

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.

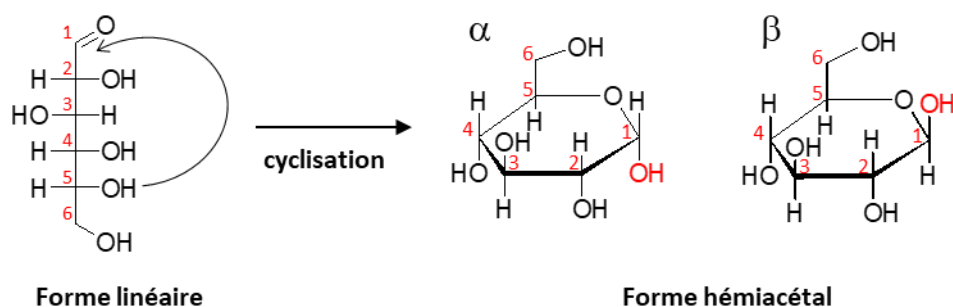


Figure 2 : Formes linéaire et cyclique (hémiacétal) du glucose.

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.

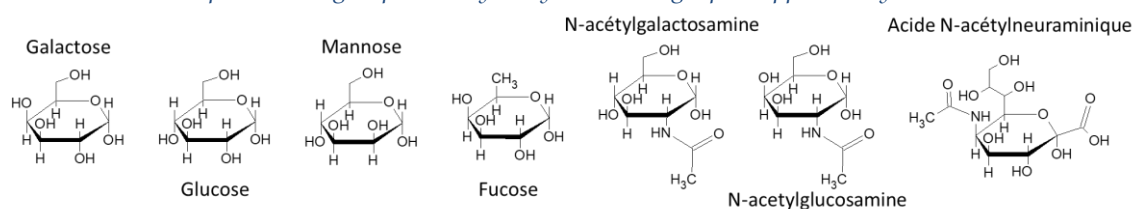


Figure 3 : Formes α -pyranosiques (hétérocycle hémiacétal à 6 atomes) des monosaccharides usuellement rencontrés dans les glycanes des mammifères.

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 glycane, 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.

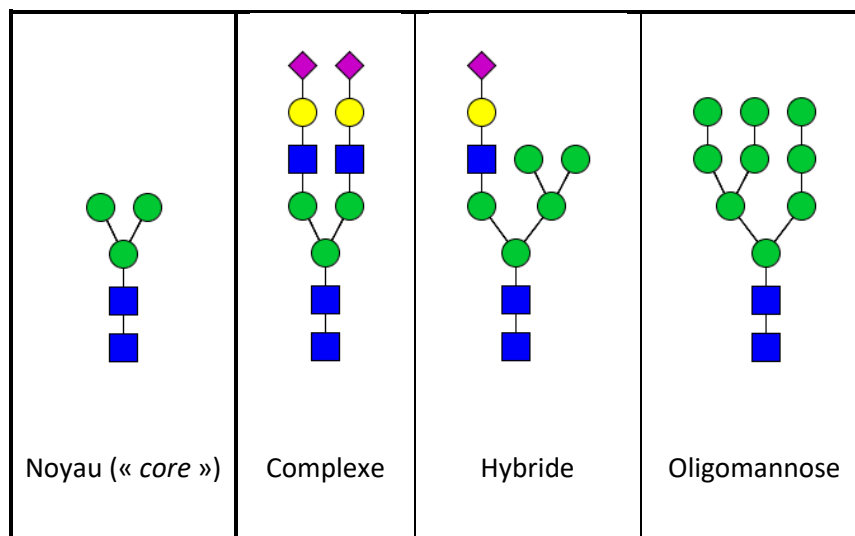


Figure 4 : N-glycosylation, représentations SNFG du noyau (ou « core ») commun aux N-glycane et des trois familles de N-glycane.

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-glycane ont en commun un noyau pentasaccharidique composé de deux N-acétylglucosamines et trois mannoses (appelé « core »), lesquels peuvent être différemment modifiés et étendus. La famille des oligomannoses (ou « high-mannose ») regroupe les N-glycane constitués du core et d'un nombre variable de mannoses 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).

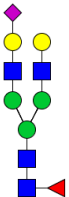
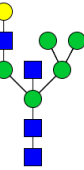
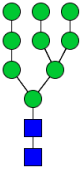
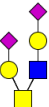
Nomenclature IUPAC	NeuAc α 2-6 Gal β 1-4 GlcNAc β 1-2 Man α 1-3 (Gal β 1-4 GlcNAc β 1-2 Man α 1-6) Man β 1-4 GlcNAc β 1-4 (Fuc α 1-6) GlcNAc	GlcNAc β 1-4 (Gal β 1-4 GlcNAc β 1-2 Man α 1-3) (Man α 1-3 (Man α 1-6) Man α 1-6) Man β 1-4 GlcNAc β 1-4 GlcNAc	Man α 1-2 Man α 1-2 Man α 1-3 (Man α 1-2 Man α 1-3 (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) GlcNAc
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.

