



HAL
open science

Ingénierie de glycoside hydrolases pour la glycosylation des protéines recombinantes

Sophie Blanchard

► **To cite this version:**

Sophie Blanchard. Ingénierie de glycoside hydrolases pour la glycosylation des protéines recombinantes. Autre. Université Joseph-Fourier - Grenoble I, 2004. Français. NNT : . tel-00008252

HAL Id: tel-00008252

<https://theses.hal.science/tel-00008252>

Submitted on 25 Jan 2005

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.

UNIVERSITÉ JOSEPH FOURIER - GRENOBLE 1

THÈSE

Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ JOSEPH FOURIER
(Arrêtés ministériels du 5 Juillet 1984 et du 30 Mars 1992)

Spécialité : **CHIMIE BIOLOGIE**

Présentée par

Sophie BLANCHARD

**INGENIERIE DE GLYCOSIDE HYDROLASES
POUR LA GLYCOSYLATION DES PROTEINES RECOMBINANTES**

Soutenue publiquement le 7 Décembre 2004

COMPOSITION DU JURY :

Dr. Philippe DEBEIRE, Rapporteur

Pr. Charles TELLIER, Rapporteur

Dr. Bernard HENRISSAT

Dr. Sylvie ARMAND

Pr. Sylvain COTTAZ

Dr. Hugues DRIGUEZ

CEntre de Recherches sur les MAcromolécules Végétales (CERMAV) - CNRS

UNIVERSITÉ JOSEPH FOURIER - GRENOBLE 1

THÈSE

Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ JOSEPH FOURIER
(Arrêtés ministériels du 5 Juillet 1984 et du 30 Mars 1992)

Spécialité : **CHIMIE BIOLOGIE**

Présentée par

Sophie BLANCHARD

**INGENIERIE DE GLYCOSIDE HYDROLASES
POUR LA GLYCOSYLATION DES PROTEINES RECOMBINANTES**

Soutenue publiquement le 7 Décembre 2004

COMPOSITION DU JURY :

Dr. Philippe DEBEIRE, Rapporteur

Pr. Charles TELLIER, Rapporteur

Dr. Bernard HENRISSAT

Dr. Sylvie ARMAND

Pr. Sylvain COTTAZ

Dr. Hugues DRIGUEZ

CEntre de Recherches sur les MAcromolécules Végétales (CERMAV) - CNRS

A mes parents

Une période de ma vie s'achève, une autre page se tourne, et celles qu'il me reste ne sont pas les plus faciles à écrire. Pendant ces trois années (et quelques mois), on pense à ce que l'on va écrire pour dire MERCI à toutes les personnes qui ont participé de près ou de loin à l'achèvement de cette étape essentielle qu'est le doctorat. Je m'excuse d'avance pour tout oubli, mais l'âge ainsi que les premières rides sont des ennemis difficiles à combattre.

Ces trois années de thèse se sont déroulées au CERMAV. Je tiens à remercier son directeur, Serge PEREZ pour m'avoir accueillie au sein de ce laboratoire.

J'adresse un grand merci et mon plus profond respect à Hugues DRIGUEZ pour m'avoir accueillie au sein de son équipe, pour le pari que vous avez fait sur moi quand, au mois de Juin 2001, vous avez proposé ce sujet, avec Sylvain et Sylvie. Vous serez à jamais gravé dans ma mémoire (olfactive) avec votre petit rituel de chaque début d'après-midi qui me permettait de profiter (à l'insu de mon plein gré) de l'atmosphère (nauséabonde) créée par vos cigares cubains.

Merci à Sylvain COTTAZ qui a co-dirigé ces travaux avec Sylvie, pour avoir su s'adapter à mon fort caractère. Je me souviendrais de la musique que je trouvais chaque matin en arrivant dans le labo, qui me donnait plutôt envie de retourner me coucher, ainsi que tes blagues (à mourir de rire...). Cela a énormément participé à la bonne humeur ambiante. Tu peux maintenant de reposer en paix, plus de menace annuelle au moment de Noël, quoique...

J'adresse toute ma reconnaissance à Sylvie ARMAND, ma formatrice officielle en biologie et biochimie, chimiste de formation que je suis! Avec ta gentillesse et ta disponibilité, j'ai pu étendre pendant cette thèse mes compétences et me former à des domaines qui m'étaient jusque-là obscurs. Et t'as quand même réussi à me faire monter les larmes aux yeux pendant ma soutenance! J'adresse également un grand merci à François, mon chef en second, pour m'avoir également transmis une petite partie de son savoir en biologie moléculaire.

Je tiens à remercier Charles TELLIER, Philippe DEBEIRE et Bernard HENRISSAT pour avoir accepté de juger ce travail.

Je remercie également Pedro COUTINHO pour m'avoir initié aux joies de la modélisation moléculaire. Merci à Jesper VIND et Anu KOIVULA pour avoir collaboré à ce projet.

Merci à Claude BOSSO et à Stéphanie BEFFY pour les analyses en spectrométrie de masse et pour la patience dont ils ont fait preuve. Et il en a fallu!

Mes sincères remerciements à Jean-Pierre UTILLE pour toutes nos petites conservations sur le tout et le rien.

Une mention (très très très)¹⁰⁰⁰ spéciale à Lapin, Coquille et Poussin qui ont été les tout premiers à m'accueillir, fraîchement débarquée que j'étais de mon grand Nord (revoyez votre géographie, les gars) et sans qui je n'aurais pas pu réussir aussi bien mon intégration dans le clan des thésards du CERMAV! Merci Lionel, pour ton inégalable sens aiguisé de la répartie que j'adore; tu as trop vite quitté Grenoble. Merci au saint Régis (de droite) et au saint David (de gauche), que j'ai dû supporter ET dans le bureau ET dans le labo (nan, j'rigole!). Merci à celui qui se surnomme le désagréable mais qui ne l'est nullement, toujours à l'écoute de mes éternels doutes, tout comme toi, le grand man in black, le seul supporter du PSG que je connaisse (une espèce en voie de disparition d'ailleurs). Je vous porterai à jamais dans mon cœur. Rendez-vous de l'autre côté de l'Atlantique!

Merci Silvia, la grande sœur que je n'ai jamais eue et que j'ai rencontrée ici. Tout au long de ces trois années, tu as su m'écouter, me reconforter, me rassurer, me faire rire à en pleurer avec tes jeux de mots (et en Français s'il vous plaît!), bref être là. Merci pour toutes nos petites discussions, de tout et de rien, et de ces fabuleuses vacances sous le soleil de Séville.

Merci à Marie-Pierre, originaire d'un coin encore plus paumé que moi (et oui ça existe!). Arrivée il n'y a pas si longtemps, mais qui tient déjà une très grande place. Je me souviendrais de ces semaines de colocation, riches de bons souvenirs et de nos longues discussions jusqu'à tard dans la nuit.

Côté gente masculine, je voudrais également remercier: Olivier dit "Dieu" (ton espièglerie ne s'arrêtera donc jamais! Ah, et un dernier mot, enfin deux : BONNE ANNEE!), le petit Grégory (je suis sûre que tu me détestes pour ce que je viens d'écrire), Fred (le seul Superman que je connaisse vouant une admiration sans faille à Britney Spears), Miguel (le Maître Jedi de l'informatique; muy buen, muy buen!), Istvan (qui a fait de mon appart' l'ambassade officielle de Hongrie à Grenoble l'espace d'une soirée), Martin (et tous nos petits cafés du matin, et toutes nos taquineries), Youssef (et son double Youyou le suédois), Fabrizio (toujours à la recherche de la touffe à la cafèt'), Stéphane (et ses deux mains gauches).

Côté gente féminine, je ne peux pas oublier mes deux compères des soirées Alias, Aurélia (mon corps se souviendra très longtemps de tes multiples impacts de balles... de squash qu'il a pu recevoir) et Hélène (la douceur et la gentillesse incarnées, merci pour les cartons du 22 Décembre!). Je tiens également à remercier Marga (la doudou de Régis, ton éternelle joie de vivre est un exemple pour moi), Lemonia (je suis encore étonnée que la Grèce ait réussi à gagner le championnat d'Europe), Carole, Claire (qui suscite mon admiration quand je vois son petit Louis grandir), Charlotte (mes premiers

cours de ski... et derniers malheureusement), Valérie, Stéphanie (pour toutes les petites discussions que nous avons partagé; vive Shrek 1&2!), Isabelle (tu me diras comment tu passes rapidement de tes degrés Calvin Klein en °C ou F), Christelle (restée si peu de temps mais nos délires resteront dans ma mémoire), Emilie (LA touche féminine du labo 319 pendant trois ans), Marie (nouvelle recrue), Helena.

Merci à Patrick (le bricoleur du CERMAV, qui m'a permis de comprendre pourquoi ma titine prenait l'eau), Robert (merci de m'adresser encore la parole; et pis non, t'es pas tout fripé!), Pierre (et son entrain communicatif pour l'éveil aux sciences), Sandra (partie trop vite de l'équipe).

Une mention particulière à Martine M. (pour avoir si bien pris soin de moi, tu resteras ma maman-CERMAV) et Magali (Madame Alarme, merci de ne pas m'avoir trop disputé). La table 6 était la plus belle ce fameux samedi soir!

Merci à Claudius (sans qui j'aurais oublié les premiers gestes de secourisme), Annie, Martine B. (et ses petits mails qui mettent de la bonne humeur partout où ils vont), Josiane (désolée de t'avoir fait si peur, ce matin-là, à 6h du mat')

Un grand merci à Perrine, Sandrine (les deux), Stéphane, Caroline, Benoît et Nathalie, pour ces petits week-ends aux quatre coins de la France qui m'ont permis de m'évader au cours de ces trois années.

J'aimerais finir sur ces quelques phrases qu'un talent grenoblois a écrit un jour : «On ne retient pas l'écume dans le creux de sa main. On sait [que] la vie se consume et [qu'] il n'en reste rien, [comme] d'une bougie qui s'allume. Tu peux encore décider du chemin, de ton chemin. Crois-tu que tout se résume au sel d'entre nos doigts? Quand plus léger qu'une plume, tu peux guider tes pas, sans tristesse ni amertume, et avancer. Avance, puisque tout s'en va!»

A tous, merci!

Fopie.

SOMMAIRE

SOMMAIRE

INTRODUCTION	- 1 -
I CONTEXTE DE L'ETUDE	- 3 -
I.1 Les protéines	- 3 -
I.1.1 Les modifications post-traductionnelles des protéines	- 5 -
I.1.2 La glycosylation des protéines	- 7 -
I.1.2.1 La N-glycosylation	- 10 -
I.1.2.2 La O-glycosylation	- 15 -
I.1.2.3 Rôle des groupements glycaniques	- 17 -
I.1.2.4 Participation de la glycosylation dans des infections et autres maladies	- 19 -
I.1.3 Utilisation des protéines : les protéines recombinantes.....	- 20 -
I.1.4 La production des protéines recombinantes	- 23 -
I.1.4.1 Les différents systèmes de production	- 24 -
I.1.4.2 Comparaison des différents systèmes	- 31 -
I.1.5 Le contrôle de la glycosylation des protéines	- 33 -
I.1.5.1 Remodelage de la glycosylation : voie chimique et in vitro	- 34 -
I.1.5.2 Ingénierie de la machinerie de glycosylation	- 36 -
I.1.5.3 Conclusion.....	- 40 -
I.2 Les glycosynthèses : un formidable outil de synthèse	- 40 -
I.2.1 Les glycoside hydrolases	- 41 -
I.2.1.1 Les classifications des glycoside hydrolases	- 41 -
I.2.1.2 Modes d'action	- 43 -
I.2.1.3 Topologie des sites actifs et distorsion de substrat.....	- 46 -
I.2.2 Le concept de glycosynthase.....	- 48 -
I.2.3 Les différents exemples de glycosynthèses.....	- 49 -
I.2.4 La glycosynthase Cel7B E197A d' <i>Humicola insolens</i>	- 54 -
I.3 L'ingénierie des protéines	- 58 -
I.3.1 L'évolution des protéines : pourquoi ?	- 59 -
I.3.2 L'évolution des protéines : comment ?	- 60 -
I.3.2.1 Les différentes modifications	- 60 -
I.3.2.2 La conception rationnelle (" <i>rational design</i> ") de mutants de protéines	- 62 -
I.3.2.3 L'évolution dirigée des protéines.....	- 64 -
I.3.3 La modification de la spécificité de substrat des enzymes.....	- 66 -
OBJECTIFS	- 69 -
II ETUDES PRELIMINAIRES SUR LA GLYCOSYNTHESE CEL7B E197A... - 73 -	
II.1 Activité de la glycosynthase vis-à-vis d'un glucopeptide accepteur	- 73 -
II.1.1 Préparation des synthons.....	- 74 -
II.1.1.1 Synthèse du glucopeptide 14	- 74 -
II.1.1.2 Préparation du fluorure d' α -lactosyle.....	- 79 -
II.1.2 Test de couplage enzymatique	- 79 -
II.2 Activité de la glycosynthase vis-à-vis d'un fluorure de 6^{II}-S-sialyl-6^{II}-thiolactosyle donneur	- 80 -
II.2.1 Synthèse du fluorure de 6 ^{II} -S-sialyl-6 ^{II} -thiolactosyle 31	- 81 -
II.2.1.1 Préparation du précurseur sialyle 20	- 82 -
II.2.1.2 Préparation du précurseur lactoside 25	- 83 -
II.2.1.3 Thioglycosylation	- 84 -

II.2.2	Test de couplage enzymatique	- 86 -
II.3	Conclusion.....	- 87 -
III	INGENIERIE DE LA GLYCOSYNTHE CEL7B E197A D'H. INSOLENS. - 89 -	
III.1	Etude de la spécificité de l'enzyme.....	- 89 -
III.1.1	La famille GH7	- 90 -
III.1.2	Etude préliminaire qualitative	- 92 -
III.1.3	Alignements de séquences	- 93 -
III.1.3.1	Principe	- 94 -
III.1.3.2	Résultats	- 94 -
III.1.4	Etudes d'arrimage moléculaire	- 96 -
III.1.4.1	Principe d'utilisation d'AutoDock.....	- 97 -
III.1.4.2	Les expériences de docking réalisées	- 101 -
III.1.4.3	Analyses des solutions de docking.....	- 102 -
III.1.5	Résumé.....	- 107 -
III.2	Etude de l'hydrolyse du tétrasaccharide Gal(β1,4)Glc(β1,4)GlcNAc(β1,4)GlcNAc 34.....	- 109 -
III.2.1	Synthèse du substrat 34.....	- 109 -
III.2.2	Hydrolyse du substrat.....	- 110 -
III.3	Etude de l'activité de transglycosylation de Cel7B de Trichoderma reesei.....	- 114 -
III.4	Mutagenèse dirigée.....	- 115 -
III.4.1	Préparation des mutants de Cel7B E197A d'H. insolens.....	- 115 -
III.4.2	Préparation de la glycosynthase de Cel7B de T. reesei.....	- 116 -
III.5	Conclusion.....	- 117 -
IV	PRODUCTION ET CARACTERISATION DES ENZYMES MUTANTES.... - 119 -	
IV.1	Production des enzymes.....	- 119 -
IV.1.1	Tentatives de surexpression de Cel7B d'H. insolens chez E. coli et B. subtilis	- 119 -
IV.1.1.1	Expression de Cel7B chez Escherichia coli	- 120 -
IV.1.1.2	Expression de Cel7B chez Bacillus subtilis	- 123 -
IV.1.1.3	Production des mutants chez le champignon Aspergillus oryzae.....	- 124 -
IV.1.2	Production de la glycosynthase Cel7B E196A de T. reesei chez Saccharomyces cerevisiae	- 124 -
IV.2	Caractérisation des enzymes.....	- 124 -
IV.2.1	Caractérisation des nouvelles glycosynthases avec un cellobiose comme accepteur	- 125 -
IV.2.2	Caractérisation de la spécificité de substrat	- 129 -
IV.2.2.1	Activité vis-à-vis de glycosides de p-nitrophényle et de disaccharides substitués.....	- 133 -
IV.2.2.2	Activité vis-à-vis du N ¹ -acétyl-2 ^{II} -azidochitobioside d'allyle 39.....	- 136 -
IV.3	Conclusion.....	- 145 -
V	PARTIE EXPERIMENTALE..... - 147 -	
V.1	Modélisation moléculaire.....	- 147 -
V.1.1	Alignements de séquence.....	- 147 -
V.1.2	Docking.....	- 147 -
V.1.3	Visualisation des structures.....	- 150 -
V.2	Synthèses chimiques et enzymatiques.....	- 150 -
V.2.1	Matériels et méthodes générales	- 150 -
V.2.1.1	Produits de départ	- 150 -
V.2.1.2	Solvants et tampons	- 151 -

V.2.1.3	<i>Enzymes utilisées</i>	- 151 -
V.2.1.4	<i>Chromatographies</i>	- 151 -
V.2.1.5	<i>Analyses physico-chimiques</i>	- 152 -
V.2.1.6	<i>Caractérisations par spectroscopie par Résonance Magnétique Nucléaire (RMN)</i>	- 152 -
V.2.2	<i>Synthèses</i>	- 153 -
V.2.2.1	<i>Synthèse peptidique</i>	- 153 -
V.2.2.2	<i>Dérivés du D-glucose</i>	- 160 -
V.2.2.3	<i>Synthèse du dérivé glucopeptidique</i>	- 162 -
V.2.2.4	<i>Dérivés du lactose</i>	- 163 -
V.2.2.5	<i>Dérivés de l'acide neuraminique</i>	- 171 -
V.2.2.6	<i>Synthèse et fonctionnalisation du trisaccharide sialylé</i>	- 173 -
V.2.2.7	<i>Couplages enzymatiques à l'échelle analytique</i>	- 177 -
V.2.2.8	<i>Préparation des substrats utilisés pour la caractérisation des enzymes</i>	- 180 -
V.2.2.9	<i>Couplage enzymatique entre le fluorure d'α-lactosyle 16 et le disaccharide 39</i>	- 183 -
V.3	Biologie moléculaire et biochimie	- 189 -
V.3.1	<i>Matériels</i>	- 189 -
V.3.1.1	<i>Produits chimiques et réactifs</i>	- 189 -
V.3.1.2	<i>Endoglucanases natives</i>	- 189 -
V.3.1.3	<i>Souches bactériennes, plasmides et milieux de culture</i>	- 190 -
V.3.1.4	<i>Enzymes</i>	- 191 -
V.3.2	<i>Travaux de biologie moléculaire</i>	- 191 -
V.3.2.1	<i>Méthodes générales</i>	- 191 -
V.3.2.2	<i>Constructions des plasmides d'expression</i>	- 192 -
V.3.2.3	<i>Mutagenèse dirigée</i>	- 195 -
V.3.3	<i>Expression de Cel7B d'<i>H. insolens</i> chez la bactérie</i>	- 197 -
V.3.3.1	<i>Méthodes générales</i>	- 197 -
V.3.3.2	<i>Etude de l'expression chez E. coli</i>	- 199 -
V.3.3.3	<i>Etude de l'expression chez B. subtilis</i>	- 199 -
V.3.4	<i>Production des enzymes dans des organismes hétérologues</i>	- 200 -
V.3.4.1	<i>Enzymes d'<i>H. insolens</i> chez Aspergillus oryzae</i>	- 200 -
V.3.4.2	<i>Enzymes de T. reesei chez Saccharomyces cerevisiae</i>	- 200 -
V.3.5	<i>Détermination des activités enzymatiques</i>	- 201 -
V.3.5.1	<i>Détection des activités</i>	- 201 -
V.3.5.2	<i>Quantification par HPLC</i>	- 201 -
	CONCLUSION ET PERSPECTIVES	- 205 -
	REFERENCES BIBLIOGRAPHIQUES	- 209 -
	ANNEXES	- 223 -
	Liste des abréviations	- 223 -
	Code à trois lettres et à une lettre des acides aminés	- 226 -
	Liste des Figures	- 227 -
	Liste des Tableaux	- 229 -
	Liste des Schémas	- 230 -

INTRODUCTION

INTRODUCTION

Grâce aux avancées scientifiques spectaculaires des vingt dernières années en matière de génie génétique et de biologie moléculaire, il est aujourd'hui envisageable de modifier le patrimoine génétique d'un organisme vivant en lui incorporant un fragment d'ADN provenant d'une espèce différente, et de lui faire exprimer cette information génétique étrangère. Ces savoir-faire nouveaux de la génétique ont permis d'ouvrir des voies de recherche et de développement nouvelles aux perspectives considérables. Il devient alors possible aujourd'hui, en s'appuyant sur ces méthodologies, de reprogrammer le code génétique d'un organisme en lui insérant un gène précis, qui lui permettra de produire une protéine recombinante précieuse à usage médical ou industriel.

