; this was expected because the
155 corresponding gene product is involved in C-di-GMP synthesis. A 3-fold increase was observed after
156 complementation with the gene encoding a functional DGC. Conversely, as expected, the intracellular
157 C-di-GMP levels were greatly elevated in *pdeA* and *pdeB*, since the C-di-GMP-degrading PDE
158 enzyme has been knocked out. Complementation with the WT gene led to a drastic 11-fold decrease in
159 intracellular C-di-GMP levels for C-*pdeA* and a 2.4-fold decrease for C-*pdeB*. In both the *akrA* and C-
160 *akrA* strains, the intracellular C-di-GMP content was similar to that of the WT strain.

161

162 ***Leptospira's* biofilm does not promote virulence but protects against environmental stresses.**

163 We next looked at whether *Leptospira*'s ability to produce a biofilm was involved in virulence or
164 colonization, using a hamster model of acute disease and a mouse model of chronic renal carriage. In
165 hamsters infected with pdeB (i.e. strain with high biofilm production but low motility), the median
166 time to death (7.5 days) was longer than in WT infections (4 days) (Fig. 5A). The median time to
167 death was also 4 days for the dgcA and akrA strains. Complementation of pdeB with the WT gene
168 substantially restored virulence, with a median time to death of 3.5 days.

169 Chronic carriage in mice was investigated by quantifying the bacterial load in urine and kidney (Fig.
170 5B). The bacterial loads in both urine and kidney were lower for pdeA and pdeB than for the WT –
171 reflecting a lower colonization capacity. In contrast, the bacterial loads for dgcA and akrA were
172 similar to those observed for the WT.

173 We investigated the biofilm's role as a protective barrier against external stresses by exposing mature
174 biofilms to simulated environmental stresses: elevated NaCl concentrations, ultraviolet light, the
175 antibiotic tetracycline, heat, and acidic/basic pH conditions (Fig 5C). The degree of protection was
176 assessed by measuring the metabolic activity of the bacteria contained in the biofilm after stress
177 exposure. Our results revealed that regardless of the type of stress, the two biofilm-over-producing
178 mutants (pdeA and pdeB) always displayed a higher level of cellular activity than biofilm-under-
179 producing mutants (dgcA and akrA) (Fig. 5C); the WT strain was positioned between these two
180 groups. Furthermore, we observed that biofilm-forming bacteria were less sensitive to stress than their
181 planktonic counterparts (Supplemental Fig 3A). In particular, we found that biofilm-over-producing
182 mutants survived exposure to the concentration of salt (36 mg/mL) found in sea water. Following
183 exposure to intense UV light (20 mJ/cm²), the metabolic activity of biofilm-forming bacteria remained
184 exceptionally high. In line with the results described above, complementation with the WT gene
185 reverted the mutants' ability to withstand NaCl and UV exposure (Fig 5C). Each strain's metabolic
186 activity after NaCl or UV exposure was positively correlated with its biofilm production capacity
187 (Supplemental Fig 3B). Lastly, we examined the biofilm produced by these four mutants at the cellular
188 level. Scanning electron micrographs (Fig. 5D) highlighted significant interstrain differences in ECM
189 production. The pdeA and pdeB mutants were deeply enmeshed within a thick matrix that prevented

190 individual bacteria from being seen, whereas individual bacteria could easily be seen in the ECM-poor
191 biofilm produced by the *dgcA* and *akrA* mutants.

192

193 **Discussion**

194 Our multidisciplinary analysis of the initial steps in biofilm formation provided detailed information
195 on the strategies used by pathogenic *L. interrogans* to create resilient biofilms and thus protect itself
196 against extracellular stresses. We also showed that the C-di-GMP pathway is functional in *L.*
197 *interrogans* and has a pivotal role in motility and biofilm formation in this bacterium.

198 Firstly, we described the sequence of events that leads to the formation of a mature biofilm (Fig 1).
199 We confirmed a previous report that these structures were indeed biofilms⁷, in line with the
200 conventional definition. The aggregates were stained by the crystal violet dye (Fig 1B) typically used
201 to reveal bacterial biofilms. Furthermore, we demonstrated conclusively that the *L. interrogans*
202 biofilms primarily contain live bacteria (Fig 1C), and are not haphazard aggregates of sedimenting
203 dead cells or cell debris.

204 We combined live and high-resolution imaging techniques to reveal differences in biofilm formation
205 in *Leptospira*, relative to well-characterized structures produced by other bacteria¹⁵ (Fig 1D, E). For
206 example, we found that the initial steps in *Leptospira* biofilm formation result from the aggregation of
207 planktonic cells rather than through the conventional, clonal growth of adherent cells. The initial
208 aggregates did not adhere to the substratum but were exceptionally motile for multicellular microbial
209 communities. The coordinated motion of large numbers of biofilm-forming bacteria was observed for
210 at least five days and stopped after a week - possibly due to a lack of space in the microscopy chamber
211 (Fig 1F and Movie 1). Although it is well known that leptospires are fast swimmers, it appears that
212 they move even more quickly when crawling on a solid surface³⁴ at speeds of up to 10 $\mu\text{m}\cdot\text{s}^{-1}$. The
213 high motility on solid surfaces observed in the present study might challenge the current paradigm of a
214 sessile biofilm attached to a surface. The motile biofilm described here might be associated with
215 several potential benefits, such as more efficient nutrient acquisition, the avoidance of toxic
216 substances, the ability to translocate towards preferred hosts and access optimal colonization sites, and
217 dispersal into the environment.

218 We observed a number of unique organizational features. The *Leptospira* in the aggregates were able
219 to synchronize with each other; the resulting morphogenesis process was similar to that observed in
220 multicellular eukaryotic organisms (Supplemental Fig 1C, D and Supplemental Movie S1), which is
221 suggestive of a particularly well-developed bacterial communication system. This synchronized
222 mobility produced a mushroom-like structure with an empty centre (Fig 1G). We consider that
223 successive fusions of these mushroom-like structures formed the large channels observed on the
224 substrate-adherent (basal) side of the mature biofilm. Imaging of both the basal and apical sides of the
225 mature biofilm (Fig 2C, E) revealed a highly polarized, structurally organized biofilm. The latter's
226 unique architectural features (such as interstitial voids and channels, resulting in a foam-like structure)
227 appears well suited to the diffusion of nutrients, gases, and other metabolites³⁵ required for bacterial
228 life (Fig 2F, G). As seen in our SEM experiments, the *Leptospira* biofilm was massive (>0.5 mm long
229 and several hundred μm thick) after 3 weeks of culture. The interconnected channels can be likened to
230 a microcirculatory system, and might favour further growth of the biofilm community by ensuring the
231 influx of nutrients and oxygen and the efflux of waste products³⁶. On the exposed side of the biofilm
232 (i.e. at the interface with the environment), the leptospire were covered by a thick matrix, rather than
233 the fibrous casing³⁷ or curli fibers³⁸ observed in other species. The intriguing presence of budding
234 structures at the surface of the biofilm might be suggestive of a dissemination mechanism via the
235 spread of the biofilm progenitor's structures (Fig 2H). We observed rod-shaped bacteria within the
236 biofilm, which contrasts with *Leptospira*'s typical spiral shape (Fig 2H). Further studies are required
237 to better characterize these rod-shaped *Leptospira* that may represent a new form of viable cells.

238 Our CLSM results showed that exDNA is one of the main ECM constituents of *Leptospira*'s biofilm
239 matrix (Fig 3A). Furthermore, the effect of DNase I treatment highlighted the exDNA's key role in
240 maintaining a dense biofilm (Fig 3B). In several pathogenic bacteria (such as *Listeria monocytogenes*,
241 *Enterococcus faecalis*, and *Pseudomonas aeruginosa*), exDNA is a major biofilm component³⁹.
242 Within single-species biofilms, autolysis is considered to be one of the sources of exDNA⁴⁰. One can
243 therefore reasonably hypothesize that the small proportion of dead bacteria observed within the
244 biofilm undergoes controlled lysis, and thus provides sufficient exDNA for the construction of a
245 biofilm for the predominant live bacteria. Our results also showed that the *L. interrogans* biofilm

246 matrix also contains extracellular polysaccharides (EPSs), although we were not able to determine the
247 type. (Fig 3B). Although alginate is hypothetically a constituent of the *Leptospira* biofilm⁴¹⁻⁴³, our use
248 of alginate lyase did not trigger biofilm dissolution - indicating that alginate is not the main EPS
249 constituent.

250 Recent work³⁰ suggested that the C-di-GMP regulatory pathway is functional in *L. interrogans*.
251 Incubation of *L. interrogans* with murine macrophages led to a decrease in C-di-GMP levels³⁰.
252 Similarly, C-di-GMP levels were downregulated in response to an infection-mimicking 28-to-37°C
253 temperature shift³⁰. The mRNA levels of four genes encoding PDEs were upregulated during the
254 temperature shift, which was consistent with the observed reduction in C-di-GMP concentrations³⁰.
255 Furthermore, the C-di-GMP pathway is known to be regulated during biofilm formation by *Leptospira*
256 *biflexa*⁴³.

257 By studying *Leptospira* mutants with defects in C-di-GMP synthesis or degradation, we demonstrated
258 that impaired C-di-GMP synthesis resulted in a low intracellular concentration of this metabolite (Fig
259 4G). In turn, this resulted in low biofilm production (Fig 4A) and enhanced motility (Fig 4F).
260 Although motility is a virulence factor in *L. interrogans*, the observed elevation in mobility was not
261 linked to greater virulence in the animal models studied here (Fig 5A). However, defects in biofilm
262 production resulted in a lower degree of protection against simulated environmental stresses (Fig 5C).
263 This sensitivity was probably due to the massive relative reduction in the quantity of ECM –thought to
264 act as a shield (Fig 5D). Conversely, low phosphodiesterase enzymatic activity led to the intracellular
265 accumulation of C-di-GMP (Fig 4G), massive biofilm production (Fig 4C, D), and low motility (Fig
266 4F). In a hamster model, pdeB showed a low level of virulence - possibly as a consequence of the loss
267 of motility (Fig5A). We had expected greater biofilm production to lead to more effective colonization
268 of the kidneys. However, chronic carriage in mice was low for the pdeA and pdeB mutants. The low
269 observed urine levels did not result from greater retention of leptospires in the kidneys because the
270 latter were less colonized than in a WT infection (Fig 5B). This finding may be partly explained by the
271 low levels of motility and virulence, which probably impaired the bacterium’s ability to infect the
272 host, spread in the body, and colonize the renal tubules⁴⁴. In general, impairments of a pathway as
273 ubiquitous as the C-di-GMP pathway will necessarily affect many secondary effectors; these changes

274 might also account for the phenotype observed here. Nevertheless, both PDE mutants were extremely
275 resistant to environmental stress (Fig5 C), which was correlated with a thick biofilm ECM (Fig 5D).

276 In conclusion, we used transposon mutants to demonstrate that DGC/PDE activities control
277 intracellular C-di-GMP levels, which regulate motility and biofilm production in *Leptospira*
278 *interrogans*. We showed that the intracellular accumulation of C-di-GMP is associated with enhanced
279 biofilm production and greater ability to withstand environmental stresses. The correlations between a
280 high intracellular C-di-GMP concentration and biofilm formation and, conversely, a low C-di-GMP
281 level and low motility are very similar to those observed in several bacterial species²⁸. Once a mature
282 biofilm has developed, it can be extremely difficult to eradicate *Leptospira*. The biofilm is a highly
283 polarized, organized structure that successfully provides life-compatible conditions - notably due to
284 the protective ECM layer and a complex network of voids for nutrient and gas exchanges. This is
285 particularly true for saprophytic species, whose biofilm production is correlated with the ability to
286 persist in the environment. Indeed, *in vitro* biofilm production is mostly prominent for *Leptospira*
287 species in the saprophyte subclade S1 (Supplemental Fig. S3C); this is very interesting from an
288 ecological viewpoint, since these species are fully adapted to an exclusively environmental lifestyle. In
289 contrast, *L. borgpetersenii* was the species that produced the least biofilm under our experimental
290 conditions (Supplemental Fig. S1A). This species belongs to the pathogenic subclade P1, and has a
291 reduced genome (lacking several genes involved in exopolysaccharide production⁴¹) as a result of full
292 adaptation to its mammal host and evolution towards a parasitic lifestyle - resulting in poor ability to
293 survive in the environment⁴⁵. The fact that multimicrobial environmental biofilms can spontaneously
294 harbour *Leptospira* (including pathogenic species) *in situ*⁹ may explain why viable, virulent
295 *Leptospira* can be found after prolonged periods of time in the environment⁴.

296 Lastly, by regulating biofilm formation, C-di-GMP signalling can also influence *Leptospira*'s ability
297 to resist a wide variety of stresses. Manipulating and subverting C-di-GMP signalling might
298 compromise biofilm formation, antimicrobial tolerance, and *Leptospira*'s persistence – thus opening
299 up perspectives for the control of leptospirosis^{46, 47}.

300

301 **Materials and Methods**

302 **Bacterial strains, culture conditions and transposon mutagenesis.**

303 *Leptospira* strains used in this study, including *Leptospira interrogans* serogroup Pyrogenes serovar
304 Manilae strain L495 designated in this study as the wild type (WT) or parental strain, are listed in
305 Supplemental table S1 and were cultivated in Ellinghausen and McCullough as modified by Johnson
306 and Harris (EMJH) growth medium at 30°C under static condition. For virulent strains, cultures were
307 regularly passaged into hamsters and used until a maximum of five *in vitro* passages. Mutants *dgcA*
308 (LMANV2_v2_150005 || Frame -2 || Begin: 1223263|| End: 1224255 || Length:993), *akrA*
309 (LMANV2_v2_50029 || Frame +3 || Begin: 242451|| End: 243287|| Length:837), *pdeA*
310 (LMANV2_v2_270021 || Frame +1 || Begin: 2272684 || End: 2274402 || Length:1719) and *pdeB*
311 (LMANV2_v2_90001 || Frame -2 || Begin: 631792 || End: 633144 || Length: 1353) were generated
312 from the parental strain *L. interrogans* serovar Manilae strain L495 by transposon mutagenesis as
313 described previously^{48, 49} and grown under the same conditions. Transposon insertion was confirmed
314 by PCR using primers flanking the insertion sites. Complementation was performed as described
315 previously by cloning genes into the spectinomycin resistant pMAORI plasmid to reintroduce the
316 functional gene under the control of its native promoter⁵⁰. Similarly, an empty pMAORI plasmid was
317 also introduced into the mutants as a control (Supplemental table S1). Complemented mutants were
318 cultivated under selection pressure (50 µg/mL of Spectinomycin).

319 **Biofilm formation**

320 Bacterial strains were grown to exponential phase and counted using a Petroff-Hausser cell-counting
321 chamber (Hausser Scientific Company, Horsham, PA, USA) under dark field microscopy. Cultures
322 were diluted in EMJH to a final concentration of 1×10^6 bacteria mL⁻¹. Under sterile conditions, 150
323 µL of diluted cultures were dispensed into the wells of a 96-well microtiter polystyrene plate (flat
324 bottom, with low evaporation lid, Nest Biotechnology). As border effect is known to occur, the
325 peripheral wells of the plate were not used, but instead filled with sterile water to minimize
326 evaporation. For quantitative assays, 18 replicate wells were used for each strain. Microtiter plates
327 were then incubated at 30°C for 3 weeks under static condition without changing the culture medium.