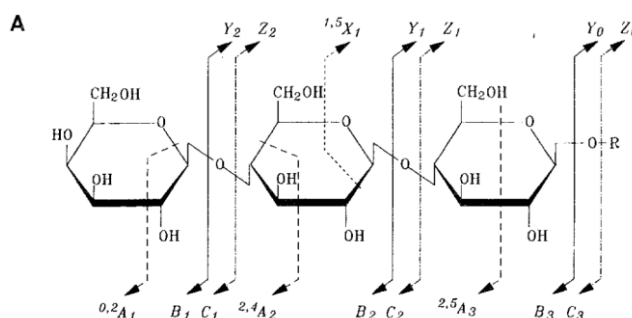


Figure 7 : Nomenclature des fragments des polysaccharides.

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 thyroéostimuline (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.

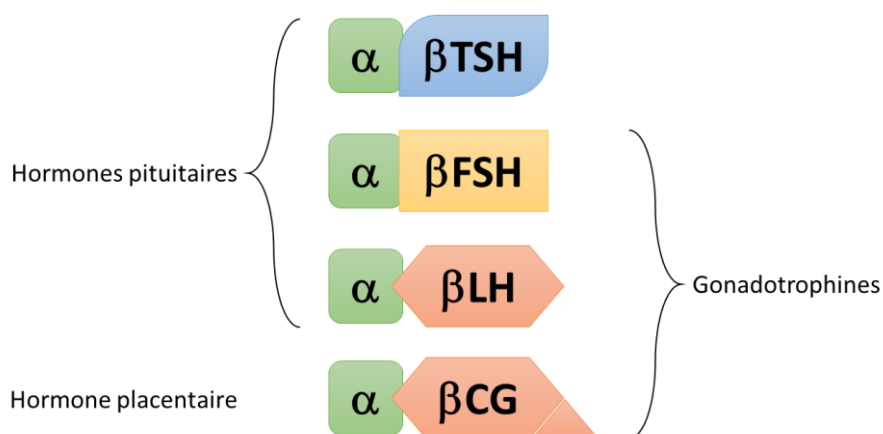


Figure 8 : Les glycoprotéines TSH, FSH, LH et CG.

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).

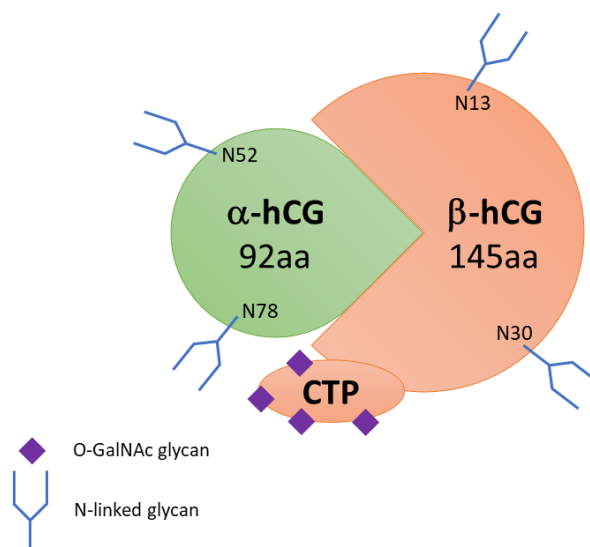


Figure 9 : Gonadotrophine chorionique humaine.

Schéma montrant les N-glycane et O-glycane. 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 glycoprotéoformes 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).