Ces procédés biotechnologiques permettent ainsi de produire des molécules trop complexes pour être synthétisées par voie chimique, et constituent une alternative plus intéressante en regard d'autres procédés d'obtention de molécules actives, comme l'extraction depuis le milieu naturel. Les biotechnologies ouvrent ainsi aujourd'hui des perspectives considérables pour le secteur de la santé. Les protéines recombinantes avec un intérêt thérapeutique occupent d'ailleurs une place primordiale dans le marché mondial des biotechnologies.

Les protéines recombinantes d'intérêt thérapeutique sont le plus souvent des glycoprotéines, où la glycosylation joue un rôle essentiel dans leurs propriétés, notamment pharmacocinétiques. Toutefois, les différents systèmes d'expression de protéines utilisés ne permettent pas de reproduire le profil de glycosylation observé à la surface des protéines humaines, ce qui peut constituer un inconvénient à l'utilisation de la protéine dans une application thérapeutique chez l'Homme. La maîtrise de la glycosylation des protéines recombinantes reste un enjeu considérable pour le développement de tels systèmes de production.

Les travaux présentés dans ce manuscrit ont été réalisés au sein de l'équipe "Glycochimie et Biotechnologie des Oligosaccharides" du CERMAV. Ils s'inscrivent dans

l'une des thématiques développée par le groupe visant à développer des méthodes de synthèse par voie enzymatique d'oligosaccharides naturels ou non naturels et de glycoconjugués.

Notre but est de développer une nouvelle approche pour contrôler la glycosylation des protéines recombinantes, basée sur l'utilisation de biocatalyseurs tels que les glycosynthases, et plus particulièrement Cel7B E197A d'*Humicola insolens*. Il s'agit ici d'accroître le potentiel de cette glycosynthase vers la préparation de glycocongués cibles.

Ce manuscrit s'articule autour de cinq chapitres.

Le premier chapitre, à caractère bibliographique, s'organise en trois volets. Tout d'abord, nous dresserons un état des connaissances acquises sur la glycosylation des protéines, en s'attachant plus particulièrement à la *N*-glycosylation qui nous intéresse dans notre étude. Nous développerons notamment les différents moyens de production des protéines recombinantes, ainsi que les méthodes envisagées pour le remodelage du profil de glycosylation de celles-ci. Ensuite, une deuxième partie présentera le concept de glycosynthase et fera un bilan des applications des différentes glycosynthases existantes. Une attention plus particulière sera portée à la glycosynthase Cel7B E197A d'*H. insolens*, utilisée au cours de ces travaux. Enfin, le dernier volet sera consacré à l'ingénierie des protéines, présentant les différentes techniques utilisées et en s'attachant plus particulièrement à la modification de spécificité de substrat des enzymes.

Dans un deuxième chapitre, nous présenterons les études préliminaires réalisées sur la glycosynthase Cel7B E197A pour sa capacité à utiliser des substrats non-naturels.

Le troisième chapitre sera consacré à l'ingénierie de la glycosynthase Cel7B E197A *via* une approche rationnelle. Des études de modélisation moléculaire permettront de déterminer des mutations ponctuelles à réaliser afin d'augmenter le potentiel synthétique de la glycosynthase. Les modifications de la glycosynthase seront alors effectuées par mutagenèse dirigée.

Le dernier chapitre montrera les résultats obtenus avec les différentes glycosynthases. La spécificité de substrat des mutants sera notamment caractérisée.

Chapitre I

CONTEXTE DE L'ETUDE

I CONTEXTE DE L'ETUDE

I.1 Les protéines

Le terme "protéine" vient d'un mot grec signifiant "de première importance" ou "qui tient la première place". En effet, les protéines sont des constituants extrêmement importants des cellules vivantes (**Figure 1**), tant d'un point de vue quantitatif (puisqu'elles représentent en général plus de la moitié du poids sec des cellules), que du point de vue qualitatif, car à côté des protéines dites structurales on trouve des protéines ayant un rôle biologique capital, en particulier les enzymes, catalyseurs biologiques indispensables au déroulement des réactions dans les cellules des organismes vivants.

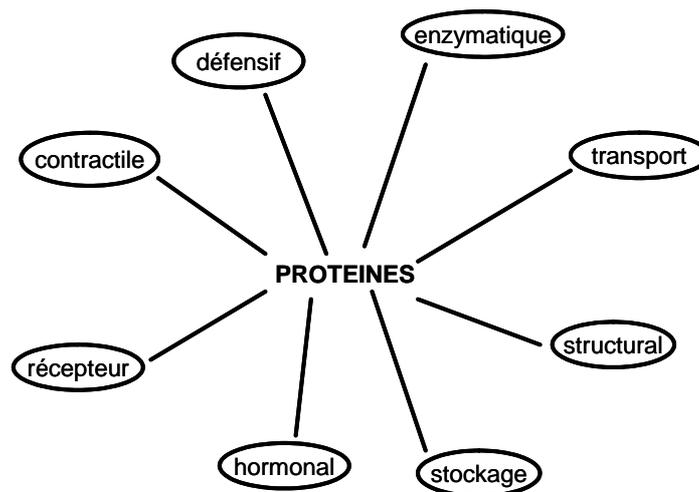


Figure 1 : Rôles biologiques des protéines des organismes vivants.

Certaines protéines stockent et transportent une grande variété de particules et de molécules comme des électrons, des cations métalliques, de l'oxygène... En tant qu'hormones, les protéines transmettent chez les organismes complexes l'information entre des cellules spécifiques et les organes. Notre système immunitaire, qui nous protège contre une variété d'agents pathogènes potentiels tels que les bactéries ou les virus, repose sur les propriétés de différentes protéines ancrées à la surface des cellules ou circulantes (les plus connues sont les anticorps), qui participent à la reconnaissance et à l'élimination de ces corps étrangers.

D'autres protéines jouent un rôle dans le contrôle de l'expression des gènes en se liant à des séquences spécifiques d'acides nucléiques, régulant ainsi la transcription/traduction du gène impliqué. Elles sont les composantes cruciales des muscles et d'autres systèmes qui assurent la conversion d'énergie chimique en énergie mécanique. Finalement, de nombreuses protéines sont "simplement" structurales, impliquées dans l'architecture des cellules ou d'autres matériels présents dans les cheveux, les ongles, les tendons ou les os.

Malgré leurs fonctions biologiques très différentes, les protéines constituent une classe relativement homogène de molécules, à savoir un polymère linéaire construit à partir de combinaisons variées des mêmes 20 acides aminés. Ces macromolécules diffèrent seulement dans l'enchaînement de ces unités.

Les protéines naturelles possèdent une taille et une séquence d'acides aminés qui leur sont propres, conformément à l'information stockée dans le génome. Leur diversité fonctionnelle repose en partie sur la diversité des caractéristiques chimiques de leurs acides aminés constitutifs, mais principalement sur la diversité de structures tridimensionnelles que l'enchaînement de ces derniers peut former.

Les protéines possèdent typiquement plusieurs niveaux d'organisation. La séquence d'acides aminés définit la structure primaire (structure covalente). Mais chaque protéine possède en réalité une structure tridimensionnelle qui lui est propre, établie et maintenue par des liaisons autres que la liaison peptidique liant les acides aminés entre eux : ce sont les liaisons hydrogènes, les liaisons de type ionique, les interactions hydrophobes ou les forces de van der Waals. Outre ces liaisons de faible énergie, on trouve également les ponts disulfures qui aident au repliement de la protéine. On dit que la protéine "native" présente une structure secondaire (conformation locale) avec des configurations de type hélice α ou brin β , une structure tertiaire (conformation globale) et même quaternaire dans certains cas.

Les protéines peuvent être groupées en deux catégories – fibreuses et globulaires. Les protéines fibreuses, ou scléroprotéines, ont une forme allongée, de structure secondaire régulière, et possèdent souvent des rôles structuraux dans la cellule et l'organisme. On peut citer les kératines, la fibroïne de la soie, les collagènes ou l'élastine. Les protéines globulaires, ou sphéroprotéines, qui constituent la plupart des protéines biologiquement actives (en particulier les enzymes) possèdent des structures tertiaires plus complexes et s'organisent sous formes compactes qui contiennent souvent des domaines bien définis. La structure tertiaire présente une importance capitale pour l'activité biologique des protéines : des acides aminés

très éloignés les uns des autres dans la séquence peuvent se trouver très proches en raison des repliements, et former ainsi des régions indispensables au fonctionnement de la protéine, par exemple le site actif (ou site catalytique) des enzymes. Des facteurs tels que l'entropie conformationnelle, l'enthalpie due aux liaisons internes non-covalentes, l'effet hydrophobe ou les ponts disulfures déterminent la stabilité de la protéine.

Enfin, plusieurs chaînes polypeptidiques peuvent s'associer de manière spécifique. Très souvent, l'activité biologique n'existe qu'au niveau de l'oligomère (dimère, tétramère, etc...) et non au niveau du monomère. Dans certains cas, la structure quaternaire résulte de l'assemblage de sous-unités peptidiques identiques, dans d'autres cas de sous-unités différentes.

I.1.1 Les modifications post-traductionnelles des protéines

La biosynthèse des protéines dans les cellules eucaryotes (**Figure 2**) s'effectue à partir de l'information génétique contenue dans l'ADN. Celle-ci est dans un premier temps transcrite en ARN messenger, étape se déroulant dans le noyau des cellules.

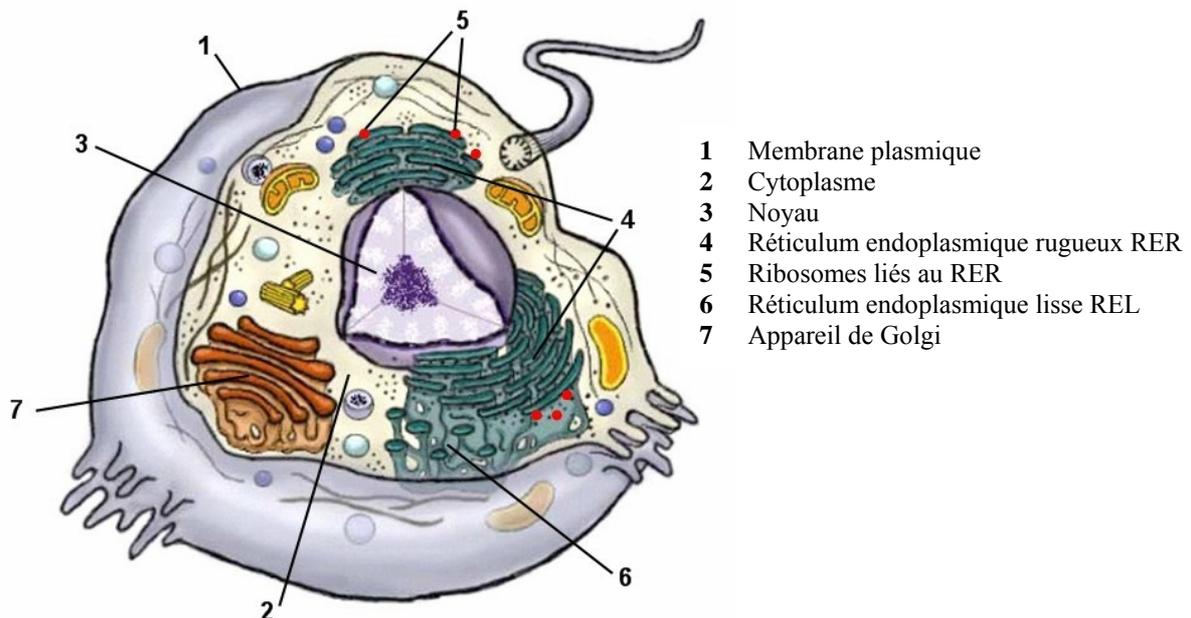


Figure 2 : Représentation schématique d'une cellule eucaryote.

Ensuite la biosynthèse se poursuit par une étape de traduction. Elle a lieu dans le cytoplasme, sur la surface externe du réticulum endoplasmique (RE) au niveau des ribosomes qui reçoivent les ARN de transfert (transporteurs des acides aminés) et l'ARN messager (qui contient l'information) nécessaires à la synthèse de la protéine. Le code génétique est la clé qui permet le passage de la séquence d'acides nucléiques à la séquence d'acides aminés. La traduction est un processus complexe qui nécessite l'action contrôlée de nombreuses macromolécules. Elle se décompose en trois étapes : l'initiation, l'élongation et la terminaison.

Dans une phase de maturation, les protéines subissent d'abondantes modifications covalentes pendant et après leur synthèse sur le ribosome. Ces modifications co- et post-traductionnelles – par la suite résumées sous le terme "modifications post-traductionnelles" – peuvent notamment servir :

- à la régulation de l'activité des protéines;
- à leur étiquetage pour une reconnaissance par des partenaires métaboliques ou par des systèmes de dégradation;
- à leur ancrage dans une membrane;
- à leur participation à des cascades de signalisation;
- à leur adressage pour qu'elles se rendent au bon endroit dans la cellule;
- à définir une identité immunologique (comme les groupes sanguins), ...

Les informations relatives à ces modifications ne sont pas portées par les acides nucléiques, mais dépendent de la nature de la séquence protéique, de la présence dans la cellule hôte de complexes enzymatiques, et parfois de systèmes membranaires.

Les modifications post-traductionnelles des protéines permettent une meilleure adaptation de la protéine à sa fonction.

Il existe différents types de modifications post-traductionnelles. Le **Tableau 1** présente une liste partielle de celles que l'on peut rencontrer dans une protéine. Cela peut être l'ajout/retrait de groupements divers, l'acylation (ajout d'acides gras divers), l'ajout/retrait de glucides (simples ou complexes) ou de polypeptides...

Modifications	Exemples
Glycosylation	<i>N</i> -glycosylation, <i>O</i> -glycosylation (détaillées par la suite)
Phosphorylation	sur des résidus Ser, Thr ou Tyr
Méthylation	sur des résidus Lys ou Arg
Addition de lipides	isoprénylation (farnésylation, géranylgeranylation) S-acylation (myristoylation, palmitoylation) O-acylation glypiation : greffage de l'ancre GPI (aussi un exemple de glycosylation)
Carboxylation	sur un résidu Glu
Hydroxylation	sur des résidus Pro ou Lys
Amidation	de l'extrémité C-terminale de la protéine
Protéolyse	du peptide signal, du résidu Met N-terminal
Sulfatation	sur un résidu Tyr
Acétylation	de l'extrémité N-terminale de la protéine ou sur un résidu Lys
Biotinylation	sur un résidu Lys
Formation des ponts disulfures	entre des résidus Cys

Tableau 1 : Exemples de modifications post-traductionnelles des protéines.

La découverte, le développement, la production et les applications cliniques des protéines recombinantes à intérêt thérapeutique chez l'Homme constituent un domaine de recherches scientifiques et médicales en plein essor. Cela a engendré un considérable intérêt dans l'étude des implications biologiques et thérapeutiques des modifications post-traductionnelles des protéines, en particulier la glycosylation, sans doute la modification la plus sophistiquée. Dès lors, de nombreuses études sont apparues dans la littérature [1].

1.1.2 La glycosylation des protéines

De très nombreuses protéines, dotées ou non d'une fonction enzymatique, sont des glycoprotéines, et la présence de glucides liés à la chaîne polypeptidique accroît encore leur diversité. La plupart des glycoprotéines sont typiquement des protéines sécrétées à l'extérieur des cellules (enzymes, hormones, transporteurs, anticorps, ...), des protéines localisées à la surface cellulaire ou dans la membrane plasmique. Elles sont très largement répandues dans les tissus animaux et végétaux, et présentes également chez les micro-organismes et les virus [2].

La glycosylation des protéines est l'une des modifications post-traductionnelles les plus courantes puisque l'on estime que près de 50% des protéines produites par la cellule

seraient glycosylées [3]. Depuis la description de la liaison GlcNAc(β 1-*N*)-Asn dans l'ovalbumine par Johansen et coll. en 1961 [4], de nombreuses autres liaisons sucre-protéine ont été décrites (**Figure 3**) [2].

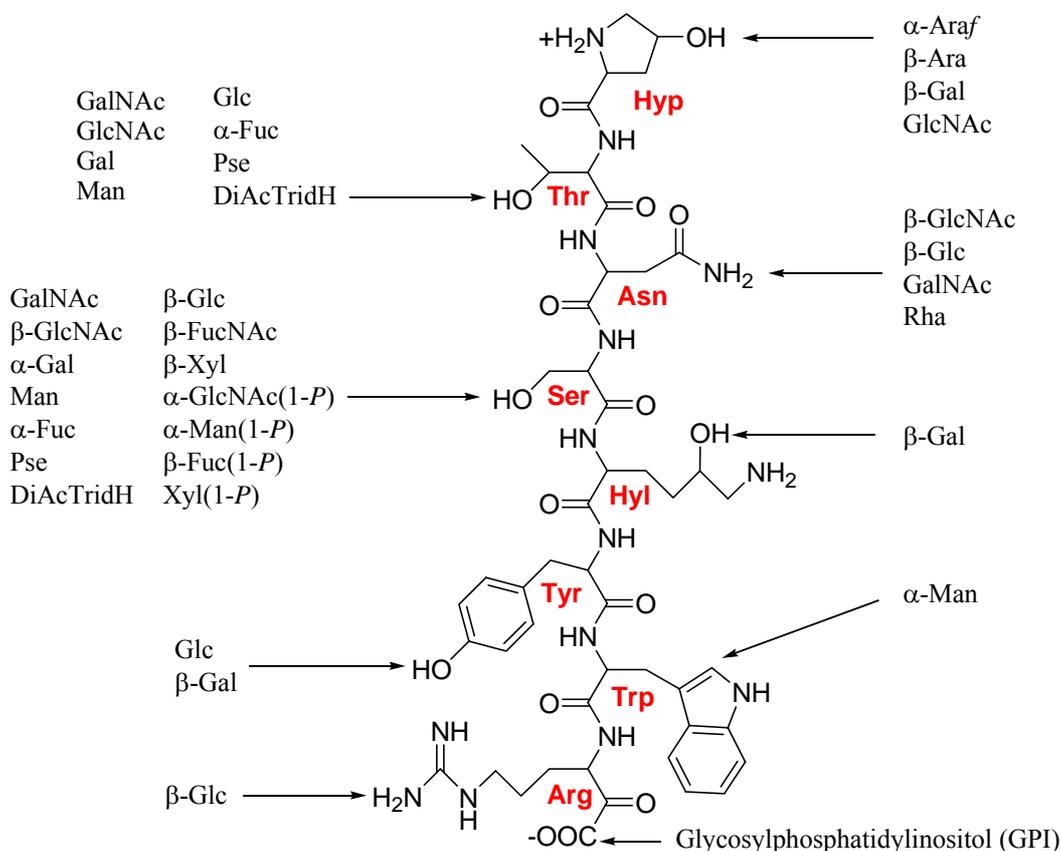


Figure 3 : Liaisons sucre-protéine connues chez les procaryotes et les eucaryotes^{a,b} [2,5].

^a Abréviations : Ara, L-arabinose; DiAcTridH, diActridésoxyhexose; *f*, furanose; Fuc, L-fucose; FucNAc, *N*-acétyl-L-fucosamine; Gal, D-galactose; GalNAc, *N*-acétyl-D-galactosamine; Glc, D-glucose; GlcNAc, *N*-acétyl-D-glucosamine; Man, D-mannose; P, phosphate; Pse, acide α -pseudaminique; Rha, L-rhamnose; Xyl, D-xylose.

^b Les deux configurations α et β peuvent être trouvées si elles ne sont pas indiquées précisément.

13 monosaccharides et 8 acides aminés différents participent à la formation des liaisons sucre-protéine. Sur la base de la liaison formée entre les parties peptidique et glycanique, on distingue différents types de glycosylation (**Figure 4**) :

- les liaisons *N*-glycosidiques, principalement sur les chaînes latérales de l'asparagine;
- les liaisons *O*-glycosidiques, sur les hydroxyles des chaînes latérales de la sérine, thréonine, ou d'acides aminés modifiés comme l'hydroxylysine (Hyl) et l'hydroxyproline (Hyp);
- les liaisons *C*-glycosidiques (*C*-mannosylation);

- les liaisons *P*-glycosidiques *via* un pont phosphodiester;
- la liaison entre l'ancre GPI et l'extrémité C-terminale de la protéine.