328 **Crystal Violet (CV) staining and quantification**

329 After three weeks of static incubation, planktonic cells were removed by gently pipetting the
330 supernatant. The biofilm was gently rinsed with 200 μL of PBS (137 mM NaCl, 2.7 mM KCl, 8 mM
331 Na_2HPO_4 , and 2 mM KH_2PO_4) at room temperature. Fixation was performed by incubation with 150
332 μL of 4% paraformaldehyde (PFA) in PBS at room temperature for 30 min. Fixative buffer was then
333 removed and biofilm was rinsed two times with 200 μL of PBS prior to staining with 175 μL of 0.1 %
334 CV solution. After 10 min of incubation, the CV solution was removed. Biofilms were rinsed two
335 times with 200 μL PBS and dried overnight. The stain was then released with 200 μL of destaining
336 solution [50% (vol/vol) ethanol, 50% (vol/vol) glacial acetic acid]. The amount of stain released was
337 quantified by measuring the absorbance at 490 nm with a microplate reader (Multiskan FC, Thermo
338 Scientific). Alternatively, phase contrast images of stained biofilms were captured using an upright
339 Leica DM4000 B microscope (Leica microsystems, Mannheim, Germany).

340 **Quantification of biofilm formation by image analysis**

341 Biofilm quantification was performed from phase contrast images using the ImageJ software⁵¹
342 (<https://fiji.sc/>). Phase contrast images of the growing biofilm were acquired using an upright Leica
343 DM4000 B microscope equipped with a 5X lens (HCX PL FLUOTAR 5X/0,15, 12 mm working
344 distance). Pictures of six replicate wells per independent experiment were acquired and further
345 analysed. Prior to quantification, the outer part of the image affected by spherical aberrations and field
346 curvature artefacts was cropped. The 2D grayscale image was then binarized and analysed using the
347 “Analyse Particles” function. The biofilm score was then express as a percentage of surface occupied
348 by the biofilm.

349 **Time-lapse imaging**

350 Time-lapse imaging of growing biofilm was acquired using the BioStation IM-Q Inverted Microscope
351 (Nikon, Tokyo, Japan). Five hundred μL of a bacterial suspension at 10^6 bacteria mL^{-1} were
352 inoculated into a 35-mm glass bottom Hi-Q4 dish (Ibidi, Martinsried, Germany). Glass bottom Hi-Q4
353 dish incorporate a plane parallel top plate, preventing light path distortion by the meniscus, therefore
354 enabling high-quality phase contrast observations. Time-lapse imaging started 24h after seeding under

355 static condition at 30°C and 95% of humidity. Phase contrast images were acquired every 30 min over
356 a period of 9 days using a 20X objective (field of view: 413*311µm).

357 **Confocal Laser Scanning Microscopy**

358 Biofilms were prepared as described in the Biofilm Formation section except inoculation performed in
359 µ-dish 35 mm (Ibidi, Martinsried, Germany). Biofilms were fixed in 4% PFA/PBS solution for 30 min
360 or stained using the LIVE/DEAD Biofilm Viability Kit (Invitrogen™), wheat-germ agglutinin (WGA,
361 10µg/mL) conjugated to Fluorescein (Invitrogen™), Concanavalin A (Con A, 10 µg/mL) conjugated
362 to Fluorescein (Invitrogen™), BOBO-3 (3 nM, Invitrogen™), or the FilmTracer SYPRO Ruby
363 Biofilm Matrix Stain (Invitrogen™), as described by the manufacturer's instructions. Staining with the
364 Vybrant® CFDA/SE Green CellTracker (10 µM, ThermoFisher Scientific) was performed prior to
365 fixation. According to the manufacturer, WGA binds N-acetyl-d-glucosamine and N-acetylneuraminic
366 acid residues, Con A selectively binds to x-mannopyranosyl and x-glucopyranosyl residues, BOBO-3
367 is a cell-impermeable DNA stain, and SYPRO Ruby stain labels most classes of proteins. The stained
368 biofilms were visualized using an inverted SP8 confocal laser-scanning microscope (Leica
369 Microsystems, Mannheim, Germany). A super continuum white light laser tuneable between 470 nm
370 and 670 nm was used for excitation and focused through a HC PL APO CS2 63x (N.A 1.4) oil
371 immersion lens. Emission signals were captured by hybrid detectors. The system was controlled by
372 Leica Application Suite X (LAS) software. 3D images were directly processed using LAS X, 3D
373 module software.

374 **Scanning Electron Microscopy**

375 *Leptospira* strains were cultured as described in the biofilm formation section with minor
376 modifications. Inoculation was performed by adding one mL of diluted culture in a 24-well plate (flat
377 bottom, with low evaporation lid, Corning) each well containing a sterile 12 mm Glass coverslips
378 (12×12 mm, Menzel-Glaser, Braunschweig, Germany). The supernatant was removed and biofilm was
379 rinsed once in PBS, then fixed in 4 % PFA / 1 % glutaraldehyde for 15 min. Coverslips were rinsed
380 again in PBS and stained with 1 % osmium tetroxide (OsO₄ in PBS, Acros Organics) for 1h. After
381 staining, samples were dehydrated through a series of ethanol concentrations (25, 50, 70, 90, and
382 100 % for 10 min each) before being desiccated using hexamethyldisilazane (HMDS, Acros

383 Organics). Before visualization, samples were submitted to carbon sputtering to increase their
384 electrical conductance (Em ACE600, Carbon Coater, Leica Microsystems, Mannheim, Germany) and
385 examined under a Jeol JSM IT-300 IntouchScope microscope (SED detector, 10 kV, Jeol LTD,
386 Tokyo, Japan).

387 **Dissociation of Biofilms by Chemical or Enzymatic Treatments**

388 WT biofilms were grown as described in the Biofilm Formation section, and after 21 days of
389 incubation, twenty μL of DNase I (6.25 to 500 $\mu\text{g}/\text{mL}$ in 150 mM NaCl, 1 mM CaCl_2 ; 10104159001;
390 Roche Diagnostics GmbH, Mannheim, Germany), 20 μL of proteinase K (6.25 to 500 $\mu\text{g}/\text{mL}$ in 50 mM
391 Tris-HCl pH 7.5, 1 mM CaCl_2 ; 05323738001; Roche Diagnostics GmbH, Mannheim, Germany), 20 μL
392 of alginate lyase (1 to 200 U in 20 mM Tris, 200 mM NaCl; A-1603-100MG, Sigma Aldrich, St-
393 Louis, MO, USA) and 20 μL of sodium periodate (1 to 100 mM in 50 mM sodium acetate; S1878-
394 25G, Sigma Aldrich, St-Louis, MO, USA) were added directly to the biofilms. Control wells were
395 treated with 20 μL of the corresponding buffer only. Wells were incubated at 30°C for 48h. Phase
396 contrast images of adherent biofilms were captured with an upright Leica DM4000 B microscope (5X
397 objective HCX PL FLUOTAR 5X/0,15,) using a DFC450 digital CCD camera (Leica Microsystems,
398 Mannheim, Germany).

399 **MALDI-TOF analysis**

400 Twenty mL of a 10^6 bacteria mL^{-1} suspension were inoculated into a 50 mL culture flask (Falcon) and
401 incubated at 30°C under static conditions. After three weeks, supernatant containing planktonic
402 bacteria was gently removed. PBS was added and flasks were incubated at 30°C for 24h in order to
403 rinse the biofilm from EMJH components that could interfere with Matrix Assisted Laser Desorption
404 Ionisation - Time of Flight (MALDI-TOF) experiments (i.e: Tween 80). Biofilm were then collected
405 and lyophilized prior to MALDI-TOF analysis. Five mg of biofilm (dry mass) was solubilized in
406 deionized water then subjected to multiple sonication, 56°C incubation, and vortex homogenization
407 cycles, until no residual biofilm particles were observable. Samples were mixed (1:1, v/v) with alpha-
408 cyano hydroxycinnamic acid (CHCA) (4mg/mL in 60:40 (v/v) acetonitrile/water) and spotted on a
409 stainless steel MALDI probe. Mass spectrometry acquisition was performed on an AB Sciex MALDI-
410 TOF/TOF 5800 system, on 500-4000 m/z range, in positive reflectron mode, using 5000 laser shots

411 per spectrum in continuous stage motion. Accuracy better than 50ppm was reached with external
412 calibration (using 13 calibration spots on the MALDI probe). MS/MS fragmentation was achieved
413 using 1kV or 2kV methods, with or without collision gas, and with either 3 m/z or 100 m/z precursor
414 selection windows.

415 **C-di-GMP extraction and quantification**

416 Forty mL of a 10^6 bacteria mL^{-1} suspension were inoculated into a 50 mL culture flask (Falcon) during
417 21 days at 30°C under static conditions. Twenty mL were used for C-di-GMP extraction and 20 mL
418 were used for normalization. Bacterial suspensions were centrifuged at 2,500 x g for 20 min at 4°C.
419 For C-di-GMP extraction, bacterial pellets were re-suspended in 300 μL of ice-cold extraction solution
420 (acetonitrile/methanol/water; 2/2/1; v/v/v) for 15 min on ice prior to heat extraction at 95°C during 10
421 min. After cooling, the suspension was centrifuged at 20,000 x g for 10 min at 4°C in order to separate
422 insoluble material from the extracted nucleotides. The extraction procedure was repeated twice. The
423 supernatant was collected and then evaporated until dryness at 30°C using a Speed Vac (Concentrator
424 plus, Eppendorf). The pellet was re-suspended in 200 μL of water using vigorous vortexing and then
425 analysed by liquid chromatography (Series 200, Perkin Elmer, Norwalk, CT, USA) coupled with an
426 API 3000 triple quadrupole mass spectrometer (Concord, Ontario, Canada) as described elsewhere⁵².

427 **Growth rate measurements**

428 Mutants, complemented and WT strains were counted using a Petroff-Hausser counting chamber
429 under dark-field microscopy. Five mL of EMJH inoculated at 1.10^6 bacteria mL^{-1} of each strain were
430 cultured over a 9-day period under static conditions. Growth was monitored at 30°C by counting
431 bacteria every day using a Petroff-Hausser counting chamber. Growth experiments were performed
432 three times on different days and using independent cultures.

433 **Bacterial speed displacement quantification**

434 Bacterial strains were grown to exponential phase and 10 μL of bacterial suspension adjusted at 10^6
435 bacteria mL^{-1} were prepared for dark field contrast microscopy using a #1.5H square coverslip.
436 Motility was assessed in liquid medium by video microscopy using a BX53 Olympus microscope
437 (OLYMPUS CORPORATION, Nishi-Shinjuku, Tokyo, Japan) equipped with a Hamamatsu 2.8 Orca
438 flash camera (HAMAMATSU PHOTONICS, Hamamatsu City, Japan). For each strain, three

439 experiments were performed on different days and using independent cultures. In total, 190 bacteria
440 per strain were recorded over 60 seconds. Trajectory analysis and speed displacement were calculated
441 using OLYMPUS CellSens Dimension imaging software.

442 **Animal infection and quantification of bacterial burden**

443 1.10^8 leptospire in EMJH were inoculated intraperitoneally into 7–8-week-old golden Syrian
444 hamsters (males and females) and 8-week-old Oncins-France 1 (OF-1, outbred) mice (males and
445 females) whose progenitors originated from Charles River Laboratories. Hamsters were monitored
446 twice daily for 21 days and subsequently euthanized by carbon dioxide inhalation. During follow-up,
447 hamsters displaying poor or no reaction upon stimulation were humanely euthanized and considered
448 dead. Kidneys were collected from animals sacrificed at day 21. Mice were monitored during 63 days
449 and urine was collected at days 7, 14, 21, 28, 35, 47, 63 after infection. At day 64 mice were
450 euthanized by carbon dioxide inhalation and kidneys were collected. DNA was extracted from urine
451 (50 μ L) and kidney (25 mg) and analysed by a real-time PCR targeting the conserved regions of the
452 16S rRNA gene⁵³. A PCR targeting the pMAORI plasmid was performed at the end of the 63 days of
453 infection to ensure that complemented mutants were still carrying the complementation plasmid
454 (primers pMAORI-1F, 5'-TCG-ATT-GGT-GTA-GTC-GGT-T-3'; pMAORI-1R, 5'-GCA-GCC-ATT-
455 CAA-TTT-CTT-GAG-TTA-3'). Four experiments (12 animals) per condition were performed on
456 different days and using independent bacterial cultures. Animal experiments were conducted
457 according to the guidelines laid out by the Animal Care and Use Committee of the Institut Pasteur of
458 Paris and of New Caledonia, and European Recommendation 2007/526/EC. Protocols and
459 experiments were approved by the Animal Care and Use Committee of the Institut Pasteur in New
460 Caledonia.

461 **Susceptibility assay**

462 *Leptospira* strains were cultured as described in the biofilm formation section. For each strain tested,
463 six wells were vigorously resuspended by pipetting to mechanically disrupt the biofilm, releasing
464 planktonic bacteria. These wells were used to compare biofilm with planktonic cells with a similar
465 metabolic status after three weeks of culture. Environmental-like stress was applied: NaCl (0, 5, 8,
466 12.5, 18 and 36 mg/mL); UV light (0, 2, 5, 8, 10, 20 mJ/cm²); tetracycline (24, 36, 48, 60, 72, 84, 96,

467 108, 120, 132, 144, 156 µg/mL); temperature (30, 37, 41, 42 °C) and pH (5.2, 7.1, 7.4, 7.9). After 24h
468 of stress exposure, 30 µL of alamarBlue (Life Technologies SAS,) was added to the well, and the
469 plates were incubated for 48 h at 30°C. The chromogenic shift was measured by absorbance at 570 nm
470 / 620 nm with a microplate reader (Multiskan FC, Thermo Scientific). Negative controls without
471 bacteria were included. Relative metabolic activity was calculated as recommended by the
472 manufacturer. Experiments were repeated at least four times.

473 **Statistical analysis**

474 All statistical analyses were performed with GraphPad Prism version 7.0 (GraphPad Software, San
475 Diego, CA). The statistical tests used and *P* values are indicated within the text and figure legends. All
476 experiments were performed a minimum of three times on separate days in duplicate or more.