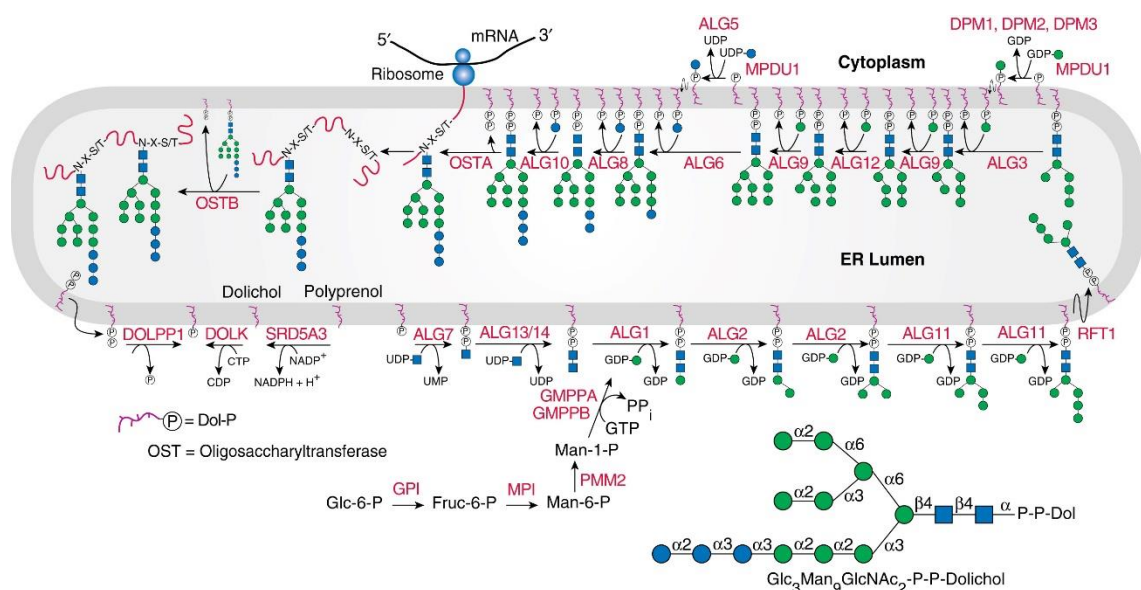


Figure 10 : Synthèse du précurseur Dol-P-P-GlcNAc₂Man₉Glc₃ à la surface externe puis dans le réticulum endoplasmique.

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.