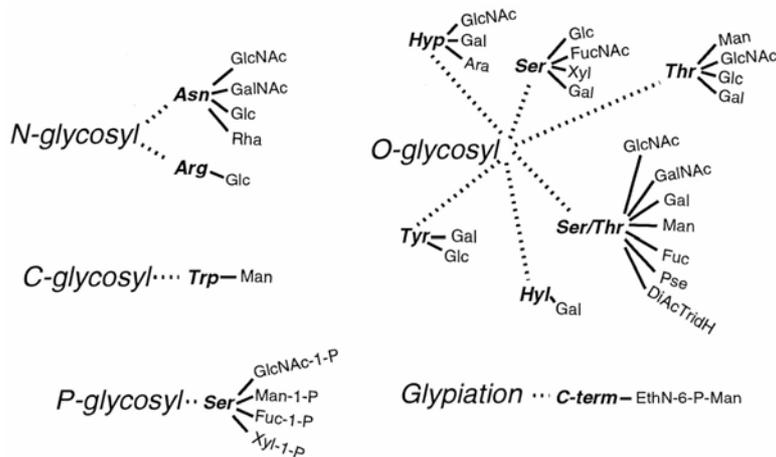


Figure 4 : Représentation des 5 types connus de glycosylation des protéines [2].

Les principaux types de glycosylation restent la *N*- et la *O*-glycosylation. Une même protéine peut porter à la fois des structures *N*- et *O*-glycaniques. La glycosylation s'effectue en des positions déterminées de la séquence, en nombre et en structure très variables d'une protéine à l'autre. Ainsi, une même protéine peut exister sous plusieurs isoformes ("glycoformes") [6] qui diffèrent seulement par leur composition en glycanes : on parle de macro- et de microhétérogénéité de la protéine [7].

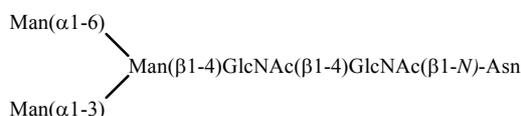
Dans les glycoprotéines courantes les glycanes sont répartis en courtes chaînes renfermant des unités caractéristiques. Les plus fréquemment rencontrées sont les suivantes :

- oses neutres : D-galactose, D-mannose, L-fucose, D-xylose, D-glucose;
- osamines (sous forme *N*-acétylées) : D-glucosamine, D-galactosamine;
- acides uroniques : acide D-glucuronique, acide L-iduronique;
- acides sialiques : acide *N*-acétylneuraminique (Neu5Ac), acide *N*-glycolylneuraminique (Neu5Gc), acides *N,O*-acétyl- et *N*-glycolyl, *O*-acétylneuraminiques.

Les unités osidiques des glycoprotéines sont liées entre elles selon une séquence et des modes de liaison définis. Cette spécificité structurale est à l'origine d'un véritable système de communication, celui des protéines des surfaces cellulaires, projetant vers l'extérieur un tapis

de filaments osidiques constituant autant d'antennes où viendront se fixer d'autres sucres ou des protéines. Ces reconnaissances spécifiques complexes constituent la base des communications entre cellules, entre cellules et certaines hormones, entre cellules et particules virales, ...

Les structures des glycanes à la surface des protéines sont très variées car elles sont spécifiques de chaque espèce et, pour une espèce donnée, de chaque type cellulaire. Toutefois, les glycanes présentent des caractères communs. Par exemple, quelle que soit leur origine, les glycanes dits de type mucine possèdent tous le "noyau" GalNAc(α 1-*O*)-Ser/Thr, et les glycanes des *N*-glycoprotéines contiennent tous le même "noyau" pentasaccharidique :



Sur ces noyaux se greffent des structures oligosaccharidiques spécifiques que l'on appelle "antennes".

Les protéines sont glycosylées d'une manière spécifique, déterminée notamment par leur cheminement à travers l'appareil de Golgi, où la phase essentielle des *N*- et *O*-glycosylations est effectuée. Cette modification est catalysée principalement par les enzymes de la famille des glycosyltransférases. La localisation cellulaire ordonnée et l'activité des enzymes impliquées dans la glycosylation définissent la séquence et la structure des glycanes associés à la protéine. A ce stade, la *N*- et la *O*-glycosylation sont traitées séparément du fait des différences existant dans la biosynthèse et les structures des glycanes.

1.1.2.1 La N-glycosylation

La *N*-glycosylation, schématisée sur la **Figure 5**, débute dans le réticulum endoplasmique – elle est donc en cela co-translationnelle – mais elle se poursuit dans l'appareil de Golgi où la maturation des *N*-glycanes de la protéine a lieu – elle est donc également post-translationnelle [8].

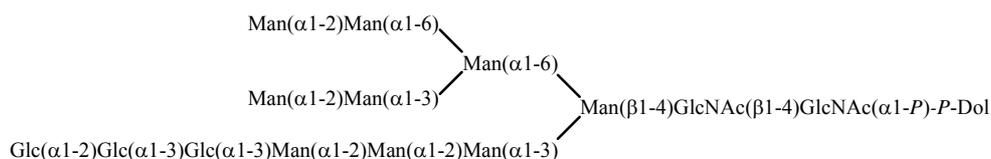


Figure 7 : Structure du précurseur tétradécasaccharidique des glycanes des *N*-glycoprotéines.

Dans un premier temps, ce précurseur est synthétisé grâce à un jeu de réactions complexes utilisant de nombreuses enzymes. La synthèse se fait par additions successives des différents monosaccharides du précurseur (**Figure 5**). Elle commence sur la face cytosolique du réticulum par le couplage entre le dolichol-phosphate (Dol-P) et le *N*-acétylglucosamine-1-phosphate (GlcNAc-P), pour donner la combinaison GlcNAc-P-P-Dol [10]. Cette réaction clé, qui amorce la préparation des *N*-glycosylations ultérieures, est inhibée par un antibiotique initialement découvert comme composé antiviral : la tunicamycine (analogue de l'UDP-GlcNAc). Les résidus suivants, une *N*-acétylglucosamine et cinq mannoses sont ajoutés de façon séquentielle par les glycosyltransférases et à partir des nucléotides-sucres correspondants. Le produit résultant est transféré par translocation de la face externe à la face interne du réticulum endoplasmique. Dans les réactions suivantes, quatre mannoses et trois glucoses sont ajoutés séquentiellement.

Le précurseur tétradécasaccharidique tri-antenné de structure $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ est ensuite transféré d'un bloc sur un résidu asparagine d'une protéine en cours de synthèse – sécrétée à travers la membrane du réticulum – par l'action d'un complexe enzymatique appelé oligosaccharyltransférase (OST) [11,12] comprenant au moins sept protéines (neuf chez la levure; seulement cinq sont identifiées chez l'homme) et lié à la membrane du réticulum endoplasmique. Cette étape enzymatique apparaît conservée chez les eucaryotes (des levures aux mammifères).

Le résidu asparagine, site de la *N*-glycosylation, est situé dans une séquence particulière reconnue par le complexe OST. Cette séquence consensus (ou séquon) est constituée du tripeptide Asn-X-Ser/Thr, où X peut être n'importe quel acide aminé excepté la proline [13] (ce dernier ne peut d'ailleurs pas être présent immédiatement après la séquence du tripeptide). Toutefois, même si la glycosylation a lieu dans le réticulum endoplasmique sur une chaîne naissante d'une protéine, la structure primaire du polypeptide n'est pas le seul

facteur déterminant de la *N*-glycosylation de la protéine car des résidus asparagine situés dans une séquence consensus et qui entrent dans le réticulum endoplasmique ne sont pas nécessairement glycosylés.

Des études mécanistiques ont été menées sur la glycosylation catalysée par le complexe OST de peptides modèles [14] et différents mécanismes sont proposés pour l'activation de l'asparagine par l'oligosaccharyltransférase [15-17]. A l'heure actuelle, de nombreux aspects restent encore très discutés.

Dès que l'oligosaccharide a été transféré sur le résidu asparagine de la protéine en cours de synthèse dans le réticulum endoplasmique, les trois résidus glucose en bout de chaîne et un résidu mannose sont éliminés par l'action des glycoside hydrolases correspondantes. La structure Glc₃ se comporte comme un signal indiquant que le précurseur est prêt pour le transfert. Son hydrolyse avertit ensuite de la fin du transfert et déclenche le début de la migration vers l'appareil de Golgi.

Après l'élimination des deux premiers résidus glucose, les glycoprotéines néosynthétisées subissent un "contrôle qualité" destiné à vérifier le bon repliement de la glycoprotéine. Ce contrôle fait intervenir des protéines chaperonnes, la calnexine et la calréticuline [18]. Les glycoprotéines mal repliées entrent alors dans un cycle de re-glycosylation/déglycosylation qui leur permet ou non d'acquérir une bonne conformation. A l'issue de ce contrôle, les molécules mal repliées sont dirigées vers le cytosol pour y être dégradées. Au contraire, les molécules correctement repliées continueront leur processus de glycosylation et seront dirigées vers l'appareil de Golgi. Une dernière étape dans le réticulum endoplasmique est l'action d'une mannosidase pour aboutir à la structure Man₈GlcNAc₂.

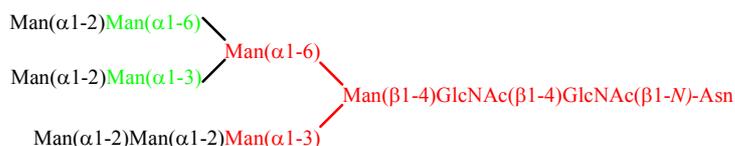
Une fois glycosylée par ce précurseur, la protéine va migrer vers l'appareil de Golgi où elle va subir un processus de maturation des *N*-glycanes – modifications de la structure oligosaccharidique – qui diffère selon le compartiment du Golgi et qui aboutit à une glycosylation terminale (par opposition à la glycosylation primaire ayant lieu dans le réticulum endoplasmique). Parmi ces modifications, on trouve l'hydrolyse de résidus mannose, l'addition d'autres monosaccharides (*N*-acétylglucosamine, galactose, fucose), l'addition d'acide sialique en dernier s'il y a lieu. Cette maturation est complexe et met en jeu une batterie d'au moins 8 à 10 enzymes (glycosyltransférases et glycoside hydrolases).

Chez les mammifères, la maturation des *N*-glycoprotéines dans l'appareil de Golgi conduit à la synthèse de trois principaux types d'oligosaccharides (**Figure 8**):

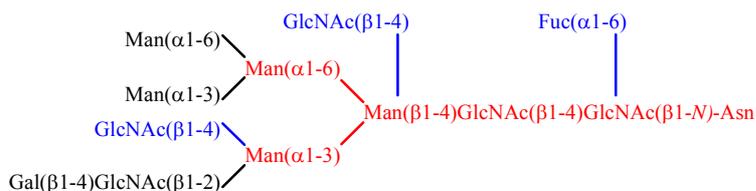
- les *N*-glycanes de type oligomannosidique;
- les *N*-glycanes de type hybride;
- les *N*-glycanes de type complexe (ou type *N*-acétyllactosaminique).

La synthèse de ces trois types de *N*-glycanes est finement régulée par l'action (voire la compétition) de glycoside hydrolases et de glycosyltransférases.

***N*-glycane de type oligomannosidique :**



***N*-glycane de type hybride :**



***N*-glycane de type complexe :**

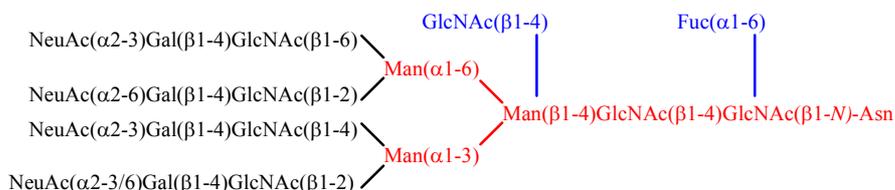


Figure 8 : Structures des trois types de *N*-glycanes rencontrés chez les mammifères.

En rouge, le noyau pentasaccharidique;
 en vert, les résidus retrouvés dans tous les *N*-glycanes de type oligomannosidique;
 en bleu, les résidus plus ou moins présents.

Mis à part le noyau chitobiosyle GlcNAc(β1-4)GlcNAc, les *N*-glycanes de type oligomannosidique contiennent presque uniquement des résidus mannose, liés entre eux par une grande variété de liaisons glycosidiques.

Les *N*-glycanes de type complexe contiennent en plus du noyau pentasaccharidique commun à toutes les *N*-glycoprotéines, des résidus galactose et *N*-acétylglucosamine et peuvent posséder à leurs extrémités des résidus galactose, *N*-acétylgalactosamine, fucose

et/ou des acides sialiques. Ces *N*-glycanes peuvent exister sous la forme de structures bi-, tri-, tétra- (cas de la **Figure 8**) ou penta-antennées.

Les *N*-glycanes de type hybride possèdent généralement les caractéristiques des *N*-glycanes de type oligomannosidique sur la branche Man(α 1-6) et des *N*-glycanes de type complexe sur la branche Man(α 1-3).

Plusieurs sites de *N*-glycosylation sur la même protéine peuvent être glycosylés de manière différente. La séquence peptidique, mais surtout la conformation de la protéine, peut influencer sur les différentes étapes réticulaires et golgiennes de la *N*-glycosylation en favorisant ou non l'accessibilité du substrat aux différentes glycosyltransférases. Le type de cellule ainsi que l'état physiologique de la cellule ont une influence sur la glycosylation [19]. D'autres facteurs incluant le métabolisme des nucléotides-sucres, leur transport dans le réticulum endoplasmique ou l'appareil de Golgi, la localisation et le taux d'expression des différentes glycosyltransférases contribuent probablement à l'hétérogénéité des structures glycaniques observées dans les *N*-glycoprotéines.

1.1.2.2 La O-glycosylation

Les résidus glycaniques à la surface des protéines peuvent également être liés par une liaison de type *O*-glycosidique à la chaîne latérale d'acides aminés tels que la sérine ou la thréonine (**Figure 3**) [20]. Dans le cas du collagène, on peut trouver également des *O*-glycanes liés à la chaîne latérale de l'hydroxylysine ou de l'hydroxyproline.

Les liaisons *O*-glycosidiques sont diverses par leur structure, et sont caractérisées par leur fragilité en milieu alcalin. La *N*-acétylgalactosamine liée en α à la thréonine est une situation fréquente, mais ce n'est pas la seule. On la trouve par exemple dans les protéines des groupes sanguins, des globulines plasmatiques, dans les glycoprotéines membranaires. Chez les levures et les champignons, c'est souvent le mannose qui est lié à la sérine ou la thréonine. Chez les plantes, l'arabinose est souvent lié à l'hydroxyproline par une liaison plus résistante à l'hydrolyse alcaline que les liaisons *O*-glycosidiques communes.

Les *O*-glycanes ne sont pas aussi faciles à classer que les *N*-glycanes, du fait de leur très grande variété. En effet, les structures liées aux protéines peuvent être aussi simples qu'un

monosaccharide (par exemple le glucose ou la *N*-acétylglucosamine) ou peuvent contenir plusieurs monosaccharides. Les structures glucidiques des groupes sanguins sont un exemple de *O*-glycanes. Ces derniers peuvent être des oligosaccharides riches en galactose et en *N*-acétylgalactosamine, où le fucose et l'acide sialique occupent ça et là des positions terminales.

La *O*-glycosylation de type mucine reste l'une des plus courantes et donc des plus étudiées. Les structures *O*-glycaniques sont classiquement décomposées en trois parties : le noyau, le squelette – qui peut être linéaire ou branché – et la périphérie. Les structures les plus couramment rencontrées peuvent être classées dans huit groupes, selon leur noyau (**Figure 9**).

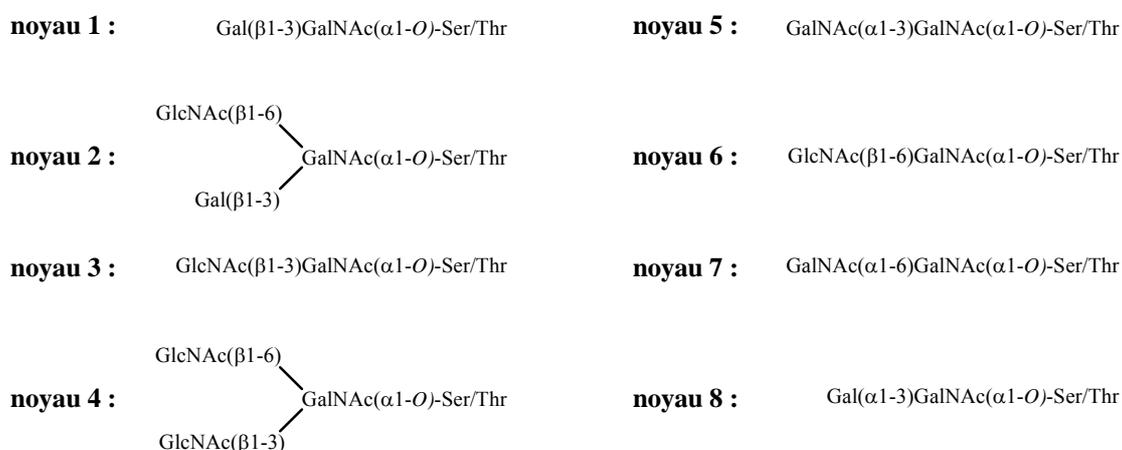


Figure 9 : Structures des noyaux identifiées à ce jour dans les chaînes *O*-glycaniques de type mucine.

Les noyaux 1 à 4 sont les plus courants.

Les signaux qui déterminent quel acide aminé sera *O*-glycosylé ne sont pas aussi bien connus que pour la *N*-glycosylation. La *O*-glycosylation a lieu sur des protéines déjà repliées et la structure tridimensionnelle de la protéine peut donc avoir une influence.

La *O*-glycosylation est une modification post-traductionnelle qui se déroule dans l'appareil de Golgi, mais la localisation exacte des enzymes impliquées dans la biosynthèse reste encore à préciser. Les monosaccharides sont transférés de manière individuelle et séquentielle sur la protéine à partir de nucléotides-sucres par des glycosyltransférases (distinctes ou non de celles intervenant dans la *N*-glycosylation).

I.1.2.3 Rôle des groupements glycaniques

Les fonctions biologiques associées aux glycanes des glycoprotéines sont aussi variées que leur diversité structurale [21]. Dans quelques cas, les oligosaccharides sont impliqués dans l'activité biologique de la protéine, mais ils sont plus souvent importants pour les propriétés physico-chimiques de la protéine, telles que la solubilité ou la stabilité.

Les caractéristiques des glycanes liés à la protéine sont :

- leurs structures variables, qui permettent une spécificité dans leurs interactions avec d'autres molécules;
- leur nature hydrophile, qui permet à la protéine de rester soluble dans des solutions aqueuses;
- le volume qu'ils occupent, qui affecte considérablement les propriétés de surface de la protéine à laquelle ils sont attachés.

On peut distinguer deux grandes familles fonctionnelles :

➤ *les fonctions structurales, de modulation et d'adressage*

Parmi les divers rôles structuraux des glycanes, on peut citer le cas des *O*-glycanes des mucines qui constituent une barrière de protection des cellules épithéliales. La forte glycosylation de certaines protéines présentes dans le sérum de poissons de l'Arctique leur confère des propriétés "anti-gel".

Les glycanes participent au maintien de la structure de la protéine dans une conformation biologiquement active, par des liaisons secondaires glycane-glycane ou glycane-protéine (de type liaison hydrogène ou ionique, interactions hydrophobes); la protéine est ainsi stabilisée (aide et contrôle du repliement de la protéine). Ils interviennent également dans l'oligomérisation de nombreuses protéines [22,23] et contribuent à leur solubilité, particulièrement dans les milieux aqueux.

Les glycanes permettent également une protection de la chaîne peptidique vis-à-vis de l'attaque d'enzymes protéolytiques. Ils modifient les propriétés antigéniques des protéines par masquage des épitopes (déterminants antigéniques) peptidiques. La glycosylation est alors un facteur influençant leur temps de demi-vie dans le sérum.

Dans certains cas, la glycosylation peut également moduler la fonction de la protéine. C'est le cas notamment de certaines hormones polypeptidiques qui ne seront fonctionnelles que si elles sont correctement glycosylées [24]. La glycosylation de la RNase est également un facteur qui module le taux de dégradation de l'ARN.