477

478 **Figures legend:**

479 **Figure 1: *Leptospira* biofilm formation**

480 (A) Phase contrast images of *Leptospira* biofilm at 1, 2 and 3 weeks post-inoculation. (B) Phase
481 contrast images of Crystal Violet-stained biofilms. (C) CLSM images of *Leptospira* biofilm stained
482 with Propidium Iodide (PI, left panel) and Syto 9 (central panel). Merged image is shown in the right
483 panel. (D) Phase contrast images of *Leptospira* biofilm formation at 1, 2, 3, 5 and 10 days post-
484 inoculation. (E) SEM images of *Leptospira* biofilm formation at 1, 7, 10, 15 and 20 days post-
485 inoculation. (F) Quantitative phase-contrast image analysis of the number of biofilm aggregates (blue
486 curve) and the biofilm area (red curve) over time between 12 and 216 h post-inoculation. The grey
487 curves represent the standard error of the mean. Insets show binarized images representative of biofilm
488 aspect at 36, 72, 132 and 180h post-inoculation. (G) Correlative phase contrast-CLSM image of
489 biofilm stained with CFDA-SE Cell tracker performed on a native non-fixed biofilm. Phase contrast
490 image (upper left panel), 3D-reconstruction (upper right panel), XY cross-section (lower left panel)
491 and ZY cross-section (lower right panel) are shown.
492

493 **Figure 2: *Leptospira* biofilm architecture and organization**

494 (A) SEM images of the apical and basal sides of a mature *Leptospira* biofilm. (B and C) Higher
495 magnification of the biofilm basal side showing large interconnected channels. (D and E) Higher
496 magnification of the biofilm apical side showing a flat surface of bacteria enmeshed within an
497 amorphous ECM. (F) ZY (upper panel) and XY (lower panel) orthogonal cross-sections of CLSM
498 images showing the foam-like architecture of *Leptospira* biofilm. (G) CLSM images at the top (upper
499 panel) and the bottom (lower panel) sides of a *Leptospira* biofilm illustrating the density gradient
500 along the baso-apical polarity axis. (H) SEM images of a mature biofilm showing budding-like
501 structures (left panel, arrows), rod shaped bacteria (upper right panel, white arrowheads) and
502 branching filaments (lower right panel, red arrow heads).
503

504 **Figure 3: Composition of the biofilm matrix.**

505 (A) CLSM micrographs of biofilms stained with BOBO-3 (exDNA), Sypro Ruby (proteins), WGA
506 (N-acetylglucosamine or sialic acid residues) and ConA (red/ α -D-glucosyl or α -D-mannosyl residues),

507 red, upper line. Bacteria were stained with a CFDA-SE cell-tracker (green, middle line). Merged
508 images are shown in the lower line. (B) Phase contrast images of biofilms left untreated or treated for
509 1, 3, 6, 24 and 48h with proteinase K (1 µg/mL), DNase I (6.25 µg/mL), alginate lyase (125 U/mL)
510 and sodium metaperiodate (5 mM).

511

512 **Figure 4: C-di-GMP intracellular level regulates *Leptospira* motility and biofilm production**

513 Mutant *akrA* is shown in orange, *dgcA* in pink, *pdeA* in blue and *pdeB* in green. Error bars represent
514 the standard errors of the mean. (A) Quantification of biofilm production at 1, 2 and 3 weeks post-
515 inoculation for WT Manilae (red), Manilae mutants (solid line) and complemented mutants (dotted
516 line): *akrA* (A), *dgcA* (B), *pdeA* (C), *pdeB* (D). Insets show a representative phase contrast image of
517 the biofilm growth for the different strains. (E) Proliferation kinetics in liquid EMJH of WT and
518 mutant strains represented as the concentration of *Leptospira* per mL over a nine-day period. (F) Mean
519 displacement speed in µm.s⁻¹ measured *in vitro* in liquid EMJH for WT and mutant Manilae strains.
520 n=190 cells per condition. (G) Quantification of intracellular C-di-GMP levels in WT, mutant and
521 complemented mutant strains. **** indicates a *p* value < 0.0001; *** indicates a *p* value < 0.001; **
522 indicates a *p* value < 0.01 in two-tailed unpaired Mann-Whitney test.

523

524 **Figure 5: *Leptospira* biofilm protects bacteria against environmental-like conditions.**

525 Solid lines represent mutant strains and dotted line complemented strains. Mutant *akrA* is shown in
526 orange, *dgcA* in pink, *pdeA* in blue and *pdeB* in green. Error bars represent the standard errors of the
527 mean. (A) Survival rate over 21 days for hamsters challenged with C-di-GMP mutants and WT strain.
528 * indicates a *p* value < 0.05 in Gehan-Breslow-Wilcoxon test. The table recapitulates the median
529 survival time for each strain. (B) Quantification of *Leptospira* excretion over 63 days for mice
530 challenged with C-di-GMP mutants and WT strain expressed as the number of *Leptospira* per mL of
531 urine after challenge (plotted on the left axis). *Leptospira* kidney load in mice challenged with C-di-
532 GMP mutants and WT strain is expressed as the number of *Leptospira* per mg of kidney at day 64
533 post-inoculation (plotted on the Right axis). (C) Relative metabolic activities of WT and C-di-GMP
534 mutant biofilms under halostress, UV stress, tetracyclin treatment, Heat shock and exposure to various
535 pH. Relative metabolic activities are expressed as a percentage of the metabolic activity of the
536 untreated WT strain (Halostress, UV stress, Tetracyclin) or WT strain under standard culture
537 conditions (Heat and pH). Mean of at least five independent experiments are shown; error bars
538 represent standard errors of the mean. (D) SEM micrograph illustrating matrix production by the
539 different C-di-GMP mutants as compared to the WT strain. * indicates a *p* value < 0.05 in two-tailed
540 unpaired Mann-Whitney test.

541

542

Supplementary figure 1.

543 (A) Phase contrast images of the biofilm of Pathogenic subclade P1 (*L. borgpetersenii*, *L. ellisii*),
544 Pathogen subclade P2 (*L. haakeii*) and Saprophyte subclade S1 (*L. meyeri*) *Leptospira* at 3 weeks
545 post-inoculation. (B) Phase contrast images of *Leptospira* planktonic bacteria at 48 and 72h (left
546 panel) and the corresponding quantification of the amount of planktonic bacteria and bacteria
547 associated to the biofilm. The difference of the absorbance of Cristal Violet (490 nm) staining
548 associated to the supernatant (planktonic bacteria) and to the substrate (biofilm-associated bacteria) is
549 plotted for *L. interrogans* Manilae (red) and the negative control EMJH (black, right panel). (C)
550 Sequence of phase contrast images of biofilm coalescence between 72 and 80h post-inoculation. (D)
551 Sequence of phase contrast images of the synchronized spreading and contraction behaviour
552 delineating cavities within biofilm aggregate (left panels) and quantification of the average perimeter
553 of biofilm aggregates. * indicates a *p* value < 0.05 in two-tailed unpaired Mann-Whitney test.

554

555

Supplementary figure 2.

556 (A) MALDI-TOF MS analysis of biofilm composition showing the intensity of the signal as a function
557 of the ratio *m/z*. The inset shows a magnification of the peak between 3,000 and 3,500 *m/z*. (B) Graph

558 showing m/z of the biofilm between 3,050 and 3,250 m/z as a function of the number of the peak. The
559 equation of the linear regression indicates that peaks are separated by 14.9 m/z. (C) MALDI-TOF MS
560 analysis of biofilm after incubation with proteinase K (PK), major peaks are PK autolysis products.

561 **Supplementary figure 3.**

562 (A) Correlation analysis between Biofilm score expressed as the percentage of well surface colonized
563 by biofilm and relative metabolic activities after exposure to 36 mg/ml of NaCl (left panel) and 20
564 mJ/cm² (right panel). (B) Relative metabolic activities of WT and C-di-GMP mutant planktonic
565 bacteria under halostress, UV stress, tetracycline treatment and Heat shock as compared to their
566 biofilm-associated counterparts. Relative metabolic activities are expressed as a percentage of the
567 metabolic activity of the untreated WT strain (Halostress, UV stress, Tetracycline) or WT strain under
568 standard culture conditions (Heat). Mean of at least five independent experiments are shown; error
569 bars represent standard errors of the mean. (C) Biofilm score expressed as the percentage of well
570 surface colonized by biofilm for subclades P1, P2 and S1. Each dot represents a different *Leptospira*
571 species. Means and standard error of the mean are plotted. * indicates a *p* value < 0.05 in two-tailed
572 unpaired Mann-Whitney test.
573

574

575 **Acknowledgements:**

576 The authors thank Johann Peltier, Nadia Benaroudj, Robert Antony Gaultney, Linda Guentas, Arnaud
577 Tarantola and Catherine Inizan for technical assistance and critical discussions. The authors are
578 indebted to Professor Jean-Marc Ghigo for critical discussions and support at the start of this project.
579 We also thank Julien Colot and the Laboratoire de Biologie Médicale at Centre Hospitalier Territorial
580 Gaston Bourret for granting access to their MALDI-TOF facilities. This work was funded by a
581 postdoctoral fellowship awarded by the AXA Research Fund (reference: 15-AXA-PDOC-037). Lastly,
582 we thank the UTechS Photonic BioImaging facility (Imagopole, C2RT, Institut Pasteur) funded by the
583 French National Research Agency (France BioImaging; ANR-10-INSB-04; Investments for the
584 Future programme).
585

586 **Author contributions:**

587 R.T., C.G. and M.P. designed the project. R.T., performed and analysed most of the experiments. E.B.
588 and D.G. assisted with the biofilm experiments. M.E.S.G. and M.K. performed the animal infection
589 and quantification of bacterial burden. H.B. performed the C-di-GMP quantification. J.F. assisted with
590 fluorescent microscopy analysis. A.D. performed the scanning electron microscopy analysis. N.E. and
591 J.V. performed the MALDI-TOF analysis. C.G., R.T. and M.P. wrote the manuscript. M.P. and C.G.
592 supervised the study and edited the manuscript. All authors reviewed and approved the manuscript.

593 **Corresponding authors**

594 Correspondence to Cyrille Goarant or Roman Thibeaux.

595

- 596 1. Bharti, A.R. et al. Leptospirosis: a zoonotic disease of global importance. *Lancet Infect Dis* **3**,
597 757-771 (2003).
- 598 2. Costa, F. et al. Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLoS*
599 *Negl Trop Dis* **9**, e0003898 (2015).
- 600 3. Adler, B. & de la Pena Moctezuma, A. *Leptospira* and leptospirosis. *Vet Microbiol* **140**, 287-
601 296 (2009).
- 602 4. Thibeaux, R. et al. Seeking the environmental source of Leptospirosis reveals durable
603 bacterial viability in river soils. *PLoS Negl Trop Dis* **11**, e0005414 (2017).

- 604 5. Andre-Fontaine, G., Aviat, F. & Thorin, C. Waterborne Leptospirosis: Survival and
605 Preservation of the Virulence of Pathogenic *Leptospira* spp. in Fresh Water. *Curr Microbiol*
606 **71**, 136-142 (2015).
- 607 6. Goarant, C. et al. in Global Water Pathogen Project (eds. Rose, J.B. & Jiménez-Cisneros, B.)
608 (Michigan State University, E. Lansing, MI, UNESCO, 2019).
- 609 7. Ristow, P. et al. Biofilm formation by saprophytic and pathogenic leptospires. *Microbiology*
610 **154**, 1309-17 (2008).
- 611 8. Brihuega, B., Samartino, L., Auteri, C., Venzano, A. & Caimi, K. In vivo cell aggregations of a
612 recent swine biofilm-forming isolate of *Leptospira interrogans* strain from Argentina. *Rev*
613 *Argent Microbiol* **44**, 138-143 (2012).
- 614 9. Vinod Kumar, K., Lall, C., Raj, R.V., Vedhagiri, K. & Vijayachari, P. Molecular detection of
615 pathogenic leptospiral protein encoding gene (lipL32) in environmental aquatic biofilms. *Lett*
616 *Appl Microbiol* **62**, 311-315 (2016).
- 617 10. Singh, R. et al. Microbial diversity of biofilms in dental unit water systems. *Appl Environ*
618 *Microbiol* **69**, 3412-20 (2003).
- 619 11. Huang, K., Zhang, X.X., Shi, P., Wu, B. & Ren, H. A comprehensive insight into bacterial
620 virulence in drinking water using 454 pyrosequencing and Illumina high-throughput
621 sequencing. *Ecotoxicol Environ Saf* **109**, 15-21 (2014).
- 622 12. Monahan, A., Callanan, J. & Nally, J. Host-pathogen interactions in the kidney during chronic
623 leptospirosis. *Vet Pathol* **46**, 792-799 (2009).
- 624 13. Sterling, C.R. & Thiermann, A.B. Urban rats as chronic carriers of leptospirosis: an
625 ultrastructural investigation. *Vet Pathol* **18**, 628-37 (1981).
- 626 14. Werts, C. Interaction of *Leptospira* with the Innate Immune System. *Curr Top Microbiol*
627 *Immunol* **415**, 163-187 (2018).
- 628 15. Hall-Stoodley, L., Costerton, J.W. & Stoodley, P. Bacterial biofilms: from the natural
629 environment to infectious diseases. *Nat Rev Microbiol* **2**, 95-108 (2004).
- 630 16. Flemming, H.C. & Wingender, J. The biofilm matrix. *Nat Rev Microbiol* **8**, 623-33 (2010).
- 631 17. Le Magrex-Debar, E., Lemoine, J., Gelle, M.P., Jacquelin, L.F. & Choisy, C. Evaluation of
632 biohazards in dehydrated biofilms on foodstuff packaging. *Int J Food Microbiol* **55**, 239-43
633 (2000).
- 634 18. McNeill, K. & Hamilton, I.R. Acid tolerance response of biofilm cells of *Streptococcus mutans*.
635 *FEMS Microbiol Lett* **221**, 25-30 (2003).
- 636 19. Espeland, E.M. & Wetzel, R.G. Complexation, Stabilization, and UV Photolysis of Extracellular
637 and Surface-Bound Glucosidase and Alkaline Phosphatase: Implications for Biofilm
638 Microbiota. *Microb Ecol* **42**, 572-585 (2001).
- 639 20. Stewart, P.S. & Costerton, J.W. Antibiotic resistance of bacteria in biofilms. *Lancet* **358**, 135-8
640 (2001).
- 641 21. Teitzel, G.M. & Parsek, M.R. Heavy metal resistance of biofilm and planktonic *Pseudomonas*
642 *aeruginosa*. *Appl Environ Microbiol* **69**, 2313-20 (2003).
- 643 22. Hengge, R. Principles of c-di-GMP signalling in bacteria. *Nat Rev Microbiol* **7**, 263-73 (2009).
- 644 23. Boyd, C.D. & O'Toole, G.A. Second messenger regulation of biofilm formation: breakthroughs
645 in understanding c-di-GMP effector systems. *Annu Rev Cell Dev Biol* **28**, 439-62 (2012).
- 646 24. Jenal, U., Reinders, A. & Lori, C. Cyclic di-GMP: second messenger extraordinaire. *Nat Rev*
647 *Microbiol* **15**, 271-284 (2017).
- 648 25. Ryjenkov, D.A., Tarutina, M., Moskvina, O.V. & Gomelsky, M. Cyclic diguanylate is a ubiquitous
649 signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. *J*
650 *Bacteriol* **187**, 1792-8 (2005).
- 651 26. Schmidt, A.J., Ryjenkov, D.A. & Gomelsky, M. The ubiquitous protein domain EAL is a cyclic
652 diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *J*
653 *Bacteriol* **187**, 4774-81 (2005).
- 654 27. Romling, U., Galperin, M.Y. & Gomelsky, M. Cyclic di-GMP: the first 25 years of a universal
655 bacterial second messenger. *Microbiol Mol Biol Rev* **77**, 1-52 (2013).