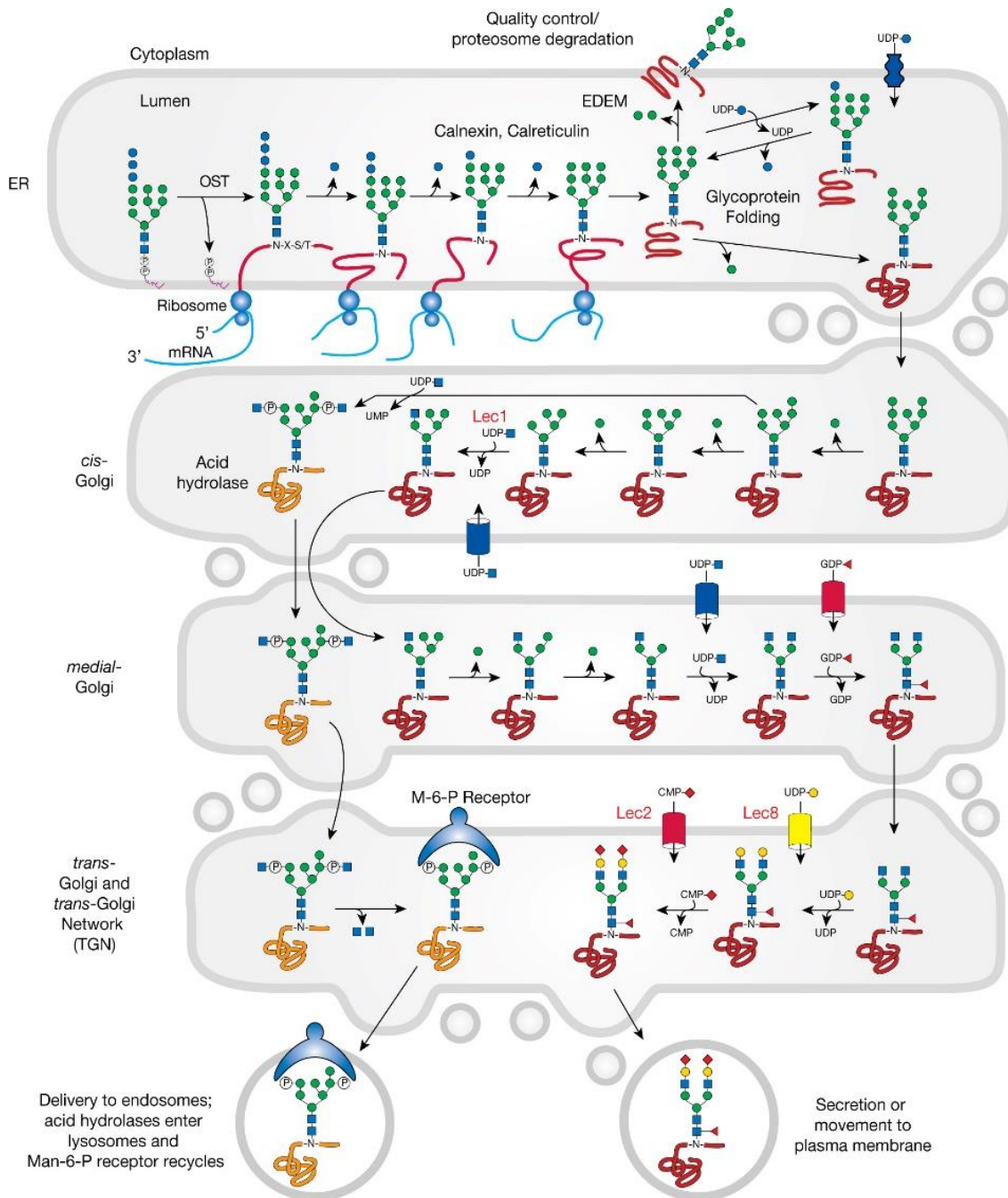


Figure 11 : Maturation d'un N-glycane dans le RE et l'appareil de Golgi.
Figure issue du livre « Essentials of Glycobiology, Third Edition, Chapter 9, Figure 4 ».

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).

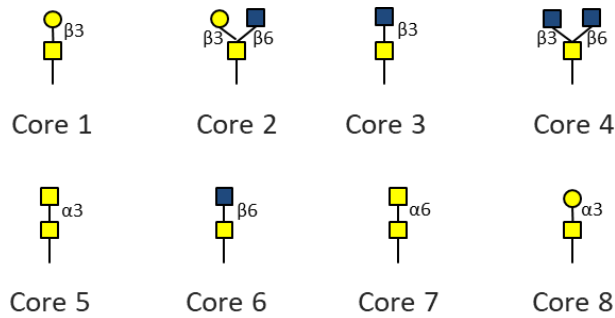


Figure 12 : Noyaux des O-glycanes de type mucine.

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-acetylglucosaminyl-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-acetylglucosaminyltransférases) 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-galactosyltransferase 1), la B3GALT5, ayant une affinité pour le *core* 3 et certains glycolipides (β 1,3-galactosyltransferase 5), la B3GNT3 (N-acétyllactosaminide β 1,3-N-acétylglucosaminyltransferase 3), la ST3GAL1 qui catalyse la synthèse de l'antigène sialyl-T à partir du *core* 1 (CMP-N-acétylneuraminate- β -galactosamide- α 2,3-sialyltransferase 1) et la ST6GALNAC1 (α N-acétylgalactosaminide α 2,6-sialyltransferase 1) qui catalyse la sialylation de l'antigène Tn.

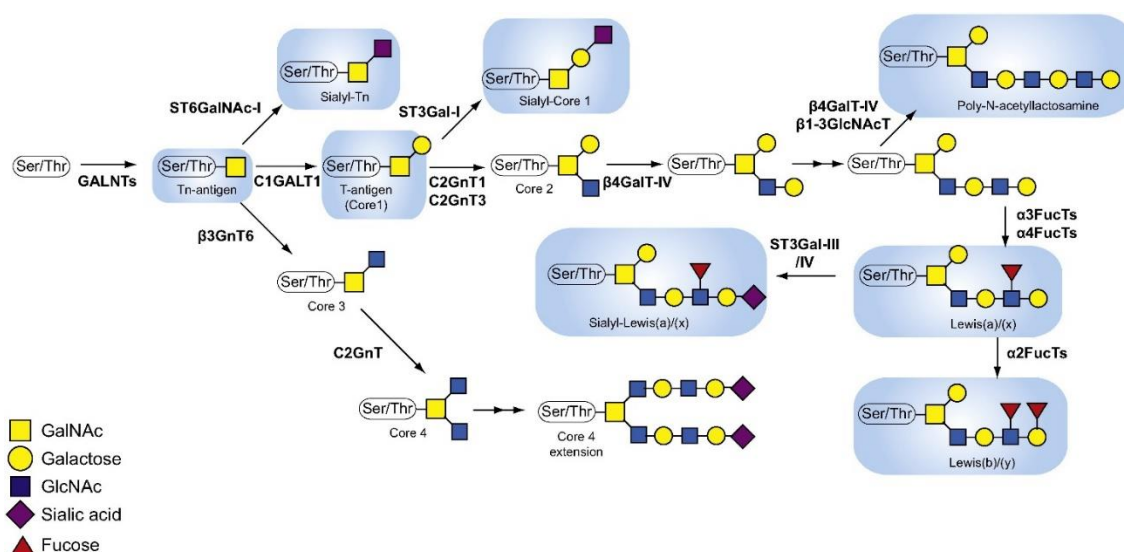


Figure 13 : Voies de synthèse des cores O-GalNAc 1 à 4 et formes de Lewis.

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 glycosyltransférase 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.