D'autre part, les glycanes participent également à l'adressage des protéines vers les différents compartiments, en assignant à la protéine une destination finale : intra-cellulaire, sécrétion vers l'extérieur ou même dégradation. L'implication directe des glycanes a été mise en évidence dans le processus d'adressage des enzymes lysosomales [25]. En effet, il a été démontré que les hydrolases destinées aux lysosomes sont des glycoprotéines dont le résidu mannose terminal est phosphorylé en position 6 (étiquette Man-6-P). Cette étiquette constitue un signal permettant le transport de la protéine de l'appareil de Golgi vers les lysosomes. Les glycanes jouent aussi un rôle dans le contrôle du développement embryonnaire et de la différenciation cellulaire.

➤ *les fonctions de reconnaissance, de nature endogène ou exogène (Figure 10)*

Les fonctions biologiques des oligosaccharides apparaissent comme essentielles à la vie des organismes pluricellulaires et de nombreuses études confortent leurs rôles dans la vie sociale des cellules [21]. Les glycoprotéines qui se trouvent à la surface des cellules sont typiquement ce que l'on appelle des protéines de communication. Par le biais de leurs glycanes, elles permettent aux cellules d'interagir avec les différents acteurs de son environnement : autre cellule, transporteur, anticorps, récepteur, hormone, enzyme, toxine, bactérie et virus. Les glycoprotéines de la surface cellulaire participent ainsi à la cohésion cellulaire [26,27], et donc à l'intégrité des tissus.

Les oligosaccharides interviennent dans de nombreux phénomènes de reconnaissance de type récepteur-ligand impliquant des lectines. Ces interactions sont le plus souvent attribuées aux unités terminales et concernent par exemple le trafic intracellulaire et la clairance des glycoprotéines sécrétées [28].

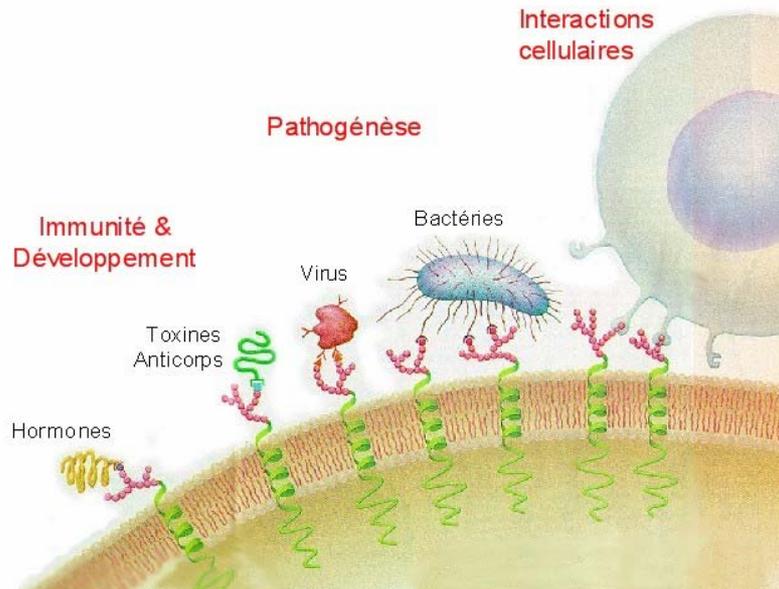


Figure 10 : Fonctions de reconnaissance des groupements glycaniques.

Les interactions des microorganismes pathogènes (virus, bactéries) avec la cellule hôte sont développées plus en détail dans le paragraphe suivant.

1.1.2.4 Participation de la glycosylation dans des infections et autres maladies

De nombreuses études ont mis en évidence les modifications à la fois qualitatives et quantitatives de la glycosylation dans de nombreux états physiologiques et pathologiques. Ces modifications sont observées notamment au cours de l'embryogénèse, au cours du vieillissement et du processus inflammatoire.

Un certain nombre de maladies sont associées à des déficiences ou défauts dans le métabolisme des glycoprotéines. Ainsi, les anomalies congénitales de glycosylation ou CDG ("Congenital Disorder of Glycosylation") [29-32] sont caractérisées par un groupe d'erreurs innées humaines affectant le processus de *N*-glycosylation. Des mutations dans la séquence de plusieurs gènes impliqués dans la biosynthèse du précurseur oligosaccharidique et son transfert (CDG de type I), ainsi que dans la maturation des *N*-glycoprotéines (CDG de type II) ont été caractérisées. En revanche, les liens entre les anomalies de glycosylation et les répercussions cliniques (notamment l'hypotonie, le retard psychomoteur, les neuropathies) n'ont pratiquement pas été explorés.

Il existe également d'autres maladies, qui ne sont peut-être pas directement associées à des erreurs dans la biosynthèse des glycoprotéines mais qui sont reliées à des motifs de glycosylation atypiques. Ainsi, des modifications de la glycosylation ont été observées dans les maladies auto-immunes et dans la majorité des cancers [33].

Cela se traduit souvent par l'apparition ou la prédominance de certains épitopes glycaniques comme les antigènes de type Le^x, sLe^x, Le^y et sLe^a qui augmentent dans les cancers du poumon [34], du colon [35] et de l'estomac [36]. Les fractions glycaniques des glycoconjugués membranaires sont profondément modifiées dans les cellules cancéreuses (variation de la masse moléculaire des glycanes des glycoprotéines et de leur taux de sialylation). Cette observation pourrait expliquer les modifications constatées de la perméabilité des membranes de ces cellules. Les glycanes interviennent également dans le phénomène d'inhibition de contact, avec un arrêt des mitoses quand les cellules normales viennent au contact les unes des autres. Ce processus est très perturbé dans les cellules cancéreuses (perte de l'inhibition de contact et diffusion métastatique).

D'autre part, un certain nombre de microorganismes pathogènes (virus, bactéries et parasites) se lient aux glycanes de la cellule hôte et/ou utilisent les glycanes pour ne pas être détectés par le système immunitaire. Les virus Influenza (grippe) se fixent sur une cellule épithéliale respiratoire grâce à un de leurs antigènes de surface, l'hémagglutinine, qui se lie à l'acide *N*-acétylneuraminique présent en position terminale des glycoprotéines de la membrane de la cellule hôte. Ces pathogènes sont également capables de modifier la glycosylation de leurs cellules hôtes par l'intermédiaire d'un autre de leurs antigènes de surface, la neuraminidase [37]. La bactérie *Helicobacter pylori*, responsable des ulcères de l'estomac, se fixerait à l'épithélium gastrique de son hôte en interagissant avec l'antigène de groupe sanguin Le^b [38]. Elle synthétise également les antigènes de groupe sanguin Le^x ou Le^y [39].

I.1.3 Utilisation des protéines : les protéines recombinantes

Avec une gamme de fonctions biologiques très large, les protéines sont communément utilisées de nos jours pour de multiples applications biotechnologiques ou médicales :

- en recherche : enzymes isolées de microorganismes utilisées dans des techniques de biologie moléculaire (par exemple les enzymes de restriction ou les ADN polymérases);
- dans l'industrie pharmaceutique : agents médicamenteux (hormone de croissance, facteur de coagulation, anticorps, ...);
- dans l'industrie agroalimentaire, dans les lessives.

Parmi les protéines ayant des applications médicales (**Tableau 2**), on peut citer l'insuline (traitement du diabète), le vaccin contre l'hépatite B, l'érythropoïétine (traitement de certaines formes d'anémies).

Types de médicaments	Fonctions
Cytokines	Régulation du système immunitaire (activation des cellules du système immunitaire pour effectuer différentes fonctions immunitaires) Exemples : interférons, interleukines...
Enzymes	Catalyse de réactions chimiques Exemples : enzymes lysosomales...
Hormones	Régulation des fonctions de l'organisme (élévation du niveau de certaines hormones) Exemples : insuline, œstrogène...
Facteurs de coagulation	Régulation de la coagulation du sang
Vaccins	Stimulation du système immunitaire pour la production d'anticorps spécifiques
Anticorps monoclonaux	Marquage d'une substance étrangère spécifique (telle que les cellules cancéreuses, les maladies infectieuses)

Tableau 2 : Types de médicaments à base de protéines.

A l'origine, les protéines d'intérêt thérapeutique (enzymes, interférons, hormones...) étaient extraites de sources naturelles telles que le sang, le placenta ou d'autres tissus humains, voire même, dans certains cas, de tissus animaux. Toutefois, cette approche est limitée par la quantité de tissus humains disponibles, et dangereuse. Les protéines extraites à partir de telles sources présentent en effet des risques de contamination par des virus, notamment ceux du sida et de l'hépatite B, ou par des prions. Les protéines d'origine animale présentent également le risque de déclencher des réactions allergiques chez l'homme.

L'avènement de la technologie de l'ADN recombinant dans les années 70, stimulée par la découverte d'outils permettant de couper (enzymes de restriction) et de lier (ligases) des fragments d'ADN [40], accompagné de méthodes de séquençage de l'ADN [41-43] a permis le développement de nouvelles approches de conception et de production de protéines, en utilisant divers systèmes d'expression. Les gènes gouvernant la synthèse des principales protéines humaines d'intérêt thérapeutique ont été identifiés et clonés, ce qui a permis une production de ces protéines en grandes quantités par des organismes hétérologues. Ces protéines recombinantes constituent alors une alternative de choix aux protéines naturelles. Aujourd'hui cette méthode est la plus utilisée pour produire des protéines d'intérêt thérapeutique.

Par exemple, l'insuline, au départ isolée du pancréas de bœuf ou de porc, peut être maintenant produite, grâce à cette technologie, dans la bactérie *Escherichia coli*. La protéine ainsi obtenue est identique à la protéine humaine et est préparée à l'échelle industrielle, sans le risque que le produit final soit par exemple contaminé par les virus présents chez les mammifères. Cela a ainsi abouti à la commercialisation de l'insuline recombinante à intérêt thérapeutique par Eli Lilly en 1982 [44]. Aujourd'hui, l'insuline utilisée pour traiter le diabète provient presque exclusivement de bactéries ou de levures génétiquement modifiées.

D'autres procédés sont susceptibles de concurrencer ces systèmes de production de molécules à intérêt thérapeutique :

- la synthèse chimique, incluant les hémisynthèses : celle-ci convient à la synthèse de peptides, avec une activité fonctionnelle équivalente aux protéines recombinantes; mais la synthèse chimique nécessite au préalable une connaissance des mécanismes d'action des protéines;
- la thérapie génique : elle consiste à administrer un gène vectorisé (vecteur viral ou vecteur de synthèse) plutôt qu'une protéine; cependant, cette thérapie se situe encore au stade de l'expérimentation clinique.

Les protéines recombinantes actuellement commercialisées sont produites majoritairement par les systèmes suivants : bactéries, levures ou cellules de mammifères. De nombreuses sociétés travaillent notamment au niveau des essais cliniques sur des protéines recombinantes produites par des animaux ou des plantes transgéniques. Toutefois, à ce jour aucun de ces produits n'a achevé son développement clinique. Actuellement, le nombre de molécules recombinantes en cours d'étude clinique s'élève à plusieurs centaines.

I.1.4 La production des protéines recombinantes

Les protéines recombinantes – d'intérêt thérapeutique ou alimentaire – sont produites à partir d'un transgène introduit dans un organisme hétérologue par génie génétique. Au sens large, un système adapté à la production d'une protéine recombinante donnée, est un procédé biotechnologique qui s'appuie principalement sur :

- l'emploi d'un vecteur d'expression (en général un plasmide ou un virus – pour les vecteurs eucaryotes –) jouant le rôle de transporteur du gène codant pour la protéine d'intérêt, avec en amont du gène inséré une séquence de contrôle de la synthèse de la protéine (promoteur inductible), très utile lorsque le produit est toxique pour la cellule;
- l'utilisation d'une cellule hôte ou d'un organisme complexe (bactéries, levures, insectes, plantes, animaux supérieurs), pour exécuter les instructions fournies par le transgène;
- une phase de production proprement dite permettant de synthétiser les protéines souhaitées;
- une phase de séparation et d'extraction de la protéine du milieu de culture ou du milieu intracellulaire, suivie de sa purification;
- enfin, une phase de caractérisation/validation des propriétés biochimiques, biophysiques, et fonctionnelles de la molécule recombinante obtenue.

Une vaste gamme de systèmes de production de protéines recombinantes est aujourd'hui disponible [45], chacun d'eux présentant des avantages et des inconvénients. Une attention particulière est portée au profil de *N*-glycosylation des protéines recombinantes obtenu dans les différents systèmes d'expression existants.

I.1.4.1 Les différents systèmes de production

I.1.4.1.1 Chez les bactéries

La bactérie *Escherichia coli* fut et reste encore le premier hôte utilisé pour l'obtention de protéines recombinantes [44,46]. On peut citer par exemple la production de l'hormone de croissance humaine, de l'insuline, de la chymosine, etc...

La génétique de cette bactérie est tout d'abord très bien connue. De nombreux vecteurs plasmidiques ont été construits et sont donc disponibles afin d'insérer et d'exprimer un gène étranger au sein de la bactérie. De nombreuses souches ont été améliorées afin d'optimiser l'expression de la protéine d'intérêt. Cette bactérie est facilement manipulable et se prête très bien à la culture de masse en fermenteur. Enfin, les taux d'expression obtenus sont élevés (jusqu'à plusieurs grammes de protéine par litre de culture).

Toutefois, la protéine synthétisée est souvent toxique pour la bactérie, empêchant d'atteindre des cultures cellulaires de haute densité. Une alternative peut être d'effectuer la production de la protéine en deux étapes : une phase de croissance de la biomasse puis une phase d'induction de la biosynthèse par ajout de l'inducteur du promoteur inséré en amont du gène codant pour la protéine.

D'autre part, il est souvent nécessaire de lyser la bactérie afin de récupérer la protéine qui est généralement peu ou pas sécrétée (ce qui induit des problèmes de purification, ou de solubilisation et de renaturation..., quelquefois au détriment des rendements). L'inconvénient majeur à son utilisation vient surtout du fait que *E. coli* n'effectue pas de modifications post-traductionnelles caractéristiques des protéines humaines (en particulier la glycosylation, la carboxylation, ...) car les enzymes responsables de ces modifications ne sont pas présentes naturellement dans la bactérie. Cela peut conduire au rejet des protéines d'intérêt thérapeutique par le système immunitaire, réduire leur durée de vie dans l'organisme voire leur activité biologique. La protéine produite peut notamment ne pas être correctement repliée.

Enfin *E. coli* étant une entérobactérie, il est nécessaire de s'assurer de l'absence d'endotoxines dans les protéines purifiées.

D'autres systèmes de production bactériens se développent également aujourd'hui : *Bacillus subtilis*, *Streptomyces*... A leur actif, ceux-ci possèdent des capacités de sécrétion supérieures à *E. coli*. Cependant, leur génétique est souvent moins connue, et le niveau de production de protéines recombinantes reste dans la plupart des cas inférieur à celui obtenu avec *E. coli*. L'utilisation d'extraits d'*E. coli in vitro* est également une option qui est envisagée.

Pour pallier les limitations du système bactérien, l'utilisation d'autres organismes producteurs non-bactériens comme les levures ou les champignons, puis de cellules de mammifères ou d'insectes, voire de plantes ou de mammifères a été développée.

I.1.4.1.2 Chez les levures et les champignons

Les levures offrent les mêmes facilités expérimentales ou industrielles que les bactéries, en particulier une culture aisée en fermenteurs à haute densité cellulaire, tout en possédant une machinerie cellulaire proche de celle d'une cellule humaine.

La levure de boulanger *Saccharomyces cerevisiae* est utilisée depuis des millénaires dans l'alimentation humaine. Des exemples de protéines recombinantes produites chez cet organisme sont l'antigène de surface du virus de l'hépatite B, l'insuline ou l'hirudine.

Le matériel génétique de cette levure est simple et elle ne présente aucune toxicité. De bons vecteurs d'expression (nombre de copies et stabilité) sont disponibles aujourd'hui. Les taux d'expression des protéines sont relativement élevés (de l'ordre de la centaine de milligrammes par litre de culture). La levure permet la production de protéines complexes et la réalisation de modifications post-traductionnelles (glycosylations simples, carboxylations, acylations, formation des ponts disulfures, rupture protéolytique...).

Néanmoins, les protéines synthétisées sont généralement localisées à l'intérieur du cytoplasme et une étape de lyse de la cellule est nécessaire afin de les récupérer. La sécrétion est possible, mais en général au détriment des rendements : elle fonctionne bien pour des petits polypeptides tels que l'insuline, mais beaucoup moins bien pour des protéines de masse moléculaire élevée.

D'autre part, la présence de protéases actives entraîne la dégradation des protéines hétérologues exprimées, réduisant ainsi les rendements.

Chez la levure *Pichia pastoris* [47], la surexpression des protéines est bien régulée. De hautes densités cellulaires peuvent être atteintes et la sécrétion est efficace. La fermentation de cet organisme est donc bon marché. Toutefois, les protéines peuvent être dégradées et la croissance des cellules est lente.

D'autres levures sont également utilisées : *Kluyveromyces lactis* (production de la chymosine, enzyme alimentaire), *Hanseluna polymorpha*, *Yarrowia lipolytica* (production de l'interféron- α) ...

Les champignons filamenteux sont des sécréteurs naturels d'un certain nombre de glycoprotéines (principalement des enzymes), souvent en quantités abondantes, et sont capables de se développer sur des milieux relativement peu coûteux. Des espèces tels que *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus oryzae* ou *Trichoderma reesei* sont particulièrement utilisées par les industries alimentaire et papetière pour la production d'enzymes extracellulaires. Des études ont donc été menées sur l'utilisation de ces organismes pour l'expression de protéines hétérologues [48].

Les levures et champignons filamenteux produisent typiquement des glycanes de type oligomannosidique, en ajoutant jusqu'à 100 résidus mannose (dans le cas des levures) au niveau du noyau pentasaccharidique (hyperglycosylation) [49,50]. Cette hypermannosylation peut engendrer une immunogénicité de la protéine chez l'homme. Chez les levures, le taux de glycosylation effectué par *Pichia pastoris* est plus faible (chaînes plus courtes) que chez *S. cerevisiae* [51]. De leur côté, les champignons filamenteux synthétisent le plus souvent des *N*-glycanes contenant un nombre relativement faible de résidus mannose (par comparaison avec les levures). Jusqu'à maintenant, aucun oligosaccharide de type complexe contenant des acides sialiques, du galactose, du fucose et de la *N*-acétylgalactosamine n'a été trouvé dans les glycoprotéines produites chez ces organismes.

Le processus de *N*-glycosylation des protéines chez les levures et les champignons filamenteux est semblable dans les étapes initiales au niveau du réticulum endoplasmique à

celui des mammifères, mais diffère de celui de l'homme au niveau de la maturation dans l'appareil de Golgi.

I.1.4.1.3 Dans les cellules de mammifères

Les cellules de mammifères ont permis de produire entre autres l'antigène du virus de l'hépatite B, l'hormone de croissance humaine, l'interleukine-2, l'interféron- β , l'érythropoïétine ou des facteurs de coagulation... Parmi les cellules de mammifères les plus utilisées, on peut citer les cellules CHO (chinese hamster ovary) et BHK (baby hamster kidney).

Les cellules CHO se prêtent très bien à la culture de masse en bioréacteur. Un avantage important de ces cellules réside dans leur capacité à synthétiser des protéines complexes (glycosylées) de masse moléculaire élevée. Certains vecteurs d'expression permettent d'introduire et de faire exprimer des gènes humains.

Toutefois les rendements sont faibles (de l'ordre de 10 milligrammes de protéines par litre de culture au maximum). En outre les cellules CHO s'avèrent fragiles et leur culture plus onéreuse.

D'autres systèmes tels que les cellules NS0 (murine myeloma) ou les cellules humaines (HeLa (human cervical carcinoma), Jurkat (human lymphocyte))... sont également étudiés dans le but d'y produire des protéines recombinantes.

L'équipement des cellules de mammifères les plus couramment utilisées (CHO et BHK) en enzymes impliquées dans la synthèse et le transport des nucléotides-sucres, ainsi qu'en glycosyltransférases, apparaît suffisant pour garantir une glycosylation des protéines recombinantes de type complexe avec un haut degré de sialylation α 2-3. Toutefois, ces systèmes manquent d'enzymes telles que les α 1,3/4-fucosyltransférases ou les α 2,6-sialyltransférases, qui transfèrent des motifs glycosidiques caractéristiques en position terminale des *N*-glycanes des tissus humains.