- 656 28. Simm, R., Morr, M., Kader, A., Nimtz, M. & Romling, U. GGDEF and EAL domains inversely
657 regulate cyclic di-GMP levels and transition from sessility to motility. *Mol Microbiol* **53**, 1123-
658 34 (2004).
- 659 29. da Costa Vasconcelos, F.N. et al. Structural and enzymatic characterization of a cAMP-
660 dependent diguanylate cyclase from pathogenic *Leptospira* species. *J Mol Biol* **429**, 2337-
661 2352 (2017).
- 662 30. Xiao, G. et al. Identification and Characterization of c-di-GMP Metabolic Enzymes of
663 *Leptospira interrogans* and c-di-GMP Fluctuations After Thermal Shift and Infection. *Front*
664 *Microbiol* **9**, 764 (2018).
- 665 31. Seshasayee, A.S., Fraser, G.M. & Luscombe, N.M. Comparative genomics of cyclic-di-GMP
666 signalling in bacteria: post-translational regulation and catalytic activity. *Nucleic Acids Res* **38**,
667 5970-81 (2010).
- 668 32. Beenken, K.E. et al. Global gene expression in *Staphylococcus aureus* biofilms. *J Bacteriol* **186**,
669 4665-84 (2004).
- 670 33. Mueller, R.S. et al. *Vibrio cholerae* strains possess multiple strategies for abiotic and biotic
671 surface colonization. *J Bacteriol* **189**, 5348-60 (2007).
- 672 34. Tahara, H. et al. The mechanism of two-phase motility in the spirochete
673 *Leptospira*: Swimming and crawling. *Science Advances* **4** (2018).
- 674 35. Stewart, P.S. Diffusion in biofilms. *J Bacteriol* **185**, 1485-91 (2003).
- 675 36. Berlanga, M. & Guerrero, R. Living together in biofilms: the microbial cell factory and its
676 biotechnological implications. *Microb Cell Fact* **15**, 165 (2016).
- 677 37. Hung, C. et al. *Escherichia coli* biofilms have an organized and complex extracellular matrix
678 structure. *MBio* **4**, e00645-13 (2013).
- 679 38. Erskine, E., MacPhee, C.E. & Stanley-Wall, N.R. Functional Amyloid and Other Protein Fibers
680 in the Biofilm Matrix. *J Mol Biol* **430**, 3642-3656 (2018).
- 681 39. Whitchurch, C.B., Tolker-Nielsen, T., Ragas, P.C. & Mattick, J.S. Extracellular DNA required for
682 bacterial biofilm formation. *Science* **295**, 1487 (2002).
- 683 40. Montanaro, L. et al. Extracellular DNA in biofilms. *Int J Artif Organs* **34**, 824-31 (2011).
- 684 41. Picardeau, M. et al. Genome Sequence of the Saprophyte *Leptospira biflexa* Provides Insights
685 into the Evolution of *Leptospira* and the Pathogenesis of Leptospirosis. *PLoS ONE* **3**, e1607
686 (2008).
- 687 42. Grassmann, A.A. et al. Discovery of Novel Leptospirosis Vaccine Candidates Using Reverse
688 and Structural Vaccinology. *Front Immunol* **8**, 463 (2017).
- 689 43. Iraola, G. et al. Transcriptome Sequencing Reveals Wide Expression Reprogramming of Basal
690 and Unknown Genes in *Leptospira biflexa* Biofilms. *mSphere* **1**, e00042-16 (2016).
- 691 44. Ratet, G. et al. Live Imaging of Bioluminescent *Leptospira interrogans* in Mice Reveals Renal
692 Colonization as a Stealth Escape from the Blood Defenses and Antibiotics. *PLoS Negl Trop Dis*
693 **8**, e3359 (2014).
- 694 45. Bulach, D.M. et al. Genome reduction in *Leptospira borgpetersenii* reflects limited
695 transmission potential. *Proc Natl Acad Sci U S A* **103**, 14560-5 (2006).
- 696 46. Opoku-Temeng, C., Zhou, J., Zheng, Y., Su, J. & Sintim, H.O. Cyclic dinucleotide (c-di-GMP, c-
697 di-AMP, and cGAMP) signalings have come of age to be inhibited by small molecules. *Chem*
698 *Commun (Camb)* **52**, 9327-42 (2016).
- 699 47. Koo, H., Allan, R.N., Howlin, R.P., Stoodley, P. & Hall-Stoodley, L. Targeting microbial biofilms:
700 current and prospective therapeutic strategies. *Nat Rev Microbiol* **15**, 740-755 (2017).
- 701 48. Bourhy, P., Louvel, H., Saint Girons, I. & Picardeau, M. Random insertional mutagenesis of
702 *Leptospira interrogans*, the agent of leptospirosis, using a mariner transposon. *J Bacteriol*
703 **187**, 3255-3258 (2005).
- 704 49. Murray, G.L. et al. Genome-wide transposon mutagenesis in pathogenic *Leptospira* spp.
705 *Infect Immun* **77**, 810-816 (2009).

- 706 50. Pappas, C.J., Benaroudj, N. & Picardeau, M. A replicative plasmid vector allows efficient
707 complementation of pathogenic *Leptospira* strains. *Appl Environ Microbiol* **81**, 3176-3181
708 (2015).
- 709 51. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**,
710 676-82 (2012).
- 711 52. Burhenne, H. & Kaefer, V. Quantification of cyclic dinucleotides by reversed-phase LC-
712 MS/MS. *Methods Mol Biol* **1016**, 27-37 (2013).
- 713 53. Merien, F., Amouriaux, P., Perolat, P., Baranton, G. & Saint Girons, I. Polymerase chain
714 reaction for detection of *Leptospira* spp. in clinical samples. *J Clin Microbiol* **30**, 2219-2224
715 (1992).
- 716
- 717

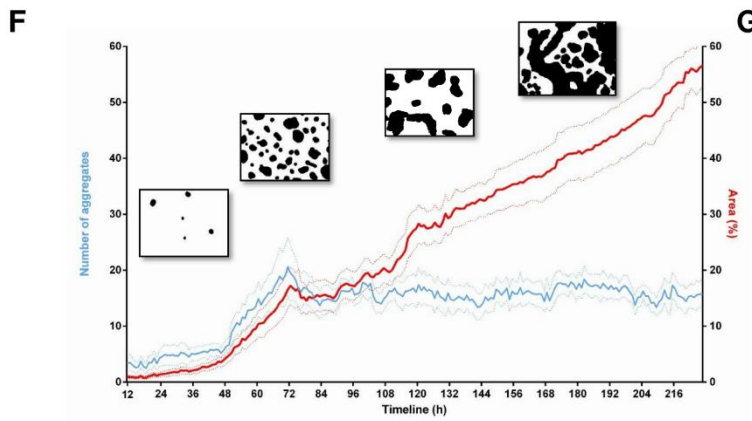
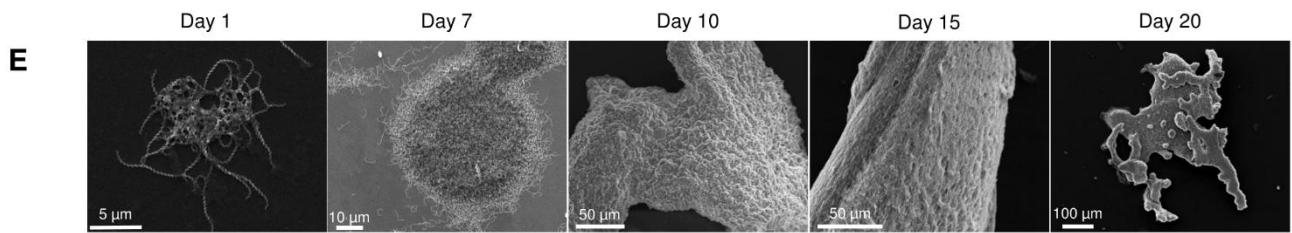
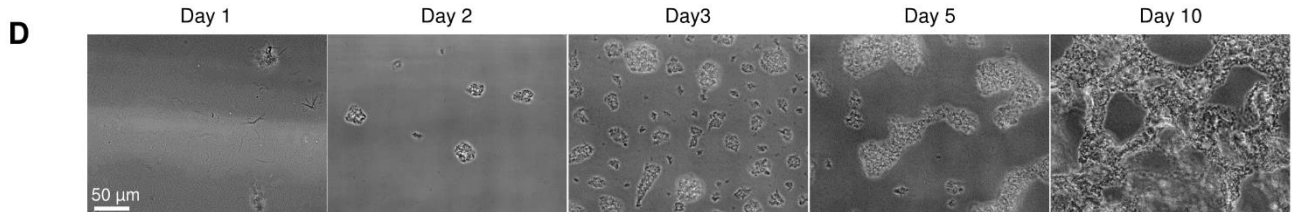
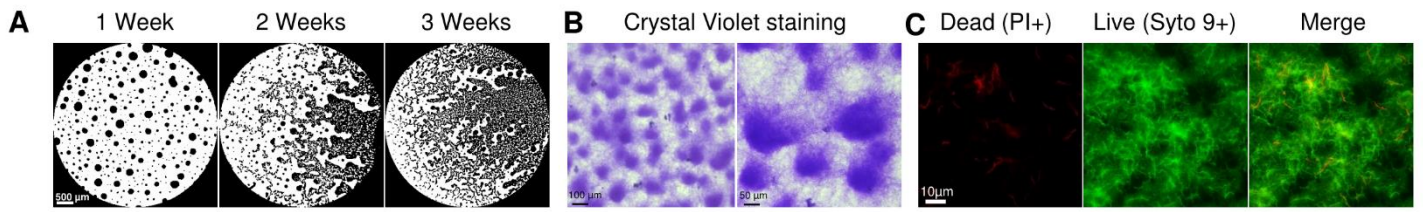
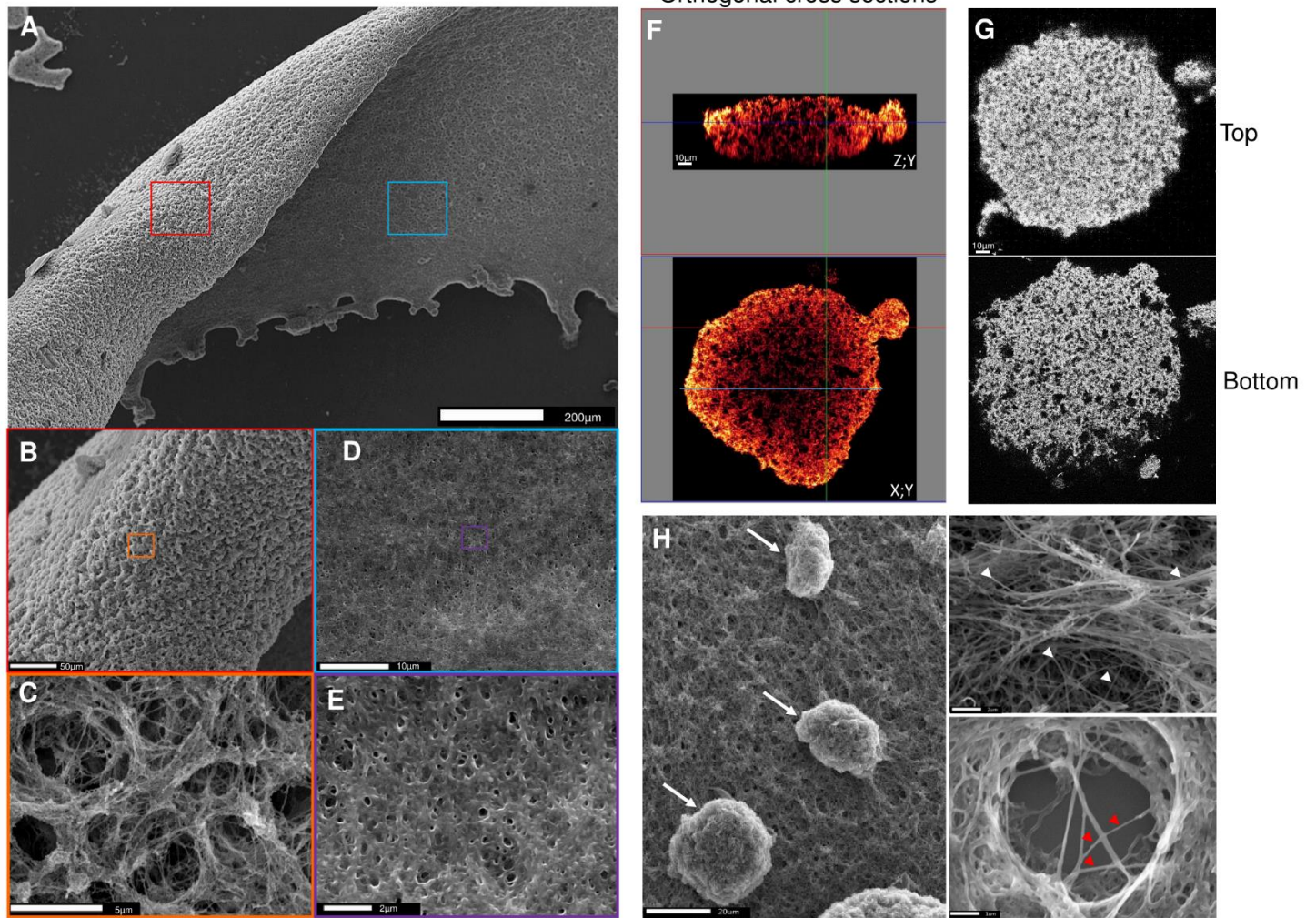
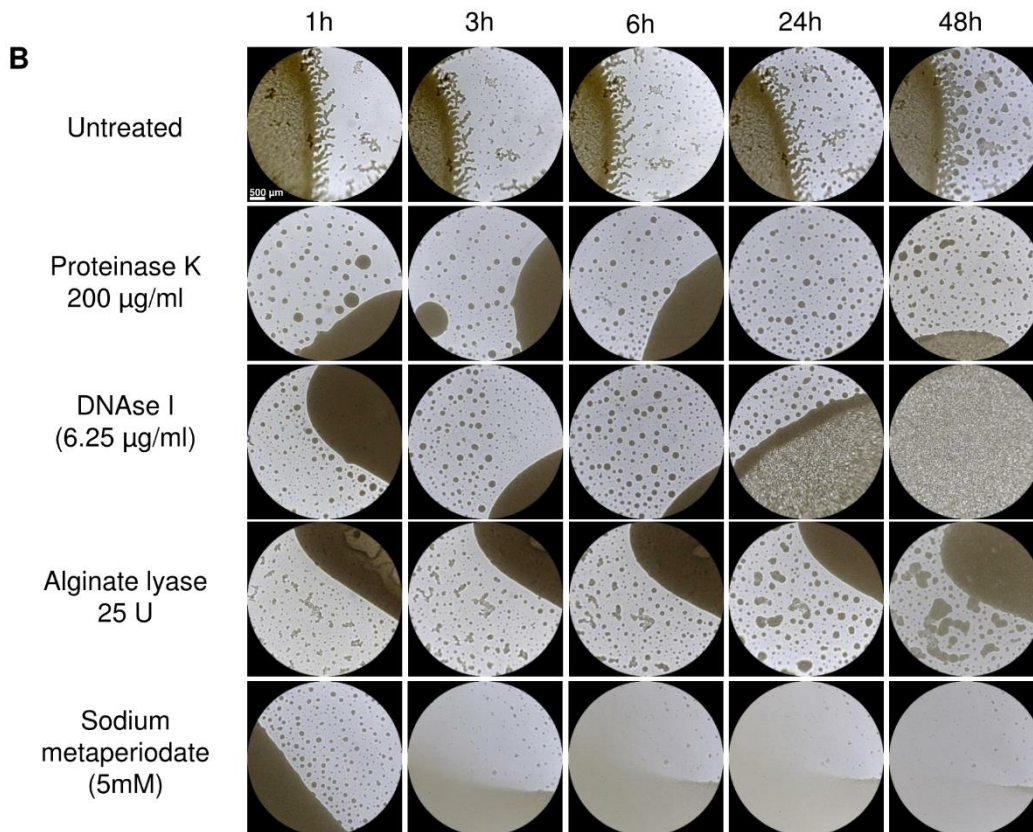
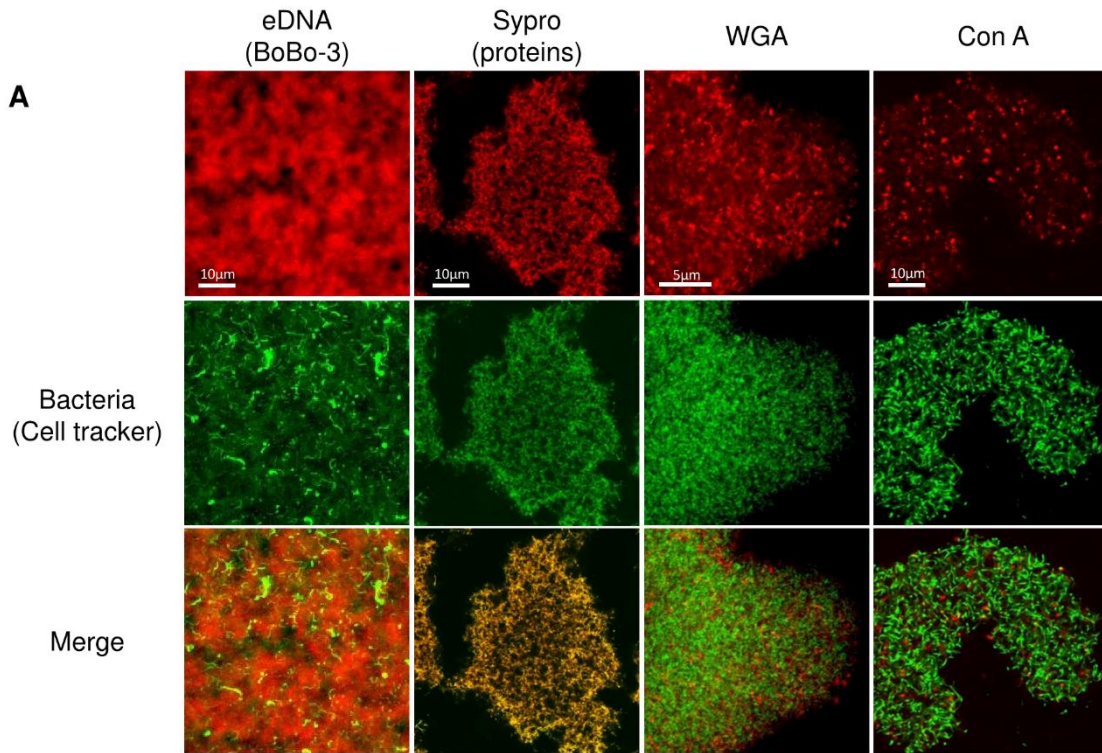
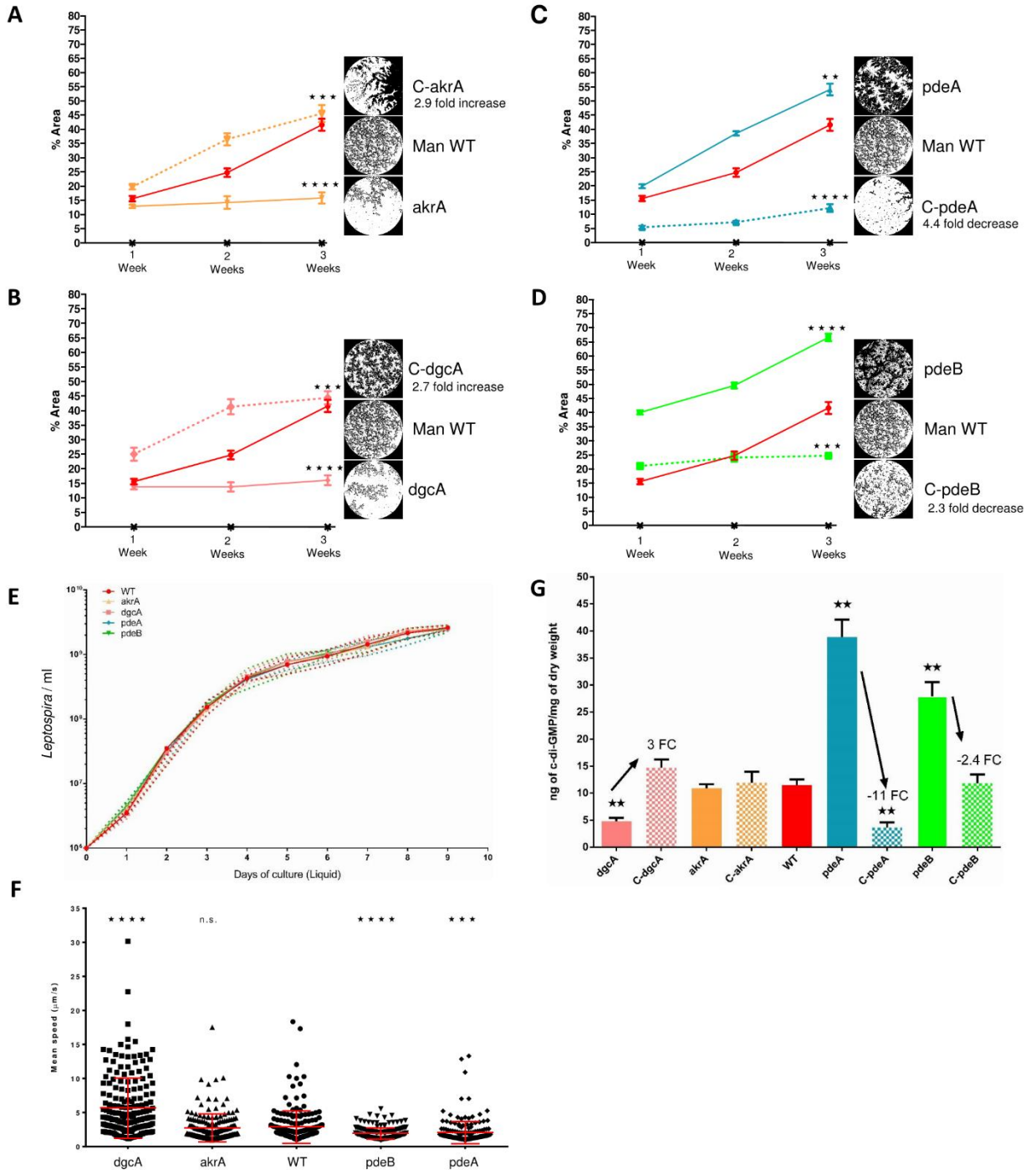
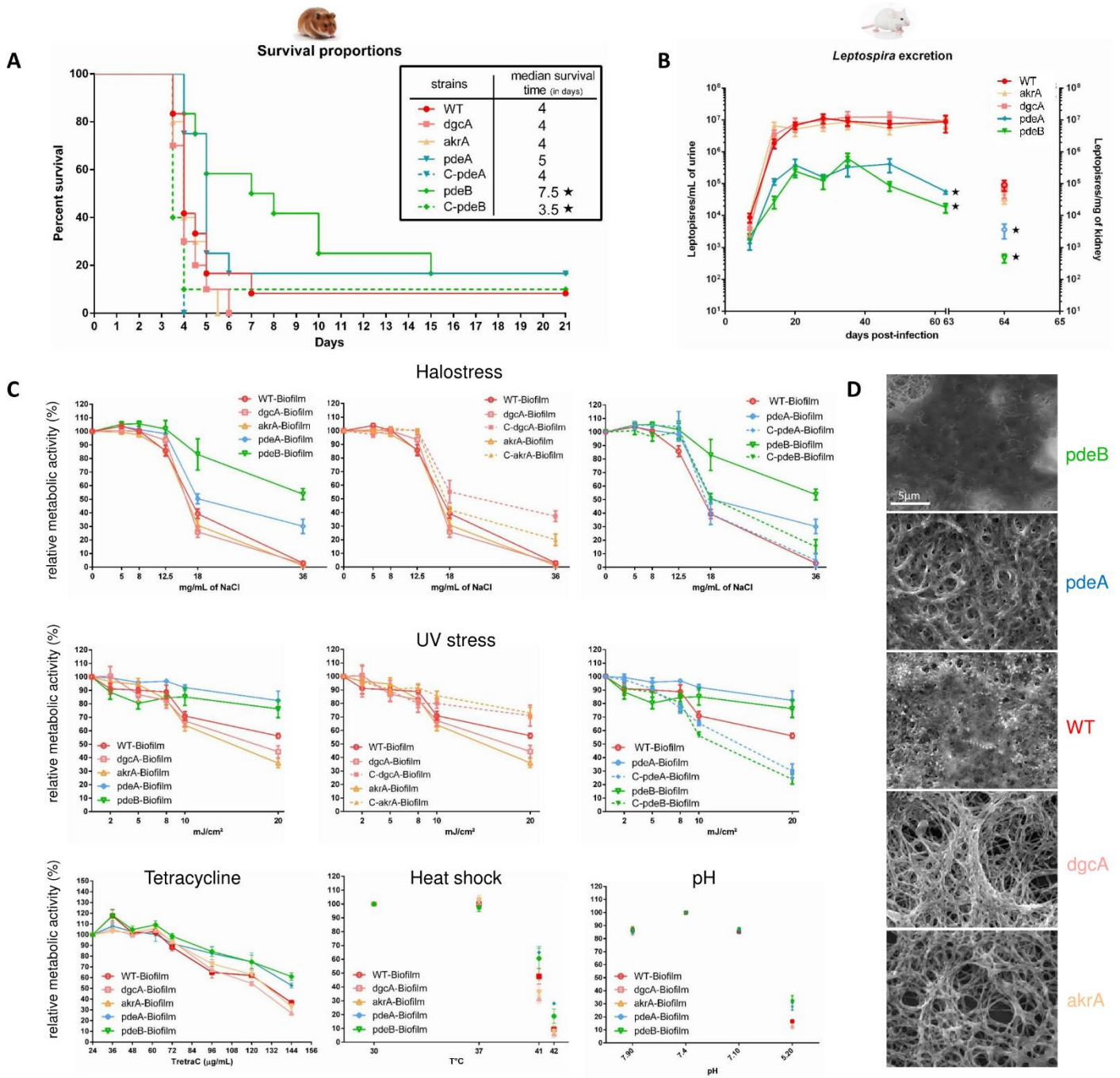


Fig.2









Bibliographie

1. Liddy KA, White MY, Cordwell SJ. Functional decorations: Post-translational modifications and heart disease delineated by targeted proteomics. *Genome Medicine*. 2013 Feb 28;5(2):20.
2. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta - Gen Subj*. 1999 Dec 17;1473(1):4–8.
3. Dell A, Galadari A, Sastre F, Hitchen P. Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int J Microbiol*. 2010;2010.
4. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, *et al.* How many human proteoforms are there? *Nature Chemical Biology*. 2018 Feb 14;14(3):206-214.
5. Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, *et al.* Symbol nomenclature for graphical representations of glycans. *Glycobiology*. 2015;25(12):1323–4.
6. Domon B, Costello CE. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj J*. 1988;5(4):397–409.
7. Roepstorff P, Fohlman J. Letter to the editors - Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biol Mass Spectrom*. 1984 Nov;11(11):601.
8. Higel F, Seidl A, Sörgel F, Friess W. N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *European Journal of Pharmaceutics and Biopharmaceutics*. 2016 Mar;100:94-100.

9. Zheng K, Bantog C, Bayer R. The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs*. 2011 Nov-Dec;3(6):568-76.
10. Wuhrer M, Selman MHJ, McDonnell LA, Kämpfel T, Derfuss T, Khademi M, *et al.* Pro-inflammatory pattern of IgG1 Fc glycosylation in multiple sclerosis cerebrospinal fluid. *J Neuroinflammation*. 2015 Dec 18;12:235.
11. Arnold JN, Wormald MR, Sim RB, Rudd PM, Dwek RA. The Impact of Glycosylation on the Biological Function and Structure of Human Immunoglobulins. *Annu Rev Immunol*. 2007;25:21-50.
12. Schroeder HW, Cavacini L. Structure and function of immunoglobulins. *J Allergy Clin Immunol*. 2010 Feb;125(2 Suppl 2):S41-52.
13. Itzkowitz SH, Yuan M, Montgomery CK, Kjeldsen T, Takahashi HK, Bigbee WL, *et al.* Expression of Tn, sialosyl-Tn, and T antigens in human colon cancer. *Cancer Res*. 1989 Jan 1;49(1):197-204.
14. Springer GF. Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy. *J Mol Med (Berl)*. 1997 Aug;75(8):594-602.
15. Bresalier RS, Ho SB, Schoeppner HL, Kim YS, Sleisenger MH, Brodt P, *et al.* Enhanced Sialylation of Mucin-Associated Carbohydrate Structures in Human Colon Cancer Metastasis. *GASTROENTEROLOGY*. 1996 May ;110(5) :1354-1367.
16. Park JJ, Lee M. Increasing the α 2, 6 sialylation of glycoproteins may contribute to metastatic spread and therapeutic resistance in colorectal cancer. *Gut Liver*. 2013 Nov;7(6):629-41.
17. Oliveira-Ferrer L, Legler K, Milde-Langosch K. Role of protein glycosylation in cancer metastasis. *Semin Cancer Biol*. 2017 Jun;44:141-152.