Chez la souris, la présence de résidus tels que l'acide *N*-glycolylneuraminique (Neu5Gc) sur les *N*-glycanes, dérivé de Neu5Ac potentiellement immunogénique, est une limitation à l'utilisation de ces cellules pour l'expression de protéines recombinantes.

Une autre voie consiste aujourd'hui à utiliser des bioréacteurs vivants en tant que système de production de protéines recombinantes. Il s'agit ici d'animaux vivants transgéniques (lapine, chèvre, brebis, vache) produisant dans leur lait, leur sang, leur urine ou même leur sperme, une protéine humaine.

Différentes techniques de transgénèse sont utilisées : micro-injection dans les pronuclei ou dans le cytoplasme de l'embryon, vecteurs rétroviraux... Les animaux transgéniques peuvent être utilisés afin de produire des protéines hétérologues : production du facteur IX de la coagulation dans le lait de brebis transgénique, de lactoferrine humaine obtenue dans le lait de vache transgénique, d'hormone de croissance humaine dans le lait de souris, d'hémoglobine humaine produite dans le sang de porc... L'intérêt d'une production de protéines recombinantes dans le lait ou le sang d'animaux transgéniques se heurte cependant à des niveaux d'investissements très lourds et à des problèmes éthiques.

I.1.4.1.4 Dans les cellules d'insectes

L'expression de protéines recombinantes dans les cellules d'insectes (*Spodoptera frugiperda* ou *Trichoplusia ni* par exemple) [52] est possible *via* l'utilisation de vecteurs d'expression développés à partir des Baculovirus pour l'infection des cellules. Ce système a l'avantage de permettre une bonne expression de la protéine, même si ensuite les rendements finaux restent faibles par rapport à ceux obtenus avec les bactéries ou les levures mais bien meilleurs que ceux obtenus avec les cellules de mammifères.

Les protéines peuvent être obtenues correctement repliées. Les cellules d'insectes, cultivées en suspension, sont capables de sécréter la protéine recombinante et d'effectuer des opérations post-traductionnelles (similaires à celles effectuées dans les cellules de mammifères), notamment la glycosylation de la protéine. Le coût de production, même s'il est plus important que pour des systèmes d'expression dans des bactéries ou des levures, reste plus abordable que la production dans des cellules de mammifères.

Toutefois, il est difficile de travailler avec ce système sensible. Il présente l'inconvénient d'être un système lytique, avec une protéolyse possible de la protéine cible, limitant ainsi les rendements.

La glycosylation dans les cellules d'insectes conduit à des structures glycaniques proches des structures d'origine humaine (type complexe) mais elle reste tout de même incomplète (types oligomannosidique et hybride), avec l'absence ou le très faible niveau de galactosyltransférases et de sialyltransférases dans les cellules. L'incapacité des cellules d'insectes à synthétiser une forme d'acide sialique reste un sujet de controverse [53]. Dans certaines lignées d'insectes, il est possible de trouver des résidus Fuc(α 1-3) qui s'avèrent être immunogènes chez l'homme, limitant l'utilisation de ce système pour la production de protéines thérapeutiques. Enfin, la présence d'une activité *N*-acétylglucosaminidase indésirable a été détectée dans ces cellules.

I.1.4.1.5 Chez les plantes transgéniques

Les industriels misent de plus en plus sur les plantes pour produire des protéines recombinantes, car elles présentent de nombreux avantages [54-57]. Les cellules végétales étant des cellules eucaryotes, elles disposent d'une machinerie cellulaire complexe et sophistiquée qui leur permet de produire des protéines ayant des propriétés thérapeutiques équivalentes aux protéines humaines.

Les différents végétaux utilisés peuvent être : le tabac, la luzerne, le colza, le maïs, le carthame, le soja, le riz ou la pomme de terre. Le feuillage ou la graine sont les cibles préférentielles de l'expression du transgène. La mise au point de plantes transgéniques à partir de végétaux permet de produire une variété de protéines recombinantes précieuses ("moléculture") : interféron, interleukine, facteur VIII de la coagulation, hirudine, anticorps ("planticorps")...

Les cellules végétales transformées peuvent être multipliées *in vitro* dans des bioréacteurs, mais elles sont de plus en plus cultivées dans des conditions qui permettent la régénération de plantes mûres, constituant ainsi l'usine de production.

Les protéines recombinantes produites chez les plantes transgéniques possèdent une très bonne qualité pharmacologique, les cellules végétales disposant des enzymes nécessaires à la maturation des protéines. Les modifications post-traductionnelles des protéines réalisées

dans des cellules de mammifères diffèrent toutefois de celles réalisées dans une cellule végétale, ce qui nécessite quelques ajustements pour éviter les risques d'allergie.

Contrairement au développement et à l'élevage d'animaux transgéniques, très coûteux, il est possible d'obtenir avec les plantes, de grandes quantités de biomasse dans des conditions économiques très compétitives (coût de production des protéines thérapeutiques réduit). En effet, l'extension de la culture des plantes productrices avec les infrastructures agricoles existantes permet une montée en puissance rapide et économique des capacités de production.

Les plantes ne sont pas porteuses des agents pathogènes couramment associés aux infections humaines. Il n'existe pas, en l'état actuel des connaissances, de pathogènes végétaux capables d'infecter l'animal ou l'homme. La synthèse de la protéine médicament peut être orientée dans les organes/tissus de la plante les plus faciles à stocker et pour lesquels les procédés d'extraction sont déjà bien maîtrisés dans le secteur agroalimentaire. Ainsi, lorsque la protéine peut être administrée oralement, il est possible d'orienter sa synthèse dans les organes comestibles de la plante : les semences de céréales, de haricot ou de pois, les tubercules de pomme de terre, les racines de carottes ou de betterave, mais également les fruits des bananiers par exemple.

Restent donc principalement à résoudre les problèmes de niveau d'expression des protéines (rendements faibles), ainsi que d'extraction et de purification. Les plantes transgéniques pourraient alors représenter un moyen de production efficace et peu coûteux.

La *N*-glycosylation des protéines chez les plantes [58] est semblable à celle des protéines humaines dans ses premières étapes. Les structures glycaniques de type complexe sont possibles. Toutefois, comme chez les cellules d'insectes, la glycosylation reste incomplète (présence des types oligomannosidique et hybride). L'addition de résidus Xyl(β 1-2) et Fuc(α 1-3) est caractéristique des *N*-glycanes des protéines chez les plantes. Ces deux résidus sont fortement immunogènes chez l'homme et compromettent donc l'utilisation de protéines thérapeutiques d'origine végétale. Enfin, les acides sialiques ne sont pas présents dans les protéines chez les plantes. Ces dernières expriment le plus souvent des structures terminales de type Le^a.

I.1.4.2 Comparaison des différents systèmes

Il est généralement reconnu qu'il n'y a pas de système d'expression universel disponible pour la production (biopharmaceutique) de protéines, et que la sélection d'un système d'expression de protéines hétérologues dépend d'un certain nombre de critères (**Figure 11**).

Tout d'abord, il y a les contraintes liées à la protéine exprimée : qualité de la protéine recombinante (repliement, modifications), ciblage cellulaire (réticulum endoplasmique, membrane cellulaire...), immunogénicité de la protéine. Ensuite il y a des contraintes d'ordre expérimental : facilité de purification, coût/rendement du système, cellules hôtes (densité de culture, temps de génération), vecteurs (stabilité, nombre de copies, promoteurs).

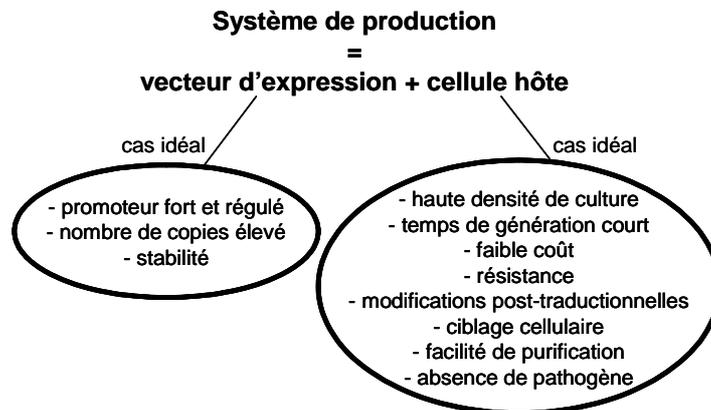


Figure 11 : Critères dans le choix du système d'expression de protéines hétérologues.

Les problèmes majeurs dans le choix du système restent le rendement, la protéolyse, la toxicité de la protéine envers la cellule hôte, les modifications post-traductionnelles et le repliement de la protéine. Aucun système d'expression ne satisfait tous les besoins, et la production d'une protéine X n'aura rien de comparable avec la production d'une protéine Y.

Un des plus importants critères à considérer dans le choix d'un système d'expression est de savoir si la protéine d'intérêt a besoin d'être (correctement) glycosylée pour son application thérapeutique [59,60]. En effet, la glycosylation module de nombreuses caractéristiques de la protéine : repliement, sécrétion, fonction biologique, temps de demi-vie, antigénicité...

Pour un certain nombre de glycoprotéines, la présence de sites spécifiques N-glycosylés est importante, mais la structure précise des glycanes n'influence apparemment pas directement la fonction biologique de la protéine.

Le glycane va donc influencer dans le choix de l'hôte (**Tableau 3**) [61]. Si la glycosylation n'a pas d'influence sur les fonctions biologiques de la protéine, alors l'expression chez la bactérie *E. coli* est préférée. Si la glycosylation n'est pas indispensable pour les fonctions biologiques mais peut aider à stabiliser la protéine, alors la levure sera choisie (utilisation de souches mutées pour éviter toutefois des chaînes oligomannosidiques trop longues).

Systèmes d'expression	Caractéristiques ^a
Bactérie	Pas ou peu de glycosylation
Levures, champignons filamenteux	Glycosylation de type oligomannosidique
Cellules d'insectes	Glycosylation de type complexe mais incomplète Résidus immunogènes
Plantes	Glycosylation de type complexe mais différente ou incomplète Résidus immunogènes
Cellules de mammifères Animaux	Glycosylation la plus proche Mais système très coûteux

Tableau 3 : Caractéristiques des profils de N-glycosylation des différents systèmes d'expression de protéines hétérologues.

^a Comparées avec le profil de glycosylation des protéines humaines

Dans le cas où la glycosylation est nécessaire à l'activité biologique de la protéine ou à l'augmentation de sa durée de vie plasmatique, les cellules de mammifères seront préférées.

Pour l'instant, les bactéries (*E. coli*) et les levures (*S. cerevisiae*) restent les systèmes de choix. Cependant certaines protéines, pour être actives, doivent présenter une glycosylation plus complexe et ne peuvent être produites que dans des organismes supérieurs, des animaux (cellules CHO), ou des végétaux.

I.1.5 Le contrôle de la glycosylation des protéines

De très nombreuses protéines, qu'elles soient dotées ou non d'une fonction enzymatique, sont des glycoprotéines, et la présence de glucides liés à la chaîne polypeptidique leur confère une plus grande diversité. Mis à part leur influence dans les propriétés physico-chimiques de la protéine, les glycanes possèdent entre autres un rôle dans la durée de vie de la protéine et dans sa fonction biologique.

La plupart des protéines thérapeutiques sont des glycoprotéines. D'une glycoprotéine à l'autre, le glycane peut avoir un rôle plus ou moins crucial et l'activité de la protéine mesurée *in vitro* ou établie *in vivo* pourra être totalement différente selon les structures glycaniques présentes.

Les résidus d'acides sialiques présents en position terminale des *N*-glycanes possèdent une grande importance dans les protéines thérapeutiques. En effet, la sialylation des protéines confère des propriétés (pharmacocinétiques) importantes aux glycoprotéines, avec des résidus de charge négative en position terminale de la structure oligosaccharidique. La machinerie nécessaire à la synthèse, à l'activation (sous forme de nucléotides-sucres) et à l'introduction des résidus sialylés (transporteurs, sialyltransférases) est faiblement représentée dans les différents systèmes d'expression des protéines, avec parfois la présence d'une activité sialidase. Ainsi, les protéines recombinantes produites sont souvent sous-sialylées (voire pas du tout) par rapport aux protéines humaines. Si elles sont produites dans des cellules de mammifères, leur sialylation diffère encore de celle effectuée chez l'homme.

Ce défaut de sialylation de la protéine augmente dramatiquement sa clairance, ce qui compromet ses effets pharmaceutiques. La désialylation des glycoprotéines circulantes provoque en effet leur reconnaissance instantanée par une lectine présente dans les membranes des hépatocytes [62] qui interagit spécifiquement avec les résidus galactose en position terminale non réductrice. Ce phénomène est suivi de l'internalisation des asialoglycoprotéines et de leur destruction par les enzymes lysosomales. Ainsi, la présence de résidus sialylés à l'extrémité non-réductrice des *N*-glycanes minimise la clairance de la protéine en masquant des résidus tels que le galactose, la *N*-acétylglucosamine ou encore le mannose qui peuvent être reconnus par des récepteurs.

Des études ont pu montrer que l'introduction de l'acide *N*-acétylneuraminique en position terminale sur des glycoprotéines permettait l'augmentation du temps de demi-vie dans l'organisme. Ainsi un analogue hypersialylé de l'érythropoïétine (EPO) possède un temps de demi-vie trois fois plus important que l'EPO humaine recombinante [63].

Un intérêt croissant a été porté à l'étude de la glycosylation des protéines, notamment pour comprendre l'influence de ces structures glycaniques à la surface des protéines dans l'activité biologique de celles-ci *in vivo*. Toutefois, ces études sont en partie freinées par la microhétérogénéité des structures glycaniques en surface.

Des recherches se sont donc orientées vers la préparation de glycoprotéines présentant des structures oligosaccharidiques bien déterminées et homogènes, très importantes dans le cas des protéines thérapeutiques. Différentes stratégies ont pour cela été développées [64,65].

De nombreuses équipes de recherche tentent également maintenant de travailler sur la modification du processus de *N*-glycosylation dans les différents systèmes de production de protéines recombinantes existants. Par un travail sur la machinerie cellulaire de glycosylation, leur but est ici "d'humaniser" le profil de glycosylation des protéines recombinantes, en se rapprochant le plus de celui présent dans les glycoprotéines naturelles et en évitant par ailleurs que les structures oligosaccharidiques ne soient immunogéniques chez l'homme.

1.1.5.1 Remodelage de la glycosylation : voie chimique et in vitro

Les diverses structures des *N*-glycanes (glycoformes) peuvent avoir différentes propriétés biophysiques et biochimiques. Toutefois, les détails des relations structure-fonction sont très peu compris, à cause de la faible disponibilité et de la microhétérogénéité des molécules disponibles. Il y a donc un besoin clair de méthodes synthétiques permettant de greffer sur des sites spécifiques de la protéine des structures oligosaccharidiques, naturelles ou modifiées, et homogènes.

Dans cet objectif, plusieurs méthodes de couplage chimiosélectives ("chemoselective ligation") ont été développées pour introduire des oligosaccharides sur des sites spécifiques de peptides ou de protéines [66]. Ainsi, la réaction sélective de β -*N*-glycosyl iodoacétamides [67,68] avec des groupements thiol de résidus cystéine, introduits par mutagenèse dirigée sur

des sites précis de glycosylation (naturels ou artificiels), a permis la synthèse de nouvelles glycoprotéines. D'autres types de greffage sur les résidus cystéine ont également été exploités, notamment avec la génération de ponts disulfures grâce à la réactivité de glycosyl méthylthio- [69] ou phénylthiosulfonates (-SSO₂R) [70], ou de glycosyl sélénylsulfides (-SSeR) [71]. La formation chimiosélective de liaisons de type oxime ou hydrazone pour la synthèse de glycoconjugués a également été développée [72], ainsi que le couplage sélectif entre un oligosaccharide activé par un groupe maléimide et la fonction thiol d'une cystéine [73].

Alternativement, l'introduction sur la protéine d'un groupement fonctionnel, avec une réactivité unique et qui n'est pas présent sur les chaînes latérales des acides aminés naturels, facilite ensuite la modification chimique sélective de la protéine. C'est le cas du groupement carbonyle d'une cétone qui peut réagir sélectivement avec des dérivés glycosidiques hydrazide et hydroxylamine. Pour cela, un acide aminé non naturel (tel que la *p*-acétyl-L-phénylalanine), portant sur sa chaîne latérale une fonction cétone, est introduit dans la séquence de la protéine au cours de sa biosynthèse [74]. Un mono- ou oligosaccharide activé réagit ensuite sélectivement avec le groupement carbonyle de la cétone.

Luchansky et coll. en 2004 [75] ont plutôt utilisé des monosaccharides modifiés, porteurs d'un groupement azide, qui sont métabolisés et incorporés dans des protéines recombinantes sécrétées par des cellules CHO. Le groupement est ensuite modifié, après récupération de la protéine d'intérêt, par un couplage de Staudinger avec un dérivé d'une phosphine.

Une autre stratégie consiste à utiliser comme structure de départ la glycoprotéine nouvellement produite contenant différentes structures glycaniques à sa surface. Il s'agit alors d'hydrolyser les oligosaccharides par une *endo-N*-acétylglucosaminidase pour obtenir un noyau simple homogène puis de les reconstruire par voie enzymatique [76]. Selon l'enzyme utilisée (EndoH ou EndoF), la protéine conserve un ou deux résidus GlcNAc du noyau pentasaccharidique. Ensuite, l'activité de transglycosylation (transfert d'oligosaccharide) des *endo-N*-acétylglucosaminidases (EndoA ou EndoM par exemple) permet la synthèse de néoglycoprotéines dont le profil de glycosylation homogène a été remodelé par transfert d'un oligosaccharide à partir d'un donneur de glycosyle (glycoprotéine) sur la protéine acceptrice. Cette approche est toutefois limitée par la spécificité de substrat de l'enzyme et par la quantité d'oligosaccharide donneur disponible.

D'autre part, une possibilité de remodelage du profil de glycosylation est l'isolement et l'utilisation *in vitro* de glycosyltransférases pour compléter les chaînes d'oligosaccharides de protéines recombinantes après purification [77]. Ces dernières sont des accepteurs pour l'action successive par exemple d'une β 1,4-galactosyltransférase et d'une sialyltransférase. Ces réactions sont effectuées sur les protéines natives (la partie glycanique pouvant avoir été hydrolysée par une *endo-N*-acétylglucosaminidase). Toutefois, dans ces conditions tous les glycanes ne seront peut-être pas accessibles aux enzymes, et des résidus antigéniques pourront encore être présents sur la glycoprotéine remodelée. Cette approche a notamment été utilisée par l'entreprise NEOSE^a dans sa technologie GlycoAdvanceTM.

Toutes ces stratégies de remodelage permettent, lors de la production de la protéine d'intérêt par un système d'expression hétérologue, de se focaliser sur l'amélioration du taux d'expression de la protéine en laissant de côté la qualité de sa glycosylation qui sera modifiée par la suite.

1.1.5.2 Ingénierie de la machinerie de glycosylation

Le concept général repose sur la modification génétique des voies métaboliques d'un organisme afin d'augmenter la production d'un métabolite d'intérêt [78-81]. Le but est ici d'améliorer la glycosylation, la productivité et la qualité des protéines recombinantes.

Dans le cas d'organismes pouvant être facilement modifiés génétiquement, la méthode envisagée serait de construire des mutants modifiés au niveau des étapes de glycosylation indésirables (pour éliminer les sucres potentiellement immunogènes) ou absentes [82], à la condition de connaître les voies de glycosylation ainsi que les gènes codant pour les enzymes impliquées (**Tableau 4**).

Il s'agit en fait de produire des structures oligosaccharidiques hybrides (type Man₅GlcNAc₂), semblables à celles des intermédiaires dans la *N*-glycosylation des protéines de mammifères, qui seront le point de départ utilisé pour la glycosylation de type complexe. Les glycosyltransférases nécessaires pour accomplir ces glycosylations de type complexe (**Tableau 4**) sont ensuite introduites par des étapes supplémentaires d'ingénierie métabolique.

^a Site Internet : <http://www.neose.com>

	Activités des enzymes dont les gènes sont surexprimés^a	Activités des enzymes dont les gènes sont délétés/inhibés
Levures	α 1,2-mannosidase (RE) transporteur de l'UDP-GlcNAc (Golgi) <i>N</i> -acétylglucosaminyltransférase (Golgi) UDP-galactose 4-épimérase (Golgi) β 1,4-galactosyltransférase (Golgi)	α 1,6-mannosyltransférase (RE) α 1,3-mannosyltransférase (RE)
Champignons filamenteux Cellules de mammifères	<i>N</i> -acétylglucosaminyltransférase <i>N</i> -acétylglucosaminyltransférase III GlcNAc 2-épimérase UDP-GlcNAc 2-épimérase UDP-glucose 4-épimérase UDP-glucose pyrophosphorylase phosphoglucomutase glucose 6-phosphatase β 1,4-galactosyltransférase transporteur du CMP-Neu5Ac α 2,6-sialyltransférase α 2,3-sialyltransférase	CMP-Neu5Ac hydroxylase sialidase
Cellules d'insectes	<i>N</i> -acétylglucosaminyltransférase II β 1,4-galactosyltransférase GlcNAc 2-épimérase <i>N</i> -acétylneuraminyl lyase CMP-Neu5Ac synthétase α 2,3-sialyltransférase α 2,6-sialyltransférase	α 1,3-fucosyltransférase <i>N</i> -acétylgalactosaminyltransférase β - <i>N</i> -acétylglucosaminidase
Plantes	α -mannosidase I et II <i>N</i> -acétylglucosaminidase II <i>N</i> -acétylglucosaminyltransférase I β 1,4-galactosyltransférase GlcNAc 2-épimérase <i>N</i> -acétylneuraminyl lyase CMP-Neu5Ac synthétase α 2,6-sialyltransférase	α 1,3-fucosyltransférase β 1,2-xylosyltransférase

Tableau 4 : Modifications génétiques utiles pour améliorer la *N*-glycosylation des protéines recombinantes selon le système d'expression utilisé [82].

^a La liste comprend tous les gènes potentiels pour la surexpression. Toutefois, il peut être seulement nécessaire de surexprimer un gène de la liste ou une sélection de plusieurs gènes pour améliorer la glycosylation sous différentes conditions de culture.

La difficulté est ici d'exprimer la glycosyltransférase hétérologue dans un compartiment spécifique de la cellule, que son activité soit adéquate et que les substrats donneurs de cette enzyme soient présents (nécessité d'un transporteur).

La sialylation terminale au sein des différents systèmes d'expression représente l'étape la plus délicate à effectuer. En effet, la machinerie cellulaire nécessaire y est incomplète voire inexistante (**Figure 12**).

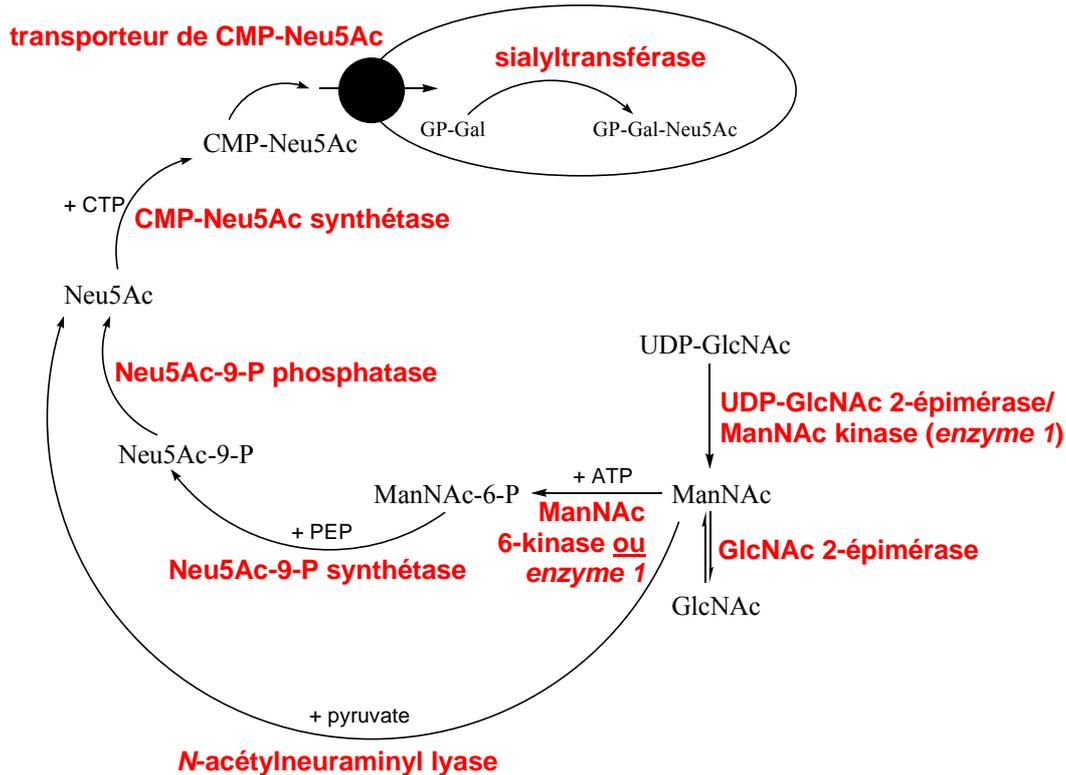


Figure 12 : Routes potentielles de la biosynthèse de l'acide *N*-acétylneuraminique.

Toutefois une alternative pourrait être de rajouter *in vitro* l'acide *N*-acétylneuraminique aux protéines recombinantes par une voie enzymatique, en utilisant par exemple une sialyltransférase. Cependant, le coût du nucléotide-sucre utilisé reste un facteur limitant, même si des procédés ont été développés pour le produire à moindre coût.

L'ingénierie de la glycosylation chez les levures [51,83] et les champignons (**Tableau 4**) permet dans un premier temps de limiter le nombre de résidus mannoses additionnés. Ensuite, des enzymes nécessaires à la galactosylation et à l'addition de GlcNAc sont ajoutées.

Des travaux ont été effectués notamment sur les champignons filamenteux *A. nidulans* ou *T. reesei* [48]. La levure *Pichia pastoris* est quant à elle la levure la plus étudiée et modifiée [84-86]. Cependant, l'ajout *in vitro* de l'acide *N*-acétylneuraminique aux protéines recombinantes demeure effectué par synthèse enzymatique. En effet, la biosynthèse *in vivo* du substrat donneur CMP-Neu5Ac reste très difficile, impliquant de nombreuses enzymes (**Figure 12**).

Dans le cas de l'expression de protéines dans les cellules animales, des travaux ont été effectués sur la surexpression d'une α 2,3-sialyltransférase et d'une β 1,4-galactosyltransférase [87], toutes deux déjà présentes dans le génome mais dont l'activité peut énormément varier avec les conditions de culture. Cela a pour conséquence une grande variabilité au niveau de la présence des résidus terminaux Gal et Neu5Ac et donc une grande hétérogénéité dans les structures glycaniques présentes à la surface des protéines. Les recherches sont pour la plupart tournées vers l'optimisation de la galactosylation et de la sialylation [88] (introduction d'une α 2,6-sialyltransférase [89]).

Dans les cellules d'insectes, les glycoprotéines produites présentent des unités potentiellement immunogènes chez l'homme, d'où la suppression des glycosyltransférases responsables de l'introduction de ces sucres. La sialylation des protéines a également été réalisée par l'addition de trois gènes supplémentaires (codant pour les enzymes GlcNAc 2-épimérase, *N*-acétylneuraminyl lyase et CMP-Neu5Ac synthétase) pour l'introduction de l'acide *N*-acétylneuraminique [90,91].

Chez les plantes transgéniques [92], une première stratégie employée pour la production de protéines avec des *N*-glycanes remodelés est de prévenir l'addition de glycanes (notamment antigéniques), en stockant la protéine d'intérêt thérapeutique dans le réticulum endoplasmique des cellules végétales (à ce moment-là, les *N*-glycanes ne seront pas de type complexe). Une autre stratégie repose sur l'inhibition de certaines glycosyltransférases de l'appareil de Golgi. Enfin, l'expression de glycosyltransférases qui pourraient compléter et/ou entrer en compétition avec la machinerie endogène de maturation des *N*-glycoprotéines permet de contrôler la glycosylation de la protéine. L'addition de résidus sialylés est effectuée de la même manière que pour les cellules d'insectes, avec l'introduction de la machinerie nécessaire pour cela.

L'utilisation de systèmes d'expression génétiquement modifiés au niveau de la glycosylation devrait permettre de produire des glycoprotéines avec une exceptionnelle homogénéité dans leurs structures glycaniques. Ces systèmes pourraient alors être utilisés pour générer des bibliothèques de glycoformes homogènes d'une seule et même protéine afin d'essayer de comprendre la relation entre une structure particulière de glycanes et son effet sur le comportement pharmacocinétique et l'activité biologique d'une glycoprotéine. L'objectif final reste bien sûr la production en grande quantité de protéines d'intérêt biomédical possédant une structure parfaitement maîtrisée.

1.1.5.3 Conclusion

Comme nous l'avons vu précédemment, le contrôle du profil de glycosylation des protéines recombinantes constitue un thème de recherche en plein essor et diverses méthodes sont en développement à l'heure actuelle. L'ingénierie de la machinerie de glycosylation du système d'expression utilisé pour la production de la protéine consiste à introduire ou éliminer certaines activités enzymatiques nécessaires à la préparation de *N*-glycoprotéines "humanisées". Le remodelage peut également être effectué par voie chimique ou catalysé par des enzymes, comme par exemple des glycosyltransférases ou des *endo-N*-acétylglycosaminidases qui permettent l'introduction de structures oligosaccharidiques homogènes à la surface de la protéine.

Une méthode alternative pourrait être l'utilisation des glycosynthases qui sont des glycoside hydrolases mutées ne possédant plus d'activité hydrolytique tout en gardant la capacité de catalyser la synthèse de liaisons glycosidiques.

1.2 Les glycosynthases : un formidable outil de synthèse

Malgré plus d'un siècle de découvertes et d'innovations, la synthèse chimique d'oligosaccharides, conjugués ou non, demeure fastidieuse, car la protection et l'activation sélective des groupements hydroxyles pour la formation d'une liaison glycosidique impliquent un grand nombre d'étapes. Les synthèses sont donc souvent longues et les rendements faibles. Une alternative consiste à utiliser *in vitro* la spécificité des enzymes pour l'établissement de ces liaisons.

Les glycosyltransférases constituent une première possibilité pour la synthèse enzymatique d'oligosaccharides. Ces enzymes utilisent des nucléotides-sucres comme donneurs activés qu'elles transfèrent sur un sucre accepteur. L'utilisation de ces enzymes, même si elles sont très répandues, est limitée à la synthèse sur de petites quantités du fait de la difficulté à obtenir des quantités suffisantes d'enzymes purifiées. De plus, les nucléotides-sucres restent chers et dans quelques cas sont instables. Finalement, les glycosyltransférases présentent souvent une capacité limitée à transférer des donneurs chimiquement modifiés et sont extrêmement limitées dans leur spécificité de substrat au niveau de l'accepteur.

Une approche alternative consiste à utiliser une autre classe d'enzymes, les glycoside hydrolases.

1.2.1 Les glycoside hydrolases

Les *O*-glycoside hydrolases sont capables d'hydrolyser les liaisons glycosidiques entre deux ou plusieurs unités sucre ou entre un sucre et son aglycone.

1.2.1.1 Les classifications des glycoside hydrolases

Dans la nomenclature de l'Union Internationale de Biochimie et de Biologie Moléculaire (IUB-MB)^b basée sur le type de réaction catalysée et la spécificité de substrat, les glycoside hydrolases possèdent un numéro du type EC 3.2.1.x. Les trois premiers chiffres indiquent qu'elles hydrolysent des liaisons *O*-glycosidiques, le dernier (x) est variable et dépend du substrat transformé.

Cette classification systématique permet de nommer précisément la spécificité de substrat d'une enzyme et possède en outre l'avantage de pouvoir classer rapidement une enzyme. Cependant, ce système ne tient pas compte des similarités de séquence protéique et donc de la structure tridimensionnelle. De plus, cette classification n'est pas systématique quant au mécanisme d'action des enzymes. Elle paraît mal adaptée aux glycoside hydrolases qui peuvent agir sur différents substrats et ne reflète pas les aspects structuraux des enzymes.

^b Site Internet : <http://www.chem.qmul.ac.uk/iubmb/enzyme/>

De nombreuses glycoside hydrolases possèdent une structure modulaire qui consiste en un domaine catalytique et un ou plusieurs domaines non-catalytiques, dont certains sont impliqués dans la fixation du substrat.

Depuis 1991 [93], une classification des domaines catalytiques des glycoside hydrolases, basée sur les similarités de séquences en acides aminés, a donc été proposée, afin de mieux refléter la structure de ces enzymes. De grandes divergences de séquences sont souvent détectées entre les glycoside hydrolases. Il est donc parfois nécessaire, pour les comparer, d'utiliser une technique différente de celles utilisées couramment et basées sur la maximisation du pourcentage en acides aminés identiques, qui ne prennent pas en compte les informations structurales. Dans la technique utilisée ici, les similarités sont appréciées par la méthode HCA ("Hydrophobic Cluster Analysis") ou encore analyse de groupes hydrophobes [94,95], qui est basée sur la détection de segments structuraux constituant le cœur hydrophobe des protéines globulaires. Grâce à cette approche, des similarités dans le repliement tridimensionnel peuvent être détectées entre des protéines possédant des identités de séquences très basses (<20%).

Cette classification diffère de la classification de l'IUB-MB par le fait que des enzymes de spécificité différente peuvent appartenir à la même famille. Elle permet de prendre en considération la structure tridimensionnelle ainsi que le mécanisme moléculaire d'action des glycoside hydrolases.

Elles sont ainsi regroupées en familles (notées GH) en fonction des similarités dans leur séquence d'acides aminés. Certaines glycoside hydrolases sont multifonctionnelles, contenant des domaines catalytiques qui appartiennent à différentes familles GH.

Puisqu'il est admis qu'il existe une relation directe entre similarités de séquence en acides aminés et similarités de repliement [96], il peut être avancé que les membres d'une même famille possèdent des repliements similaires, ce qui permet de prédire la structure tridimensionnelle générale et la structure du site actif des membres de chaque famille, si cette information est connue pour un ou plusieurs de ses représentants. D'autre part, comme le mécanisme d'action d'une enzyme est dicté par la structure du site actif et la position des différents groupes fonctionnels, le mécanisme et la stéréochimie de la réaction sont conservés dans une famille (exception faite de la famille GH 4). Toutes les glycoside hydrolases agissent généralement par un mécanisme qui implique deux résidus importants. Chez ces enzymes, les acides aminés catalytiques sont le plus souvent des aspartates et/ou des glutamates [97]. Les comparaisons de séquences par la méthode HCA permettent de déterminer les aspartates et

glutamates strictement conservés dans une même famille et donc de prédire les résidus catalytiques potentiels.

A l'heure actuelle, 97 familles GH différentes sont répertoriées et peuvent être consultées sur Internet^c [98]. Par ailleurs, beaucoup de ces familles, présentant des similarités dans le repliement tridimensionnel des enzymes qu'elles contiennent, ont été regroupées pour former des clans ou super-familles [99] (actuellement au nombre de 14). Réactualisée en permanence [100,101], la classification CAZy ("Carbohydre-Active enZYmes") est maintenant étendue à d'autres classes d'enzymes actives sur les sucres, par exemple les glycosyltransférases ou les polysaccharide lyases, ainsi qu'à leurs modules non-catalytiques.

1.2.1.2 Modes d'action

Le mécanisme d'hydrolyse enzymatique des liaisons glycosidiques le plus commun s'opère par une catalyse au cours de laquelle une paire d'acides carboxyliques intervient. Cependant, certaines glycoside hydrolases comme les myrosinases ou les chitinases et les *endo-N*-acétylglucosaminidases présentent un mécanisme moléculaire d'hydrolyse n'impliquant l'intervention que d'un seul acide aminé catalytique [102]. D'autres impliquent des mécanismes d'oxydoréduction et/ou d'élimination [103-105]. Ces mécanismes particuliers ne seront pas abordés dans ce chapitre.

Dans le cas du mécanisme le plus commun, un acide aminé joue le rôle du donneur de proton acide/base et l'autre agit en tant que base/nucléophile. Les glycoside hydrolases peuvent alors agir avec deux modes d'action distincts proposés en 1953 par Koshland [106].

Certaines enzymes agissent avec rétention de configuration et libèrent un produit dont la stéréochimie du carbone anomère est la même que celle du substrat (après deux inversions successives). Les autres, qui libèrent un oligosaccharide de configuration opposée, agissent avec inversion (par simple substitution). Le mécanisme de rétention conduit souvent à la transglycosylation avec le transfert du résidu glycosylé sur un accepteur différent d'une

^c Coutinho, P.M. & Henrissat, B. (1999) Carbohydrate-Active Enzymes server : <http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html>

molécule d'eau (alcool, unité glycosidique), alors que le mécanisme d'inversion conduit exclusivement à l'hydrolyse [107].

La différence de mécanisme entre les deux classes d'enzymes se traduit par des modifications structurales au sein du site actif et notamment sur la distance entre les acides aminés catalytiques. En effet, dans le cas d'une enzyme agissant avec rétention de configuration, ils sont distants en moyenne de 5.5 Å alors que la distance les séparant est de l'ordre de 10 Å dans une enzyme agissant avec inversion [108,109]. Cet écart plus important est nécessaire au mécanisme d'inversion puisqu'une molécule d'eau et le substrat doivent être présents simultanément au voisinage des deux acides aminés catalytiques.

Le mécanisme d'hydrolyse avec inversion de configuration, schématisé sur la **Figure 13** dans le cas des β -D-glycoside hydrolases, se déroule en une seule étape. L'atome d'oxygène glycosidique est protoné par le résidu acide/base et le départ de l'aglycone s'accompagne de l'attaque d'une molécule d'eau activée par la base. Cette simple substitution conduit donc à un produit de configuration opposée à celle du substrat initial.

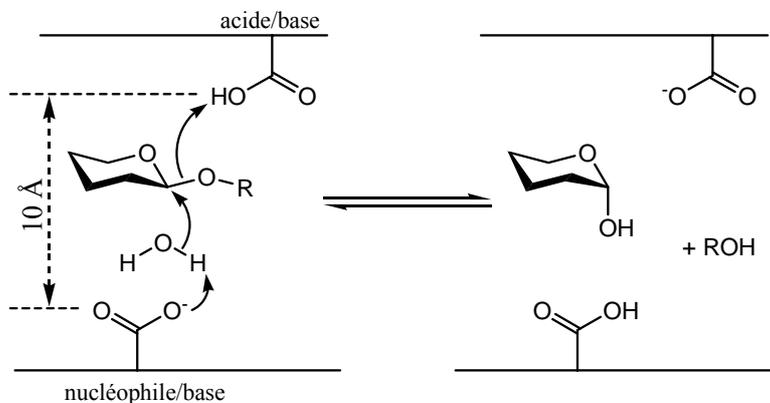


Figure 13 : Mécanisme d'une β -D-glycoside hydrolase agissant avec inversion de configuration.

Dans le mécanisme de rétention de configuration (**Figure 14**, cas des β -D-glycoside hydrolases), la première étape de l'hydrolyse est la protonation de l'atome d'oxygène glycosidique par le résidu acide/base et le départ de l'aglycone assisté par le nucléophile pour conduire à la formation d'un glycosyl-enzyme. Au cours d'une seconde étape, l'intermédiaire covalent est hydrolysé par une molécule d'eau conduisant à un produit de même stéréochimie que le substrat initial.

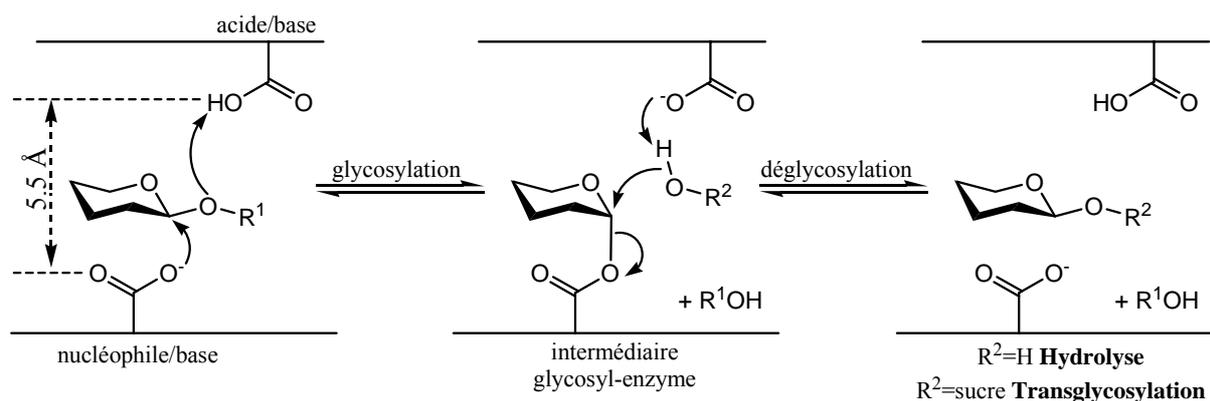


Figure 14 : Mécanisme d'une β -D-glycoside hydrolase agissant avec rétention de configuration.