18. More SK, Srinivasan N, Budnar S, Bane SM, Upadhya A, Thorat RA, *et al.* N-glycans and metastasis in galectin-3 transgenic mice. *Biochem Biophys Res Commun.* 2015 May 1;460(2):302–7.
19. Cole LA, Butler SA. Hyperglycosylated human chorionic gonadotropin and human chorionic gonadotropin free beta-subunit: tumor markers and tumor promoters. *J Reprod Med.* 2008 Jul;53(7):499–512.
20. Stowell SR, Ju T, Cummings RD. Protein Glycosylation in Cancer. *Annu Rev Pathol Mech Dis.* 2015 Jan 26;10(1):473–510.
21. Magon N, Kumar P. Hormones in pregnancy. *Niger Med J.* 2013;53(4):179.
22. Neulen J, Raczek S, Pogorzelski M, Grunwald K, Yeo TK, Dvorak HF, *et al.* Secretion of vascular endothelial growth factor/vascular permeability factor from human luteinized granulosa cells is human chorionic gonadotrophin dependent. *Mol Hum Reprod.* 1998;4(3):203–6.
23. Yagel S, Geva TE, Solomon H, Shimonovitz S, Reich R, Finci-Yeheskel Z, *et al.* High levels of human chorionic gonadotropin retard first trimester trophoblast invasion in vitro by decreasing urokinase plasminogen activator and collagenase activities. *J Clin Endocrinol Metab.* 1993 Dec;77(6):1506-11.
24. Schumacher A, Heinze K, Witte J, Poloski E, Linzke N, Woidacki K, *et al.* Human Chorionic Gonadotropin as a Central Regulator of Pregnancy Immune Tolerance. *J Immunol.* 2013;190(6):2650–8.
25. Blithe DL. N-linked oligosaccharides on free α interfere with its ability to combine with human chorionic gonadotropin- β subunit. *J Biol Chem.* 1990 Dec 15;265(35):21951-6.
26. Cahoreau C, Klett D, Combarnous Y. Structure-function relationships of glycoprotein hormones and their subunits' ancestors. *Front Endocrinol (Lausanne).* 2015 Feb 26;6:26.

27. Thotakura NR, Bahl OP. Role of carbohydrate in human chorionic gonadotropin: Deglycosylation uncouples hormone-receptor complex and adenylate cyclase system. *Biochem Biophys Res Commun.* 1982 Sep 16;108(1):399-405.
28. Koistinen H, Koel M, Peters M, Rincken A, Lundin K, Tuuri T, *et al.* Hyperglycosylated hCG activates LH/hCG-receptor with lower activity than hCG. *Mol Cell Endocrinol.* 2019 Jan 5;479:103-109.
29. Fournier T, Guibourdenche J, Evain-Brion D. Review: hCGs: Different sources of production, different glycoforms and functions. *Placenta.* 2015;36(S1):S60–5.
30. AFSAPPS. Les médicaments inducteurs de l'ovulation-les gonadotrophines LES MEDICAMENTS INDUCTEURS DE L'OVULATION : LES GONADOTROPHINES RECOMMANDATIONS Actualisation 2007.
31. Eftekhari M, Khalili MA, Rahmani E. The efficacy of recombinant versus urinary HCG in ART outcome. *Iran J Reprod Med.* 2012 Nov;10(6):543-8.
32. Gautam A. Recombinant or urinary human chorionic gonadotropin in ovulation induction? *J Obstet Gynaecol India.* 2011 Dec; 61(6): 621–623.
33. Nwabuobi C, Arlier S, Schatz F, Guzeloglu-Kayisli O, Lockwood CJ, Kayisli UA. hCG: Biological functions and clinical applications. *Int J Mol Sci.* 2017;18(10):1–15.
34. Elliott MM, Kardana A, Lustbader JW, Cole LA. Carbohydrate and peptide structure of the α - and β -subunits of human chorionic gonadotropin from normal and aberrant pregnancy and choriocarcinoma. *Endocrine.* 1997;7(1):15–32.
35. Clerc F, Reiding KR, Jansen BC, Kammeijer GSM, Bondt A, Wuhrer M. Human plasma protein N-glycosylation. *Glycoconj J.* 2016;33(3):309–43.

36. Liu J, Dolikun M, Štambuk J, Trbojević-Akmačić I, Zhang J, Wang H, *et al.* The association between subclass-specific IgG Fc N-glycosylation profiles and hypertension in the Uygur, Kazak, Kirgiz, and Tajik populations. *J Hum Hypertens.* 2018 Sep 1;32(8–9):555–63.
37. Menzer C, Schams D. Radioimmunoassay for PMSG and its application to in-vivo studies. *Reproduction.* 1979;55(2):339–45.
38. Sherman GB, Wolfe MW, Farmerie TA, Clay CM, Threadgill DS, Sharp DC, *et al.* A single gene encodes the beta-subunits of equine luteinizing hormone and chorionic gonadotropin. *Mol Endocrinol.* 1992 Jun;6(6):951-9.
39. Dausset J, Moullec J, Bernard J. Acquired hemolytic anemia with polyagglutinability of red blood cells due to a new factor present in normal human serum (Anti-Tn). *Blood.* 1959 Oct;14:1079-93.
40. Strous GJ, Dekker J. Mucin-type glycoproteins. *Crit Rev Biochem Mol Biol.* 1992;27(1-2):57-92.
41. Jensen PH, Kolarich D, Packer NH. Mucin-type O-glycosylation - Putting the pieces together. *FEBS J.* 2010 Jan;277(1):81-94.
42. Hounsell EF, Davies MJ, Renouf D V. O-linked protein glycosylation structure and function. *Glycoconjugate Journal.* 1996 Feb;13(1):19-26.
43. Taniguchi N, Honke K, Fukuda M, Narimatsu H, Yamaguchi Y, Angata T. Handbook of glycosyltransferases and related genes, second edition. *Handbook of Glycosyltransferases and Related Genes, Second Edition.* 2014.
44. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, *et al.* *Essentials of glycobiology*, third edition. Cold Spring Harbor Laboratory Press. 2017.
45. Niwa Y, Suzuki T, Dohmae N, Simizu S. Identification of DPY19L3 as the C -mannosyltransferase of R-spondin1 in human cells . *Mol Biol Cell.* 2016;27(5):744–56.

46. Tian R, Zhang H, Chen H, Liu G, Wang Z. Uncovering the Binding Specificities of Lectins with Cells for Precision Colorectal Cancer Diagnosis Based on Multimodal Imaging. *Adv Sci*. 2018;5(6):1–9.
47. Haab BB. Using lectins in biomarker research: Addressing the limitations of sensitivity and availability. *Proteomics Clin Appl*. 2012 Aug;6(7-8):346-50.
48. Schubert M, Walczak MJ, Aebi M, Wider G. Posttranslational modifications of intact proteins detected by NMR spectroscopy: Application to glycosylation. *Angew Chem Int Ed Engl*. 2015 Jun 8;54(24):7096-100.
49. Ardá A, Jiménez-Barbero J. The recognition of glycans by protein receptors. Insights from NMR spectroscopy. *Chem Commun (Camb)*. 2018 May 8;54(38):4761-4769.
50. Jiménez-Barbero J, Canales A, Boos I, Perkams L, Karst L, Lubert T, *et al*. Breaking the Limits in Analyzing Carbohydrate Recognition by NMR Spectroscopy: Resolving Branch-Selective Interaction of a Tetra-Antennary N-Glycan with Lectins. *Angew Chemie - Int Ed*. 2017;56(47):14987–91.
51. Takahashi N. Demonstration of a new amidase acting on glycopeptides. *Biochem Biophys Res Commun*. 1977 Jun 20;76(4):1194-201.
52. Plummer TH, Elder JH, Alexander S, Phelan AW, Tarentino AL. Demonstration of peptide:N-glycosidase F activity in endo- β -N-acetylglucosaminidase F preparations. *J Biol Chem*. 1984 Sep 10;259(17):10700-4.
53. Norris GE, Stillman TJ, Anderson BF, Baker EN. The three-dimensional structure of PNGase F, a glycosyl asparaginase from *Flavobacterium meningosepticum*. *Structure*. 1994 Nov 15;2(11):1049-59.
54. Altmann F, Schweiszer S, Weber C. Kinetic comparison of peptide: N-glycosidases F and A reveals several differences in substrate specificity. *Glycoconj J*. 1995 Feb;12(1):84-93.

55. Tretter V, Altmann F, März L. Peptide - N₄ - (N - acetyl - β - glucosaminy)asparagine amidase F cannot release glycans with fucose attached α 1 \rightarrow 3 to the asparagine - linked N - acetylglucosamine residue. *Eur J Biochem.* 1991 Aug 1;199(3):647-52.
56. Wilson IBH, Zeleny R, Kolarich D, Staudacher E, Stroop CJM, Kamerling JP, *et al.* Analysis of asn-linked glycans from vegetable foodstuffs: Widespread occurrence of Lewis a, core α 1,3-linked fucose and xylose substitutions. *Glycobiology.* 2001;11(4):261-74.
57. Sun G, Yu X, Bao C, Wang L, Li M, Gan J, *et al.* Identification and characterization of a novel prokaryotic peptide: N-Glycosidase from *Elizabethkingia meningoseptica*. *J Biol Chem.* 2015 Mar 20;290(12):7452-62.
58. Wang T, Cai ZP, Gu XQ, Ma HY, Du YM, Huang K, *et al.* Discovery and characterization of a novel extremely acidic bacterial N -glycanase with combined advantages of PNGase F and A . *Biosci Rep.* 2014;34(6):673-84.
59. Skoza L, Mohos S. Stable thiobarbituric acid chromophore with dimethyl sulphoxide. Application to sialic acid assay in analytical de- O - acetylation . *Biochem J.* 1976 Dec 1; 159(3): 457-462.
60. Plummer TH Jr, Phelan AW, Tarentino AL. Detection and quantification of peptide-N₄-(N-acetyl-beta-glucosaminy)asparagine amidases. *Eur J Biochem.* 1987 Feb 16;163(1):167-73.
61. Ftouhi-Paquin N, Hauer CR, Stack RF, Tarentino AL, Plummer TH. Molecular cloning, primary structure, and properties of a new glycoamidase from the fungus *Aspergillus tubigenis*. *J Biol Chem.* 1997 Sep 5;272(36):22960-5.
62. Hossain MA, Nakano R, Nakamura K, Kimura Y. Molecular identification and characterization of an acidic peptide:N-glycanase from tomato (*Lycopersicon esculentum*) fruits. *J Biochem.* 2010 Feb;147(2):157-65.

63. Küster B, Wheeler SF, Hunter AP, Dwek RA, Harvey DJ. Sequencing of N-linked oligosaccharides directly from protein gels: In-gel deglycosylation followed by matrix-assisted laser desorption/ionization mass spectrometry and normal-phase high-performance liquid chromatography. *Anal Biochem.* 1997 Jul 15;250(1):82-101.
64. Wilson NL, Schulz BL, Karlsson NG, Packer NH. Sequential analysis of N- and O-linked glycosylation of 2D-PAGE separated glycoproteins. *J Proteome Res.* 2002 Nov-Dec;1(6):521-9.
65. Naldi M, Tramarin A, Bartolini M. Immobilized enzyme-based analytical tools in the -omics era: Recent advances. *J Pharm Biomed Anal.* 2018 Oct 25;160:222-237.
66. Zhou H, Briscoe AC, Froehlich JW, Lee RS. PNGase F catalyzes de-N-glycosylation in a domestic microwave. *Anal Biochem.* 2012 Aug 1;427(1):33-5.
67. Kalay H, Ambrosini M, Chiodo F, van Kooyk Y, García-Vallejo JJ. Enhanced glycan nanoprofiling by weak anion exchange preparative chromatography, mild acid desialylation, and nanoliquid chromatography-mass spectrometry with nanofluorescence detection. *Electrophoresis.* 2013 Aug;34(16):2350-6.
68. Varki A, Freeze HH, Manzi AE. Preparation and Analysis of Glycoconjugates. *Current Protocols in Molecular Biology.* 2009 Oct;88(1):17.0.1-17.0.12.
69. Cooper CA, Packer NH, Redmond JW. The elimination of O-linked glycans from glycoproteins under non-reducing conditions. *Glycoconj J.* 1994 Apr;11(2):163-7.
70. Kang P, Mechref Y, Novotny M V. High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Commun Mass Spectrom.* 2008;22(5):721-34.

71. Shajahan A, Supekar N, Heiss C, Azadi P. High-Throughput Automated Micro-permethylation for Glycan Structure Analysis. *Anal Chem.* 2019 Jan 15;91(2):1237-1240.
72. Harvey DJ. Derivatization of carbohydrates for analysis by chromatography; electrophoresis and mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2011 May 15;879(17-18):1196-225.
73. Naven TJP, Harvey DJ. Cationic derivatization of oligosaccharides with Girard's T reagent for improved performance in matrix-assisted laser desorption/ionization and electrospray mass spectrometry. *Rapid Commun Mass Spectrom.* 1996;10(7):829-34.
74. Rohmer M, Meyer B, Mank M, Stahl B, Bahr U, Karas M. 3-Aminoquinoline Acting As Matrix and Derivatizing Agent for Maldi Ms Analysis of Oligosaccharides. *Anal Chem.* 2010;82(9):3719-26.
75. Wheeler SF, Domann P, Harvey DJ. Derivatization of sialic acids for stabilization in matrix-assisted laser desorption/ionization mass spectrometry and concomitant differentiation of $\alpha(2 \rightarrow 3)$ - and $\alpha(2 \rightarrow 6)$ -isomers. *Rapid Commun Mass Spectrom.* 2009 Jan 30;23(2):303-312.
76. Kailemia MJ, Ruhaak LR, Lebrilla CB, Amster IJ. Oligosaccharide analysis by mass spectrometry: A review of recent developments. *Anal Chem.* 2014 Jan 7;86(1):196-212.
77. Kronewitter SR, De Leoz MLA, Peacock KS, McBride KR, An HJ, Miyamoto S, *et al.* Human serum processing and analysis methods for rapid and reproducible N-glycan mass profiling. *J Proteome Res.* 2010 Oct 1; 9(10): 4952-4959.
78. Verostek MF, Lubowski C, Trimble RB. Selective organic precipitation/extraction of released N-glycans following large-scale enzymatic deglycosylation of glycoproteins. *Anal Biochem.* 2000 Feb 15;278(2):111-22.

79. Totten SM, Kullolli M, Pitteri SJ. Multi-lectin affinity chromatography for separation, identification, and quantitation of intact protein glycoforms in complex biological mixtures. *Methods Mol Biol.* 2017;1550:99-113.
80. Qiu R, Regnier FE. Use of multidimensional lectin affinity chromatography in differential glycoproteomics. *Anal Chem.* 2005 May 1;77(9):2802-9.
81. Mao X, Qin J, Lin B. Miniaturized Lectin Affinity Chromatography. *Lectins: Analytical Technologies.* 2007 Jul 18;213-238.
82. Madera M, Mann B, Mechref Y, Novotny MV. Efficacy of glycoprotein enrichment by microscale lectin affinity chromatography. *J Sep Sci.* 2008 Aug; 31(14): 2722–2732.
83. Gbormittah FO, Lee LY, Taylor K, Hancock WS, Iliopoulos O. Comparative studies of the proteome, glycoproteome, and N-glycome of clear cell renal cell carcinoma plasma before and after curative nephrectomy. *J Proteome Res.* 2014;13(11):4889–900.
84. Keogh D, Thompson R, Larragy R, McMahon K, O'Connell M, O'Connor B, *et al.* Generating novel recombinant prokaryotic lectins with altered carbohydrate binding properties through mutagenesis of the PA-IL protein from *Pseudomonas aeruginosa*. *Biochim Biophys Acta.* 2014 Jun;1840(6):2091-104.
85. Hirabayashi J, Yamada M, Kuno A, Tateno H. Lectin microarrays: Concept, principle and applications. *Chem Soc Rev.* 2013 May 21;42(10):4443-58.
86. Ruhaak LR, Xu G, Li Q, Goonatileke E, Lebrilla CB. Mass Spectrometry Approaches to Glycomic and Glycoproteomic Analyses. *Chem Rev.* 2018 Sep 12;118(17):7886-7930.
87. Vreeker GCM, Wuhrer M. Reversed-phase separation methods for glycan analysis. *Anal Bioanal Chem.* 2017 Jan;409(2):359-378.

88. Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J Chromatogr.* 1990 Jan 19;499:177-96.
89. Hemström P, Irgum K. Hydrophilic interaction chromatography. *J. Sep. Sci.* 2006. 29;1784-1821
90. Selman MHJ, Hemayatkar M, Deelder AM, Wührer M. Cotton HILIC SPE microtips for microscale purification and enrichment of glycans and glycopeptides. *Anal Chem.* 2011 Apr 1;83(7):2492-9
91. Ruhaak LR, Huhn C, Waterreus WJ, De Boer AR, Neusüss C, Hokke CH, *et al.* Hydrophilic interaction chromatography-based high-throughput sample preparation method for N-glycan analysis from total human plasma glycoproteins. *Anal Chem.* 2008 Aug 1;80(15):6119-26.
92. Lauber MA, Yu YQ, Brousmiche DW, Hua Z, Koza SM, Magnelli P, *et al.* Rapid preparation of released N-glycans for HILIC analysis using a labeling reagent that facilitates sensitive fluorescence and ESI-MS detection. *Anal. Chem.* 2015 Apr 15;87(10):5401-5409
93. Ying QY, Gilar M, Kaska J, Gebler JC. A rapid sample preparation method for mass spectrometric characterization of N-linked glycans. *Rapid Commun Mass Spectrom.* 2005;19(16):2331-6.
94. Aich U, Liu A, Lakbub J, Mozdzanowski J, Byrne M, Shah N, *et al.* An Integrated Solution-Based Rapid Sample Preparation Procedure for the Analysis of N-Glycans from Therapeutic Monoclonal Antibodies. *J Pharm Sci.* 2016 Mar;105(3):1221-32.
95. Mancera-Artu M, Giménez E, Barbosa J, Sanz-Nebot V. Identification and characterization of isomeric N-glycans of human alpha-acid-glycoprotein by stable isotope labelling and ZIC-HILIC-MS in combination with exoglycosidase digestion. *Anal Chim Acta.* 2016 Oct 12;940:92-103.

96. Giménez E, Balmaña M, Figueras J, Fort E, Bolós C de, Sanz-Nebot V, *et al.* Quantitative analysis of N-glycans from human alfa-acid-glycoprotein using stable isotope labeling and zwitterionic hydrophilic interaction capillary liquid chromatography electrospray mass spectrometry as tool for pancreatic disease diagnosis. *Anal Chim Acta.* 2015 Mar 25;866:59-68.
97. Veillon L, Huang Y, Peng W, Dong X, Cho BG, Mechref Y. Characterization of isomeric glycan structures by LC-MS/MS. *Electrophoresis.* 2017 Sep;38(17):2100-2114.
98. Badhan A, Wang Y, McAllister TA. Analysis of Complex Carbohydrate Composition in Plant Cell Wall Using Fourier Transformed Mid-Infrared Spectroscopy (FT-IR). *Methods Mol Biol.* 2017;1588:209-214.
99. Wang JR, Gao WN, Grimm R, Jiang S, Liang Y, Ye H, *et al.* A method to identify trace sulfated IgG N-glycans as biomarkers for rheumatoid arthritis. *Nat Commun.* 2017 Sep 20;8(1):631.
100. Stavenhagen K, Hinneburg H, Kolarich D, Wührer M. Site-specific N- and O-glycopeptide analysis using an integrated C18-PGC-LC-ESI-QTOF-MS/MS approach. *Methods Mol Biol.* 2017;1503:109-119.
101. Wang H, Chen X, Zhang X, Zhang W, Li Y, Yin H, *et al.* Comparative Assessment of Glycosylation of a Recombinant Human FSH and a Highly Purified FSH Extracted from Human Urine. *J Proteome Res.* 2016;15(3):923–32.
102. West C, Elfakir C, Lafosse M. Porous graphitic carbon: A versatile stationary phase for liquid chromatography. *J Chromatogr A.* 2010 May 7;1217(19):3201-16.
103. Polyakova YL, Row KH. Retention of Some Five-Membered Heterocyclic Compounds on a Porous Graphitized Carbon, Hypercarb™. *Chromatographia.* 2007 Jan;65(1-2):59–63.
104. Hanai T. Separation of polar compounds using carbon columns. *J Chromatogr A.* 2003 Mar 14;989(2):183-96.

105. Walker SH, Carlisle BC, Muddiman DC. Systematic comparison of reverse phase and hydrophilic interaction liquid chromatography platforms for the analysis of N-linked glycans. *Anal Chem.* 2012 Oct 2;84(19):8198-206.
106. Mauko L, Lacher NA, Pelzing M, Nordborg A, Haddad PR, Hilder EF. Comparison of ZIC-HILIC and graphitized carbon-based analytical approaches combined with exoglycosidase digestions for analysis of glycans from monoclonal antibodies. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2012 Dec 12;911:93-104.
107. Mauko L, Nordborg A, Hutchinson JP, Lacher NA, Hilder EF, Haddad PR. Glycan profiling of monoclonal antibodies using zwitterionic-type hydrophilic interaction chromatography coupled with electrospray ionization mass spectrometry detection. *Anal Biochem.* 2011 Jan 15;408(2):235-41.
108. Reusch D, Habegger M, Maier B, Maier M, Kloseck R, Zimmermann B, *et al.* Comparison of methods for the analysis of therapeutic immunoglobulin G Fc-glycosylation profiles - Part 1: Separation-based methods. *MAbs.* 2015;7(1):167-79.
109. Melmer M, Stangler T, Premstaller A, Lindner W. Comparison of hydrophilic-interaction, reversed-phase and porous graphitic carbon chromatography for glycan analysis. *J Chromatogr A.* 2011 Jan 7;1218(1):118-23.
110. Stephens E, Maslen SL, Green LG, Williams DH. Fragmentation Characteristics of Neutral N-Linked Glycans Using a MALDI-TOF/TOF Tandem Mass Spectrometer. *Anal Chem.* 2004 Apr 15;76(8):2343-54.
111. Fukuyama Y. Liquid Matrices in MALDI-MS. *Applications of Ionic Liquids in Science and Technology*, Handy S (Ed.). 2011 Sep 22;6:361-374.
112. Tholey A, Heinzle E. Ionic (liquid) matrices for matrix-assisted laser desorption/ionization mass spectrometry-applications and perspectives. *Anal Bioanal Chem.* 2006 Sep;386(1):24-37.

113. Fukuyama Y, Nakaya S, Yamazaki Y, Tanaka K. Ionic liquid matrixes optimized for MALDI-MS of sulfated/sialylated/neutral oligosaccharides and glycopeptides. *Anal Chem.* 2008 Mar 15;80(6):2171-9.
114. Laremore TN, Zhang F, Linhardt RJ. Ionic liquid matrix for direct UV-MALDI-TOF-MS analysis of dermatan sulfate and chondroitin sulfate oligosaccharides. *Anal Chem.* 2007 Feb 15;79(4):1604-10.
115. Fukuyama Y, Funakoshi N, Takeyama K, Hioki Y, Nishikaze T, Kaneshiro K, *et al.* 3-aminoquinoline/p-coumaric acid as a MALDI matrix for glycopeptides, carbohydrates, and phosphopeptides. *Anal Chem.* 2014;86(4):1937-42.
116. Iwamoto S, On MLW, Ying LP, Noor SM, Takaki K, Kotani E. Mass spectrometric profiling of N-linked glycan of recombinant Newcastle disease virus F and HN proteins produced by insect cell expression system. *J Insect Biotechnol Sericology.* 2018;87(1):1_001-1_008.
117. Kawaguchi-Sakita N, Kaneshiro-Nakagawa K, Kawashima M, Sugimoto M, Tokiwa M, Suzuki E, *et al.* Serum immunoglobulin G Fc region N-glycosylation profiling by matrix-assisted laser desorption/ionization mass spectrometry can distinguish breast cancer patients from cancer-free controls. *Biochem Biophys Res Commun.* 2016 Jan 22;469(4):1140-5.
118. Nishikaze T, Tsumoto H, Sekiya S, Iwamoto S, Miura Y, Tanaka K. Differentiation of Sialyl Linkage Isomers by One-Pot Sialic Acid Derivatization for Mass Spectrometry-Based Glycan Profiling. *Anal Chem.* 2017 Feb 21;89(4):2353-2360.
119. Thennati R, Singh SK, Nage N, Patel Y, Bose SK, Burade V, *et al.* Analytical characterization of recombinant hCG and comparative studies with reference product. *Biologics.* 2018 Jan 30;12:23-35.
120. Grunow D, Blanchard V. Enzymatic Release of Glycoprotein N-Glycans and Fluorescent Labeling. *Methods Mol Biol.* 2019;1934:43-49.

121. Glyco-engineering: Methods and protocols. Castilho A (Ed.). *Methods in Molecular Biology*. 2015;1321.
122. Ashwood C, Lin CH, Thaysen-Andersen M, Packer NH. Discrimination of Isomers of Released N- and O-Glycans Using Diagnostic Product Ions in Negative Ion PGC-LC-ESI-MS/MS. *J Am Soc Mass Spectrom*. 2018 Jun;29(6):1194-1209.
123. Hofmeister GE, Zhou Z, Leary JA. Linkage Position Determination in Lithium-Cationized Disaccharides: Tandem Mass Spectrometry and Semiempirical Calculations. *J Am Chem Soc*. 1991 Jul 1;113(16):5964-5970
124. Zaia J. Mass Spectrometry and the Emerging Field of Glycomics. *Chem Biol*. 2008 Sep 22;15(9):881-92.
125. Zhou W, Håkansson K. Electron capture dissociation of divalent metal-adducted sulfated N-glycans released from bovine thyroid stimulating hormone. *J Am Soc Mass Spectrom*. 2013 Nov;24(11):1798-806.
126. Yu X, Jiang Y, Chen Y, Huang Y, Costello CE, Lin C. Detailed glycan structural characterization by electronic excitation dissociation. *Anal Chem*. 2013 Nov 5;85(21):10017-21.
127. Harvey DJ, Scarff CA, Edgeworth M, Pagel K, Thalassinou K, Struwe WB, *et al*. Travelling-wave ion mobility mass spectrometry and negative ion fragmentation of hybrid and complex N-glycans. *J Mass Spectrom*. 2016 Nov;51(11):1064-1079.
128. Pu Y, Ridgeway ME, Glaskin RS, Park MA, Costello CE, Lin C. Separation and Identification of Isomeric Glycans by Selected Accumulation-Trapped Ion Mobility Spectrometry-Electron Activated Dissociation Tandem Mass Spectrometry. *Anal Chem*. 2016 Apr 5;88(7):3440-3.
129. Hinneburg H, Hofmann J, Struwe WB, Thader A, Altmann F, Varón Silva D, *et al*. Distinguishing N-acetylneuraminic acid linkage isomers on

glycopeptides by ion mobility-mass spectrometry. *Chem Commun (Camb)*. 2016 Mar 21;52(23):4381-4.

130. Moh ESX, Thaysen-Andersen M, Packer NH. Relative versus absolute quantitation in disease glycomics. *Proteomics - Clinical Applications*. 2015;9(3-4):368-382.

131. Chen Z, Zhong X, Tie C, Chen B, Zhang X, Li L. Development of a hydrophilic interaction liquid chromatography coupled with matrix-assisted laser desorption/ionization-mass spectrometric imaging platform for N-glycan relative quantitation using stable-isotope labeled hydrazide reagents. *Anal Bioanal Chem*. 2017 Jul;409(18):4437-4447.

132. Liang HC, Lahert E, Pike I, Ward M. Quantitation of protein post-translational modifications using isobaric tandem mass tags. *Bioanalysis*. 2015;7(3):383-400.

133. Bahl OP. Human chorionic gonadotropin. I. Purification and physicochemical properties. *J Biol Chem*. 1969 Feb 25;244(4):567-74.

134. Bell JJ, Canfield RE, Sciarra JJ. Purification and Characterization of Human Chorionic Gonadotropin. *Endocrinology*. 1969 Feb 1;84(2):298-307.

135. Khademi F, Hamzehee K, Mostafaie A, Hajihossaini R. Purification of three major forms of β -hCG from urine and production of polyclonal antibodies against them. *Clin Biochem*. 2009 Sep;42(13-14):1476-82.

136. Rodríguez AM, Rodríguez OZ, Conde IB, Del Pino YR, Pérez LH, Cisnero RL, *et al*. Purification of Human Chorionic Gonadotropin from Pregnancy Urine by Immunoaffinity Chromatography Using a Monoclonal Antibody Anti- β Chain hCG. *Hybridoma (Larchmt)*. 2005 Oct;24(5):258-62.

137. Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev*. 2013 Apr 10;113(4):2343-94.

138. Shevchenko A, Tomas H, Havliš J, Olsen J V., Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc.* 2006;1(6):2856-60.
139. Fournier ML, Gilmore JM, Martin-Brown SA, Washburn MP. Multidimensional separations-based shotgun proteomics. *Chem Rev.* 2007 Aug;107(8):3654-86.
140. Ongay S, Boichenko A, Govorukhina N, Bischoff R. Glycopeptide enrichment and separation for protein glycosylation analysis. *J Sep Sci.* 2012 Sep;35(18):2341-72.
141. Mysling S, Palmisano G, Hojrup P, Thaysen-Andersen M. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal Chem.* 2010 Jul 1;82(13):5598-609.
142. Jensen PH, Mysling S, Højrup P, Jensen ON. Glycopeptide enrichment for MALDI-TOF mass spectrometry analysis by hydrophilic interaction liquid chromatography solid phase extraction (HILIC SPE). *Methods Mol Biol.* 2013;951:131-44.
143. Alagesan K, Khilji SK, Kolarich D. It is all about the solvent: on the importance of the mobile phase for ZIC-HILIC glycopeptide enrichment. *Anal Bioanal Chem.* 2017 Jan;409(2):529-538.
144. Cao L, Qu Y, Zhang Z, Wang Z, Prytkova I, Wu S. Intact glycopeptide characterization using mass spectrometry. *Expert Rev Proteomics.* 2016 May;13(5):513-22.
145. Wohlgemuth J, Karas M, Eichhorn T, Hendriks R, Andrecht S. Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. *Anal Biochem.* 2009 Dec 15;395(2):178-88.