Si un autre sucre ou une molécule hydroxylée se trouve dans le milieu, alors une réaction de transglycosylation entre en compétition avec la réaction d'hydrolyse, pouvant conduire à la formation d'un nouvel oligosaccharide. Cette réactivité peut être utilisée à des fins synthétiques dans la préparation d'oligosaccharides.

Sur la base de ce mécanisme, il s'agit de favoriser la réaction de transglycosylation par rapport à la réaction d'hydrolyse. Pour cela, le meilleur moyen est d'utiliser des sucres donneurs portant d'excellents groupes partants et ensuite de réduire la concentration effective en eau par l'addition de co-solvants. L'utilisation des glycoside hydrolases dans la synthèse d'oligosaccharides présente de nombreux avantages par rapport à celle des glycosyltransférases. Ces enzymes sont souvent bon marché et abondantes; ce sont typiquement des enzymes solubles qui peuvent être facilement obtenues par la technologie de l'ADN recombinant et beaucoup sont disponibles commercialement. Les donneurs de glycosyle sont facilement préparés. D'autre part, les glycoside hydrolases présentent souvent des spécificités de substrat (donneurs et accepteurs) plus larges que les glycosyltransférases.

Toutefois, le produit obtenu est aussi un substrat pour l'enzyme et peut donc être hydrolysé ou utilisé dans une nouvelle réaction de transglycosylation. Il est donc nécessaire d'optimiser les conditions expérimentales afin d'obtenir le meilleur rendement en produit désiré.

I.2.1.3 Topologie des sites actifs et distorsion de substrat

Malgré la diversité des repliements, la topologie des sites actifs des glycoside hydrolases appartient seulement à trois classes différentes (**Figure 15**), indépendamment du type de mécanisme [97].

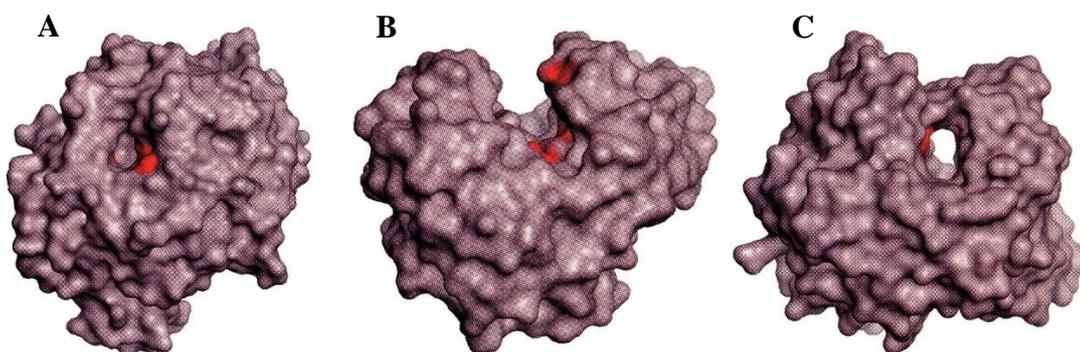


Figure 15 : Topologie du site actif des glycoside hydrolases [97].

A : Poche; B : Crevasse; C : Tunnel.

La topologie de type poche ou cratère est optimale pour la reconnaissance de l'extrémité non-réductrice ou réductrice d'un sucre. Il s'agit notamment de la topologie du centre actif commune aux *exo*-enzymes, c'est-à-dire des enzymes capables d'hydrolyser un glycoside ou un oligo- ou polysaccharide à partir d'un bout de chaîne (par exemple les glucoamylases).

La topologie de type crevasse ou sillon est une structure ouverte qui permet la fixation de plusieurs unités saccharidiques et est présente chez les *endo*-enzymes. Ces glycoside hydrolases sont capables d'hydrolyser une chaîne oligosaccharidique de façon aléatoire (par exemple les endoglucanases). Généralement, le site actif est constitué de 2 à 4 sous-sites^d de part et d'autre du site de catalyse.

Enfin la topologie de type tunnel résulte de la précédente, quand la protéine développe de longues boucles qui couvrent une partie de la crevasse. Elle est notamment rencontrée chez les enzymes qui hydrolysent les polysaccharides de façon processive (par exemple les cellobiohydrolases).

^d Par convention les sous-sites de reconnaissance du substrat sont numérotés à partir du site de coupure. Les sous-sites du côté non réducteur de la liaison à hydrolyser sont numérotés négativement et ceux du côté réducteur sont numérotés positivement [110].

L'étude et la meilleure compréhension du mode d'action des glycoside hydrolases ont été possibles grâce à l'utilisation conjointe d'inhibiteurs irréversibles ou réversibles et de techniques de cristallographie et de biologie moléculaire. Des inhibiteurs peuvent être des analogues de substrats non hydrolysables par les enzymes, de type C-oligosaccharides ou thio-oligosaccharides, ou des analogues de l'état de transition.

L'utilisation de thio-oligosaccharides a ainsi permis de mettre en évidence les changements conformationnels subis par l'unité saccharidique au niveau du sous-site -1 de certaines glycoside hydrolases. En 1996 [111], la structure aux rayons X du complexe entre la cellulase Cel7B de *Fusarium oxysporum* et un thio-cellopentaoside de méthyle (inhibiteur) a pour la première fois montré dans une cellulase la distorsion du substrat dans le sous-site -1 vers une conformation sofa (**Figure 16**), amenant l'aglycone en position semi-axiale et facilitant ainsi l'hydrolyse de la liaison glycosidique.

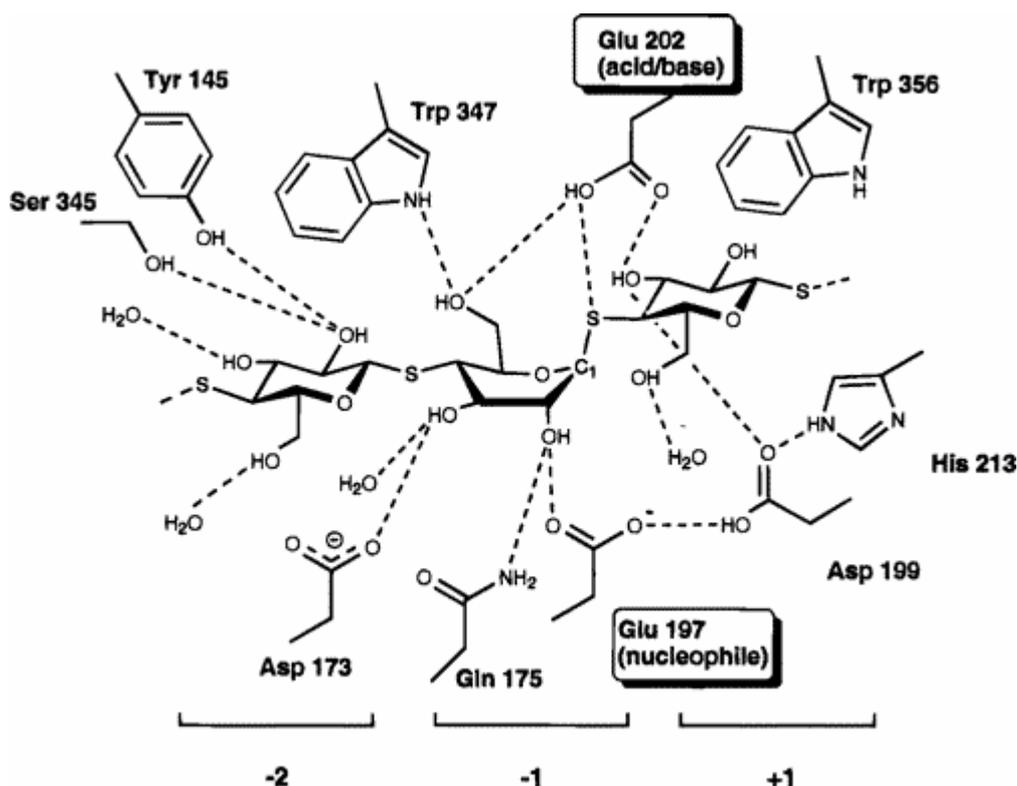


Figure 16 : Structure aux RX du complexe entre Cel7B de *Fusarium oxysporum* et le thio-cellopentaoside de méthyle (inhibiteur) [111].

Le même phénomène, observé avec d'autres enzymes, laisse supposer que ce changement conformationnel est caractéristique des β -glycoside hydrolases qui agissent avec rétention de configuration.

1.2.2 Le concept de glycosynthase

L'utilisation de glycoside hydrolases natives en milieu hydro-organique permet la synthèse d'oligosaccharides de degrés de polymérisation variables (DP 3, 4, ...). Cependant, la synthèse de dérivés de DP supérieurs est très difficile avec des glycoside hydrolases natives car les produits formés (qui ont une meilleure affinité pour l'enzyme que le substrat de départ) sont rapidement utilisés à leur tour comme substrats, conduisant à des mélanges complexes de produits. Cette méthode est donc limitée par les activités d'hydrolyse et de transglycosylation des enzymes, qui entraînent la dégradation des produits recherchés pendant la réaction.

Ce problème peut désormais être évité par l'utilisation de glycosynthases. En 1998, Withers et coll. [112,113] ont présenté ce concept (**Figure 17**), basé sur la modification du fonctionnement d'une enzyme par la mutation d'un des acides aminés catalytiques.

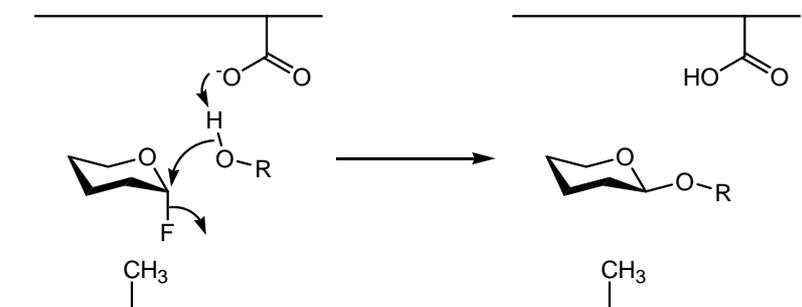


Figure 17 : Mécanisme d'action d'une glycosynthase.

Les glycosynthases sont ainsi des glycoside hydrolases dont l'acide aminé nucléophile est remplacé par un résidu non-nucléophile par mutagenèse dirigée. Ces enzymes conservent leur stéréosélectivité et leur capacité à effectuer des réactions de transglycosylation en présence de composés mimant l'intermédiaire glycosyl-enzyme. Le donneur de glycosyle, activé en position anomère par un bon groupe partant (généralement de type fluorure), est convenablement positionné dans le site actif pour réagir avec l'accepteur. L'acide aminé

responsable de la catalyse acide/base est conservé et permet d'activer l'hydroxyle de l'accepteur de glycoside. Privées de l'acide aminé nucléophile, ces enzymes sont incapables d'hydrolyser le substrat ainsi que le produit formé qui s'accumule dans le milieu.

1.2.3 Les différents exemples de glycosynthases

La première glycosynthase obtenue dérive de la β -glucosidase d'*Agrobacterium* sp. [113], enzyme possédant une forte activité de transglycosylation [114]. Withers et coll. ont transformé cette enzyme en une glycosynthase par mutation du résidu Glu-358, qui est le résidu catalytique nucléophile, en résidu Ala. Cette enzyme modifiée est capable dans un milieu tamponné (pour neutraliser l'acide fluorhydrique libéré) de condenser les fluorures donneurs d' α -D-glucosyle et d' α -D-galactosyle sur divers accepteurs avec d'excellents rendements.

Depuis, plusieurs autres glycosynthases ont été obtenues et sont utilisées pour la synthèse d'oligosaccharides. Les glycosynthases répertoriées à ce jour sont décrites dans le **Tableau 5** pour les *exo*-glycosynthases et dans le **Tableau 6** pour les *endo*-glycosynthases [115]. L'intérêt des *endo*-glycosynthases (provenant d'endoglycosidases) est qu'elles présentent une excellente régiosélectivité, contrairement aux *exo*-glycosynthases qui conduisent généralement à des mélanges de régioisomères.

Le pouvoir catalytique de la glycosynthase originale d'*Agrobacterium* a été amélioré grâce à des études de mutagenèse dirigée et une méthodologie générale a été développée pour permettre le criblage rapide de bibliothèques de mutants glycosynthases avec une meilleure activité [116]. Ainsi, le remplacement du nucléophile par une sérine plutôt qu'une alanine améliore l'efficacité du mutant d'*Agrobacterium* [117]. Pour certaines enzymes, le résidu nucléophile a été remplacé par une glycine [116].

Le concept des glycosynthases semblait, dans un premier temps, limité aux enzymes hydrolysant les liaisons glycosidiques de type β -D avec rétention de configuration. Mais en 2002, Okuyama et coll. [118] ont publié la première glycosynthase permettant la formation de liaisons de type α -D, obtenue après mutation du résidu nucléophile Asp-481 de l' α -glucosidase de *Schizosaccharomyces pombe*.

Les applications potentielles des glycosynthases n'ont pas encore été toutes explorées à ce jour. On peut citer l'utilisation du mutant glycosynthase de la β 1,3-glucanase d'*Hordeum vulgare* [119] qui a permis d'obtenir des β 1,3-glucanes linéaires avec un degré de polymérisation de 30-34 qui sont des agents thérapeutiques potentiels grâce notamment à leurs propriétés immunostimulantes.

Tolborg et coll. ont également utilisé le mutant glycosynthase d'*Agrobacterium* sp. pour la synthèse en phase solide d'oligosaccharides complexes ou de glycopeptides [120].

Organisme	Enzyme native	GH	Mutations	Liaisons formées	Donneurs ^{a,b} /Accepteurs ^a	Références
<i>Agrobacterium</i> sp.	β-glucosidase Abg	1	E358A E358S E358G	β1-4; β1-3	Donneurs : GlcαF; GalαF Accepteurs : mono- et disaccharides	[113]
<i>Sulfolobus solfataricus</i>	β-glycosidase Ss-β-Gly	1	E387G	β1-3; β1-6; β1-4	Donneurs : GlcαF; Glcβ- <i>o</i> NP; Galβ- <i>o</i> NP; Fucβ- <i>o</i> NP Accepteurs : mono- et disaccharides	[121-124]
<i>Cellulomonas fimi</i>	β-mannosidase Man2A	2	E519A E519S	β1-3; β1-4	Donneurs : ManαF Accepteurs : mono- et disaccharides	[125]
<i>Schizosaccharomyces pombe</i>	α-glucosidase SPG	31	D481G	α1-4; α1-6	Donneurs : GlcβF Accepteurs : monosaccharides	[118]
<i>Escherichia coli</i>	β-galactosidase LacZ	2	E537S (+G794D)	β1-6	Donneurs : GalαF Accepteurs : Glcβ- <i>p</i> NP; Glc(β1-4)Glcβ- <i>p</i> NP	[126]
<i>Thermosphaera aggregans</i>	β-glycosidase Ta-β-Gly	1	E386G	β1-3; β1-4; β1-6	Donneurs : Glcβ- <i>o</i> NP; Galβ- <i>o</i> NP Accepteurs : mono- et disaccharides	[123,124]
<i>Pyrococcus furiosus</i>	β-glycosidase CelB	1	E372A	β1-3; β1-4; β1-6	Donneurs : Glcβ- <i>o</i> NP Accepteurs : mono- et disaccharides	[123]
<i>Streptomyces</i> sp.	β-glucosidase	1	E383A	β1-3; β1-4	Donneurs : GlcαF; GalαF Accepteurs : mono- et disaccharides	Planas (résultats non publiés)
<i>Thermus thermophilus</i>	β-glycosidase Tt-β-Gly	1	E338A E338S E338G	β1-3	Donneurs : GlcαF; GalαF Accepteurs variés	Tellier (résultats non publiés)

Tableau 5 : Exo-glycosynthases connues.

^a Abréviations : GlcαF, fluorure d'α-D-glucopyranosyle; GlcβF, fluorure de β-D-glucopyranosyle; Glcβ-*o*NP, β-D-glucopyranoside d'*ortho*-nitrophényle; Glcβ-*p*NP, β-D-glucopyranoside de *para*-nitrophényle; GalαF, fluorure d'α-D-galactopyranosyle; Galβ-*o*NP, β-D-galactopyranoside d'*ortho*-nitrophényle; ManαF, fluorure d'α-D-mannopyranosyle; Glc(β1-4)Glcβ-*p*NP, β-D-glucopyranosyl-(1-4)-β-D-glucopyranoside de *para*-nitrophényle; Fucβ-*o*NP, β-D-fucopyranoside d'*ortho*-nitrophényle.

^b Présence de formate de sodium (nucléophile externe) pour les donneurs non fluorés.

Organisme	Enzyme native	GH	Mutations	Liaisons formées	Donneurs ^a /Accepteurs	Références
<i>Bacillus licheniformis</i>	β1,3-1,4-glucanase (lichenase)	16	E134A	β1-4	Donneurs : Glc(β1-3)GlcαF Accepteurs : mono- et disaccharides	[127-129]
<i>Humicola insolens</i>	β1,4-glucanase Cel7B (cellulase)	7	E197A	β1-4	Donneurs : Glc(β1-4)GlcαF; Gal(β1-4)GlcαF Accepteurs : mono-, di- et oligosaccharides	[130]
<i>Hordeum vulgare</i>	β1,3-glucanase (laminarinase)	17	E231G	β1-3	Donneurs : Glc(β1-3)GlcαF; Gal(β1-3)GlcαF Accepteurs : mono- et disaccharides	[119,131]
<i>Cellvibrio japonicus</i>	β-mannanase Man26A	26	E320G	β1-4	Donneurs : Man(β1-4)ManαF Accepteurs : mono- et disaccharides	[132]
<i>Pyrococcus furiosus</i>	β1,3-glucanase LamA (laminarinase)	16	E170A	β1-3 β1-4	Donneurs : Glc(β1-3)GlcαF Accepteurs : mono- et disaccharides	Planas, van der Oost (résultats non publiés)

Tableau 6 : Endo-glycosynthèses connues.

^a Abréviations : Glc(β1-3)GlcαF, fluorure de β-D-glucopyranosyl-(1-3)-α-D-glucopyranosyle; Glc(β1-4)GlcαF, fluorure de β-D-glucopyranosyl-(1-4)-α-D-glucopyranosyle; Gal(β1-4)GlcαF, fluorure de β-D-galactopyranosyl-(1-4)-α-D-glucopyranosyle; Gal(β1-3)GlcαF, fluorure de β-D-galactopyranosyl-(1-3)-α-D-glucopyranosyle; Man(β1-4)ManαF, fluorure de β-D-mannopyranosyl-(1-4)-α-D-mannopyranosyle.

Les glycosynthases conservent la stéréospécificité des glycoside hydrolases dont elles sont issues, sont régiosélectives ou régiospécifiques, et sont capables d'utiliser un ensemble varié de substrats, aussi bien donneurs qu'accepteurs. A l'heure actuelle, il existe un intérêt croissant pour ce concept du point de vue de la découverte de nouvelles enzymes possédant une forte activité de transglycosylation et pouvant être converties en glycosynthases, mais également concernant la modification de leurs spécificités de substrat ou du type de liaison formée. Jahn et coll. ont ainsi développé les thioglycoligases [133,134] et les thioglycosynthases [135] (**Figure 18**), enzymes capables de former des liaisons thioglycosidiques, après mutation d'un ou des deux résidus catalytiques.