146. Yang W, Shah P, Hu Y, Toghi Eshghi S, Sun S, Liu Y, *et al.* Comparison of Enrichment Methods for Intact N- and O-Linked Glycopeptides Using Strong Anion Exchange and Hydrophilic Interaction Liquid Chromatography. *Anal Chem.* 2017 Nov 7;89(21):11193-11197.
147. Perchepied S, Eskenazi N, Giangrande C, Camperi J, Fournier T, Vinh J, *et al.* Development of Immobilized Enzyme Reactors for the characterization of the glycosylation heterogeneity of a protein. *Talanta.* 2020 Jan 1;206:120171.
148. Liu X, Chan K, Chu IK, Li J. Microwave-assisted nonspecific proteolytic digestion and controlled methylation for glycomics applications. *Carbohydr Res.* 2008 Nov 24;343(17):2870-7.
149. Clowers BH, Dodds ED, Seipert RR, Lebrilla CB. Site determination of protein glycosylation based on digestion with immobilized nonspecific proteases and fourier transform ion cyclotron resonance mass spectrometry. *J Proteome Res.* 2007 Oct;6(10):4032-40.
150. Froehlich JW, Barboza M, Chu C, Lerno LA, Clowers BH, Zivkovic AM, *et al.* Nano-LC-MS/MS of glycopeptides produced by nonspecific proteolysis enables rapid and extensive site-specific glycosylation determination. *Anal Chem.* 2011 Jul 15;83(14):5541-7.
151. Dodds ED, Seipert RR, Clowers BH, German JB, Lebrilla CB. Analytical performance of immobilized pronase for glycopeptide footprinting and implications for surpassing reductionist glycoproteomics. *J Proteome Res.* 2009 Feb;8(2):502-12.
152. Zauner G, Koeleman CAM, Deelder AM, Wuhrer M. Protein glycosylation analysis by HILIC-LCMS of Proteinase K-generated N- and Oglycopeptides. *J Sep Sci.* 2010 Mar;33(6-7):903-10.
153. Hoffmann M, Pioch M, Pralow A, Hennig R, Kottler R, Reichl U, *et al.* The Fine Art of Destruction: A Guide to In-Depth Glycoproteomic Analyses—

Exploiting the Diagnostic Potential of Fragment Ions. *Proteomics*. 2018 Dec;18(24):e1800282.

154. Kozlik P, Goldman R, Sanda M. Study of structure-dependent chromatographic behavior of glycopeptides using reversed phase nanoLC. *Electrophoresis*. 2017 Sep;38(17):2193-2199.

155. Ji ES, Lee HK, Park GW, Kim KH, Kim JY, Yoo JS. Isomer separation of sialylated O- and N-linked glycopeptides using reversed-phase LC-MS/MS at high temperature. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2019 Mar 15;1110-1111:101-107.

156. Hua S, Nwosu CC, Strum JS, Seipert RR, An HJ, Zivkovic AM, *et al.* Site-specific protein glycosylation analysis with glycan isomer differentiation. *Anal Bioanal Chem*. 2012;403(5):1291-302.

157. Kozlik P, Goldman R, Sanda M. Hydrophilic interaction liquid chromatography in the separation of glycopeptides and their isomers. *Anal Bioanal Chem*. 2018 Aug;410(20):5001-5008.

158. Lu J, Fu D, Yu L, Cao C, Zou L, Liang X. Determination of N-Glycopeptides by Hydrophilic Interaction Liquid Chromatography and Porous Graphitized Carbon Chromatography with Mass Spectrometry Detection. *Anal Lett*. 2017;50(2):315-324.

159. Stavenhagen K, Plomp R, Wuhrer M. Site-Specific Protein N- and O-Glycosylation Analysis by a C18-Porous Graphitized Carbon-Liquid Chromatography-Electrospray Ionization Mass Spectrometry Approach Using Pronase Treated Glycopeptides. *Anal Chem*. 2015;87(23):11691-9.

160. Kammeijer GSM, Jansen BC, Kohler I, Heemskerk AAM, Mayboroda OA, Hensbergen PJ, *et al.* Sialic acid linkage differentiation of glycopeptides using capillary electrophoresis - Electrospray ionization - Mass spectrometry. *Sci Rep*. 2017 Jun 16;7(1):3733.

161. Khatri K, Klein JA, Haserick JR, Leon DR, Costello CE, McComb ME, *et al.* Microfluidic Capillary Electrophoresis-Mass Spectrometry for Analysis of Monosaccharides, Oligosaccharides, and Glycopeptides. *Anal Chem.* 2017 Jun 20;89(12):6645-6655.
162. Hofmann J, Pagel K. Glycan Analysis by Ion Mobility–Mass Spectrometry. *Angew Chemie - Int Ed.* 2017;56(29):8342–9.
163. Barroso A, Giménez E, Konijnenberg A, Sancho J, Sanz-Nebot V, Sobott F. Evaluation of ion mobility for the separation of glycoconjugate isomers due to different types of sialic acid linkage, at the intact glycoprotein, glycopeptide and glycan level. *J Proteomics.* 2018 Feb 20;173:22-31.
164. Reiding KR, Bondt A, Franc V, Heck AJR. The benefits of hybrid fragmentation methods for glycoproteomics. *TrAC - Trends in Analytical Chemistry.* 2018 Nov;108:260-268.
165. Ford KL, Zeng W, Heazlewood JL, Bacic A. Characterization of protein N-glycosylation by tandem mass spectrometry using complementary fragmentation techniques. *Front Plant Sci.* 2015 Aug 28;6:674.
166. Khatri K, Pu Y, Klein JA, Wei J, Costello CE, Lin C, *et al.* Comparison of Collisional and Electron-Based Dissociation Modes for Middle-Down Analysis of Multiply Glycosylated Peptides. *J Am Soc Mass Spectrom.* 2018 Jun 1;29(6):1075–85.
167. Dotz V, Haselberg R, Shubhakar A, Kozak RP, Falck D, Rombouts Y, *et al.* Mass spectrometry for glycosylation analysis of biopharmaceuticals. *TrAC - Trends in Analytical Chemistry.* 2015 Nov;73:1-9.
168. Tonie Wright H, Urry DW. Nonenzymatic deamidation of asparaginyl and glutaminyl residues in protein. *Crit Rev Biochem Mol Biol.* 1991;26(1):1–52.
169. Palmisano G, Melo-Braga MN, Engholm-Keller K, Parker BL, Larsen MR. Chemical deamidation: a common pitfall in large-scale N-linked

glycoproteomic mass spectrometry-based analyses. *J Proteome Res.* 2012 Mar 2;11(3):1949–57.

170. Zhu Z, Go EP, Desaire H. Absolute quantitation of glycosylation site occupancy using isotopically labeled standards and LC-MS. *J Am Soc Mass Spectrom.* 2014 Jun;25(6):1012–7.

171. Cao L, Diedrich JK, Kulp DW, Pauthner M, He L, Park S-KR, *et al.* Global site-specific N-glycosylation analysis of HIV envelope glycoprotein. *Nat Commun.* 2017;8:14954.

172. Sajic T, Liu Y, Aebersold R. Using data-independent, high-resolution mass spectrometry in protein biomarker research: Perspectives and clinical applications. *Proteomics Clin Appl.* 2015 Apr;9(3-4):307-21.

173. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: A critical review. *Anal Bioanal Chem.* 2007 Oct;389(4):1017–31.

174. Toby TK, Fornelli L, Kelleher NL. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu Rev Anal Chem.* 2016 Jun 12;9(1):499–519.

175. Frese CK, Altelaar AFM, van den Toorn H, Nolting D, Griep-Raming J, Heck AJR, *et al.* Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal Chem.* 2012 Nov 20;84(22):9668–73.

176. Brunner AM, Lössl P, Liu F, Huguet R, Mullen C, Yamashita M, *et al.* Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteoform characterization. *Anal Chem.* 2015 Apr 21;87(8):4152–8.

177. Li S, Zhou Y, Xiao K, Li J, Tian Z. Selective fragmentation of the N-glycan moiety and protein backbone of ribonuclease B on an Orbitrap Fusion Lumos

Tribrid mass spectrometer. *Rapid Commun Mass Spectrom.* 2018 Dec 15;32(23):2031–9.

178. Cole LA. Hyperglycosylated hCG and pregnancy failures. *J Reprod Immunol.* 2012 Mar;93(2):119–22.

179. Bernardini L, Balmaceda J, Gianaroli L, Ferraretti A, Rojas F, Moretti-Rojas I, *et al.* Signal Transduction of hCG Induces Decidualization and Uterine Receptivity. *J Fertil Biomarkers.* 2017 Aug 19;1(1):1–19.

180. Combarous Y. Structure et relations structure-activite des hormones folliculo-stimulantes recombinantes humaines. *Medecine/Sciences.* 1999;15(2):167–74.

181. Kane N, Kelly R, Saunders PTK, Critchley HOD. Proliferation of uterine natural killer cells is induced by human chorionic gonadotropin and mediated via the mannose receptor. *Endocrinology.* 2009;150(6):2882–8.

182. Kaneshiro K, Fukuyama Y, Iwamoto S, Sekiya S, Tanaka K. Highly sensitive MALDI analyses of glycans by a new aminoquinoline-labeling method using 3-aminoquinoline/ α -cyano-4-hydroxycinnamic acid liquid matrix. *Anal Chem.* 2011 May 15;83(10):3663–7.

183. Zhu H, Qiu C, Ruth AC, Keire DA, Ye H. 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.* 2017;19(3):846–55.

184. Bai X, Li D, Zhu J, Guan Y, Zhang Q, Chi L. From individual proteins to proteomic samples: Characterization of O-glycosylation sites in human chorionic gonadotropin and human-plasma proteins. *Anal Bioanal Chem.* 2015;407(7):1857–69.

185. Omasits U, Ahrens CH, Müller S, Wollscheid B. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics.* 2014 Mar 15;30(6):884–6.

186. Irie A, Koyamat S, Kozutsumi Y, Kawasaki T, Suzuki A. The molecular basis for the absence of N-glycolylneuraminic acid in humans. *J Biol Chem.* 1998 Jun 19;273(25):15866–71.
187. Tangvoranuntakul P, Gagneux P, Diaz S, Bardor M, Varki N, Varki A, *et al.* Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proc Natl Acad Sci U S A.* 2003 Oct 14;100(21):12045–50.
188. Yu J, Schorlemer M, Gomez Toledo A, Pett C, Sihlbom C, Larson G, *et al.* Distinctive MS/MS Fragmentation Pathways of Glycopeptide-Generated Oxonium Ions Provide Evidence of the Glycan Structure. *Chem - A Eur J.* 2016 Jan 18;22(3):1114–24.
189. Toghi Eshghi S, Yang W, Hu Y, Shah P, Sun S, Li X, *et al.* Classification of Tandem Mass Spectra for Identification of N- and O-linked Glycopeptides. *Sci Rep.* 2016;6(May):1–8.
190. Halim A, Westerlind U, Pett C, Schorlemer M, RüU, Brinkmalm G, *et al.* Assignment of Saccharide Identities through Analysis of Oxonium Ion Fragmentation Profiles in LC–MS/MS of Glycopeptides. *J Proteome Res.* 2014 Dec 5;13(12):6024–32.
191. Böhm E, Seyfried BK, Dockal M, Graninger M, Hasslacher M, Neurath M, *et al.* Differences in N-glycosylation of recombinant human coagulation factor VII derived from BHK, CHO, and HEK293 cells. *BMC Biotechnol.* 2015 Sep 18;15:87.
192. Ribela MTCP, Damiani R, Silva FD, Lima ER, Oliveira JE, Peroni CN, *et al.* N-glycoprofiling analysis for carbohydrate composition and site-occupancy determination in a poly-glycosylated protein: Human thyrotropin of different origins. *Int J Mol Sci.* 2017 Feb 3;18(2):131.
193. Yu Q, Wang B, Chen Z, Urabe G, Glover MS, Shi X, *et al.* Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact

Glycopeptide/Glycoproteome Characterization. *J Am Soc Mass Spectrom.* 2017 Sep 1;28(9):1751–64.

194. Yu Q, Canales A, Glover MS, Das R, Shi X, Liu Y, *et al.* Targeted Mass Spectrometry Approach Enabled Discovery of O-Glycosylated Insulin and Related Signaling Peptides in Mouse and Human Pancreatic Islets. *Anal Chem.* 2017 Sep 5;89(17):9184–91.

195. Giangrande C, Auberger N, Rentier C, Papini AM, Mallet JM, Lavielle S, *et al.* Multi-Stage Mass Spectrometry Analysis of Sugar-Conjugated β -Turn Structures to be Used as Probes in Autoimmune Diseases. *J Am Soc Mass Spectrom.* 2016;27(4):735–47.

196. Hu Y, Mechref Y. Comparing MALDI-MS, RP-LC-MALDI-MS and RP-LC-ESI-MS glycomic profiles of permethylated N-glycans derived from model glycoproteins and human blood serum. *Electrophoresis.* 2012 Jul;33(12):1768–77.

197. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics.* 2002 Nov;1(11):845–67.

Caractérisation structurale de la glycosylation des protéines : de la glycomique à la glycoprotéomique, une stratégie utile pour l'étude des gonadotrophines

RÉSUMÉ

La composante sucrée des glycoprotéines joue un rôle majeur dans de nombreux processus physiologiques et pathologiques. Leur analyse structurale nécessite des outils performants capables d'évaluer la forte hétérogénéité de ces glycanes. Des approches complémentaires ont été développées : l'analyse des N-glycanes par MALDI-TOF et l'approche protéomique par LC-ESI-FTMS.

La glycomique consiste en l'analyse des sucres indépendamment de la protéine, pouvant livrer une empreinte globale de la glycosylation au niveau d'une protéine, d'un fluide *etc.*... Le développement de nouvelles conditions de préparation d'échantillon et de matrice incluant la modification chimique des sucres nous a permis d'améliorer leur détection et leur fragmentation par spectrométrie de masse.

La considération des glycanes comme partie des glycoprotéines et glycopeptides permet une sélectivité accrue et une caractérisation de l'hétérogénéité intramoléculaire de ces glycosylations. L'intégration de protéases aspécifiques et de fragmentation à énergie variable (SCE) nous a conduit à plus de diversité et de fiabilité dans les glycoformes identifiées.

De par leur implication dans la grossesse et le développement de l'embryon, les glycohormones gonadotropes ont un fort intérêt médical et vétérinaire. La combinaison des méthodes analytiques a été entreprise dans le but de caractériser les motifs de glycosylation spécifiques à différentes populations.

MOTS CLÉS

Glycosylation ; Spectrométrie de masse ; Protéomique

ABSTRACT

Polysaccharides from glycoproteins play an important part in numerous biological processes and pathologies. Their structural elucidation requires performant and sensitive tools capable of assessing the heterogeneity of such glycans. Different approaches were developed: N-glycan analysis by MALDI-TOF and LC-ESI-FTMS proteomics.

Glycomics is the analysis of glycans independently of their linked-protein, and can be using to obtain a glycosylation fingerprint of a protein, fluid *etc.*... Developments of innovative sample and matrix preparation conditions, including glycan chemical modification, improved their mass spectrometry detection and fragmentation.

Glycopeptide and glycoprotein analysis permits selective assessment of the glycan population across the protein hence allowing the characterization of molecular microheterogeneity of the different glycoforms. The development and integration of non-specific protease digestion and stepped-energy fragmentation produced more diverse and reliable identifications.

Because of their critical role in gestation and embryo development, glycosylated gonadotropins are of major medical and veterinary interest. Combining analytical methodologies granted access to the characterization of population-specific glycosylation motifs.

KEYWORDS

Glycosylation ; Mass spectrometry ; Proteomics