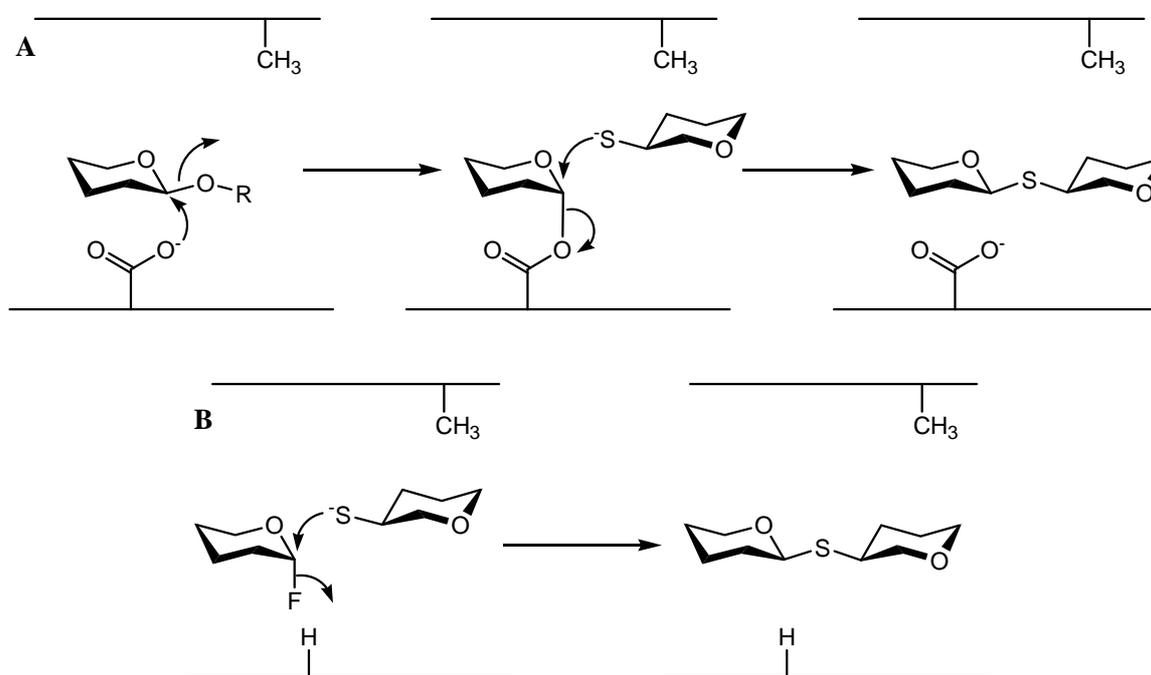


Figure 18 : Mécanisme d'une thioglycoligase (A) et d'une thioglycosynthase (B).

R = dinitrophényle.

Dernièrement, Stick et coll. [136] ont réussi à modifier la régiosélectivité du mutant glycosynthase de la β -glucosidase d'*Agrobacterium* en utilisant des accepteurs de glycoside porteurs de groupements de type éthers d'aryle ou esters d'aryle. Ces derniers modifient le positionnement de l'accepteur dans les sous-sites de l'enzyme, le substrat présentant alors à l'acide aminé catalytique acide/base un autre groupe hydroxyle que l'hydroxyle habituel.

1.2.4 La glycosynthase Cel7B E197A d'*Humicola insolens*

Le champignon *Humicola insolens* produit une grande variété de cellulases et d'hémicellulases [137]. Il sécrète au moins 2 cellobiohydrolases, 5 endoglucanases et une cellobiose déshydrogénase, actives à pH alcalin. Les 7 cellulases ont été clonées et exprimées chez un hôte non cellulolytique, *Aspergillus oryzae*, et leurs domaines catalytiques respectifs sont classés dans 5 familles différentes.

L'endo- β 1,4-glucanase Cel7B native (auparavant EG I) appartient à la famille 7 des glycoside hydrolases (clan GH-B). Elle possède une forte similarité de séquence avec l'endoglucanase Cel7B de *Trichoderma reesei* (GH 7) [138], malgré l'absence de domaine de fixation à la cellulose (CBD). Sa grande similarité de séquence (57%) avec l'endoglucanase Cel7B de *Fusarium oxysporum* (GH 7) [139] laisse supposer des interactions identiques entre le substrat et le site actif de ces deux enzymes.

La famille GH 7 regroupe des enzymes de spécificité différente et contient à la fois des enzymes identifiées comme des cellobiohydrolases et des endoglucanases. Cependant, les enzymes de la famille GH 7 ont en commun les caractéristiques suivantes :

- elles agissent avec rétention de configuration sur des liaisons Glc(β 1-4)Glc;
- les acides aminés catalytiques sont :
 - un résidu Glu pour l'acide aminé nucléophile;
 - un résidu Glu comme résidu catalytique acido-basique.
- elles possèdent un repliement caractéristique " β -jelly roll";

Les premiers résultats concernant la cristallisation et l'analyse aux rayons X de Cel7B d'*H. insolens* ont été publiés en 1992 [140] et sa structure tridimensionnelle est maintenant connue [141]. La chaîne polypeptidique présente un site actif en forme de crevasse qui lui permet de fixer aléatoirement la chaîne de cellulose et de dégrader efficacement les celluloses substituées comme la CMC qu'elle hydrolyse à 33%, alors qu'elle ne possède aucune activité sur la cellulose cristalline. Des études cinétiques ont également montré que le site actif de Cel7B contient 4 sous-sites distincts pour la fixation du substrat [142].

L'enzyme native possède une forte activité de transglycosylation. En effet, aucune hydrolyse n'est observée lorsqu'elle est incubée avec une solution concentrée de cellotétritol

(cellotétraose réduit). Par contre des oligomères insolubles de degrés de polymérisation plus élevés sont rapidement formés [142].

Fort et coll. ont publié en 2000 [130] l'étude de la glycosynthase Cel7B E197A d'*Humicola insolens* utilisée pour la synthèse de liaisons de type β 1-4 entre des unités D-glycosyles. En 2003, Ducros et coll. [143] ont rapporté l'étude des caractéristiques structurales et cinétiques de deux mutants glycosynthase de cette endoglucanase, où le résidu nucléophile avait été muté en alanine ou en sérine.

Une étude complète concernant la spécificité de substrat de cette enzyme a été effectuée [130]. Elle a été réalisée en utilisant le fluorure d' α -lactosyle comme donneur en présence de différents accepteurs mono- et disaccharidiques (**Tableau 7**). Il a été montré que les hydroxyles 2 et 6 du glycoside situé dans le sous-site +1 ne sont pas impliqués dans des interactions prépondérantes (Entrées **1**, **2** et **3**), confirmant les résultats des structures aux rayons X obtenues.

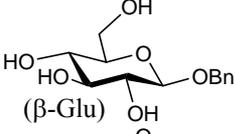
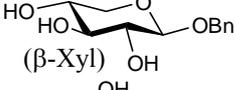
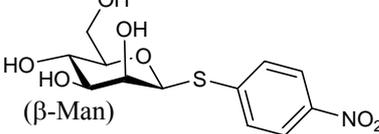
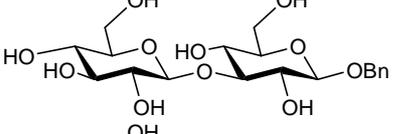
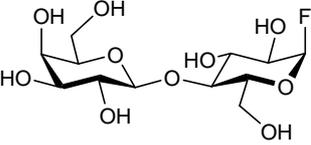
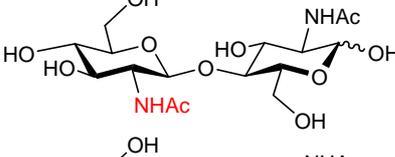
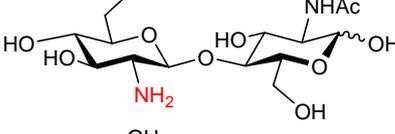
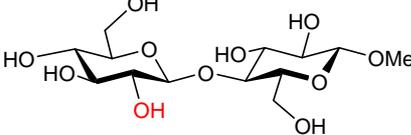
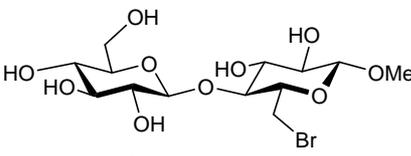
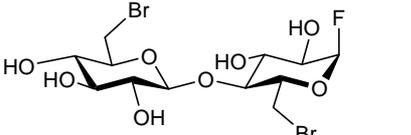
Entrée	Donneur	Accepteurs	Rendement
1			83% ^a
2			51% ^a
3			61% ^a
4			100% ^b
5			0%
6			65% ^a
7			100% ^b
8			100% ^b
9			80% ^a

Tableau 7 : Exploration de la spécificité de la glycosynthase dérivée de Cel7B d'*H. insolens* au niveau de l'accepteur [130].

^a Le produit de couplage est isolé sous forme acétylée; ^b Le produit de couplage est isolé sous forme libre.

Alors que l'autocondensation du fluorure d' α -lactosyle sur lui-même n'est pas possible, ce donneur possédant un hydroxyle axial sur le C-4 du sucre non-réducteur, le fluorure d' α -cellobiosyle peut être à la fois donneur et accepteur. Ce dernier, en présence de la glycosynthase, est polymérisé en cellulose de type II. D'autres donneurs cellobiosidiques substitués en 6^{II} (**Schéma 1**) ont permis l'obtention de celluloses modifiées ou branchées de faible poids moléculaire.

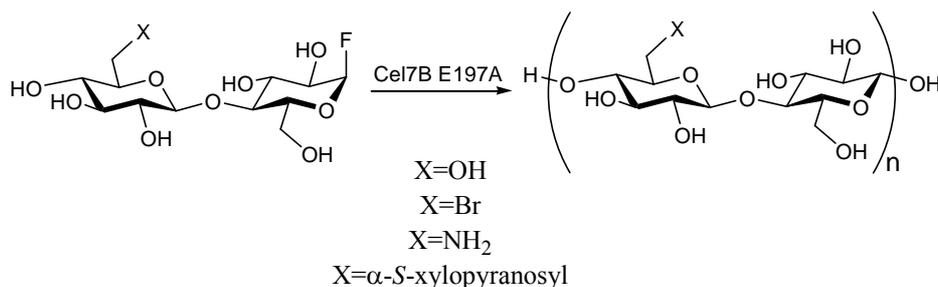


Schéma 1 : Polymérisation enzymatique de fluorures de cellobiosyle substitués catalysée par la glycosynthase Cel7B E197A d'*H. insolens* [130].

La possibilité d'une substitution en 6^{II} confirme les résultats cristallographiques qui avaient permis de prédire une absence de gêne stérique au niveau de l'hydroxyle sur le C-6 de l'extrémité non-réductrice. Par la suite, Fort et coll. ont développé une méthode pour contrôler l'autocondensation du fluorure d' α -cellobiosyle [144]. L'introduction d'un groupement protecteur judicieusement choisi (groupement tétrahydropyranyle THP), toléré par l'enzyme, facilement déprotégé et capable de supporter les conditions de préparation du fluorure (HF, pyridine), a permis d'obtenir des oligomères de cellulose de DP contrôlé.

Le spectre étendu de produits obtenus par l'utilisation de cette glycosynthase démontre bien l'intérêt d'une telle enzyme pour la synthèse. Cette glycosynthase est donc capable de condenser divers substrats, donneurs et/ou accepteurs, modifiés par rapport à son substrat d'origine qu'est la cellulose.

Les équipes d'H. Driguez et d'A. Planas ont utilisé de concert les glycosynthases dérivées de Cel7B d'*Humicola insolens* et de la β 1,3-1,4-glycosyltransférase de *Bacillus licheniformis* dans le but de synthétiser "one pot" un hexasaccharide comportant des liaisons mixtes β 1-3 et β 1-4 avec un excellent rendement [145]. Boyer et coll. [146] ont également utilisé Cel7B E197A d'*H. insolens* pour la synthèse d'un hexasaccharide bifonctionnalisés par des groupements fluorescents permettant le dosage d'activités enzymatiques de cellulases de différentes familles. Dernièrement, les potentialités de la glycosynthase Cel7B E197A ont été démontrées pour la synthèse d'oligosaccharides branchés complexes de type xylogluco-oligosaccharidique^e.

^e Thèse Régis FAURE (2004) "Synthèses chimio-enzymatiques de xylogluco-oligosaccharides pour l'étude des enzymes intervenant dans le métabolisme des xyloglucanes".

1.3 L'ingénierie des protéines

Le terme "ingénierie des protéines" regroupe la conception et la construction de protéines possédant des propriétés améliorées ou même totalement nouvelles. Les cibles visées par l'ingénierie peuvent être la protéine en elle-même ou le gène codant pour la protéine étudiée.

Parmi toutes les protéines existantes, les enzymes sont de fascinants biocatalyseurs, par leur capacité à accélérer des réactions chimiques avec une haute spécificité (le plus souvent de façon régio- et stéréosélective) dans des conditions douces. Grâce aux progrès récents dans les biotechnologies, l'utilisation de ces biocatalyseurs est en plein développement. Dans l'industrie notamment, ils permettent de développer des procédés moins polluants, d'effectuer les réactions dans des conditions plus douces et sont énantiosélectifs. Toutefois, les enzymes ne sont pas adaptées pour fonctionner dans des procédés industriels classiques de synthèse organique, mais dans un environnement biologique, c'est-à-dire dans une gamme de pH, température et salinité bien définie, et généralement dans l'eau comme unique solvant. De plus, dans leur milieu naturel, elles fonctionnent dans un système impliquant des relations complexes entre les enzymes, mais aussi avec les substrats ou les produits des réactions. En fait, peu d'enzymes correspondent exactement aux besoins du chercheur ou de l'industriel, qui souhaite souvent modifier des substrats non-naturels avec l'enzyme, dans des conditions où elle serait instable voire inactive.

Une première solution consiste à explorer des environnements extrêmes pour isoler des enzymes extrémophiles. Toutefois, aucun environnement naturel ne permet d'isoler à coup sûr des enzymes capables de fonctionner par exemple dans le diméthylformamide ou de catalyser la transformation énantiosélective d'un substrat qui n'existe pas à l'état naturel. D'où l'idée de faire évoluer les enzymes en laboratoire, de façon à mieux répondre au cahier des charges des procédés industriels [147-150].

I.3.1 L'évolution des protéines : pourquoi ?

Dans certains cas, il est nécessaire de modifier une enzyme afin d'améliorer ses performances. De telles modifications peuvent être introduites avec des objectifs divers (**Tableau 8**): augmenter sa stabilité et sa solubilité, accroître sa résistance à de hauts pH ou à des agents agressifs tels que l'eau de Javel ou encore améliorer son activité vis-à-vis d'un substrat d'intérêt.

Les premiers essais de modifications des protéines étaient basés sur des approches chimiques ou enzymatiques, mais plus tard l'introduction des outils d'ingénierie génétique a permis de réaliser des modifications au niveau du gène, rendant possible des changements précis tels que les substitutions, délétions ou insertions/extensions d'acides aminés seuls ou de larges séquences.

Propriétés modifiées
- Taux d'expression dans un hôte
- Stabilité thermique
- Adaptation à différentes températures, différents pH
- Performances dans un environnement non naturel ou extrême (stabilité et activité)
Exemples : dans les solvants organiques; exigences vis-à-vis des ions métalliques, de co-facteurs
- Activité vis-à-vis de substrats naturels ou non-naturels : Elimination de fonctions enzymatiques non-désirées Spécificité de substrat
- Enantiosélectivité
- Mécanisme de réaction

Tableau 8 : Exemples de propriétés des enzymes pouvant être améliorées ou modifiées.

Une des caractéristiques fondamentales des enzymes est leur spécificité vis-à-vis de leur substrat. Modifier cette spécificité, en augmentant par exemple la variété de molécules dont une enzyme peut catalyser la transformation, revêt une importance particulière [**151-154**].

I.3.2 L'évolution des protéines : comment ?

I.3.2.1 Les différentes modifications

La **Figure 19** regroupe différents exemples d'ingénierie des protéines à la disposition du chercheur.

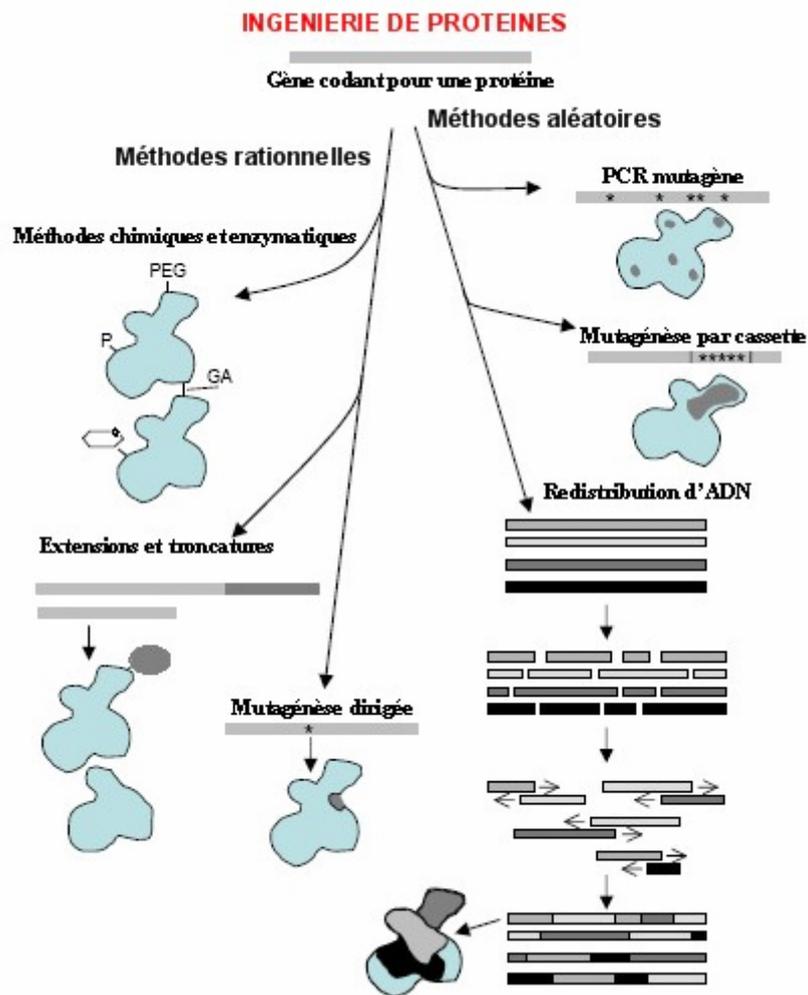


Figure 19 : Exemples de principes d'ingénierie des protéines disponibles à l'heure actuelle^a.

^a Abréviations : GA, glutaraldéhyde; P, phosphate; PEG, polyéthylène glycol.

On distingue les modifications par des voies chimiques ou enzymatiques, et les modifications effectuées par génie génétique.

I.3.2.1.1 Modifications chimiques et enzymatiques

Un premier exemple de modification d'une protéine, très simple et de nature non-covalente, a été l'addition d'ions zinc dans des préparations d'insuline, qui ont permis d'augmenter la stabilité de la protéine pendant le stockage et après son injection *in vivo*.

Des modifications chimiques covalentes ont également été utilisées pour altérer les propriétés de protéines par des méthodes spécifiques ou non [155]. Par exemple, l'utilisation du glutaraldéhyde, un agent chimique de liaison bifonctionnel qui réagit avec les fonctions amines primaires, a permis en 1964 de stabiliser la carboxypeptidase, facilitant ainsi sa détermination structurale par cristallographie des rayons X [156] mais en perdant presque toute son activité après le traitement.

Un autre agent utilisé pour la modification chimique des protéines est le polymère amphiphile polyéthylène glycol (PEG). Le PEG a notamment été utilisé pour augmenter la solubilité des protéines dans des solvants organiques, pour réduire leur antigénicité ou pour prolonger la demi-vie de certaines protéines dans le sérum [157,158].

Des méthodes enzymatiques peuvent également être utilisées pour modifier les protéines, en ciblant des sites naturellement présents ou introduits par ingénierie génétique. Les protéases peuvent ainsi être utilisées pour digérer les protéines au niveau de sites spécifiques. On peut citer ici l'exemple de la trypsine, utilisée en combinaison avec la carboxypeptidase pour la digestion *in vitro* de la pro-insuline en insuline. Les protéases telles que la papaïne et la pepsine ont été utilisées pendant longtemps pour digérer les anticorps en sous-fragments. D'autres exemples de modifications enzymatiques peuvent permettre la biotinylation des protéines [159] ou la phosphorylation/déphosphorylation, en utilisant les kinases et les phosphatases, respectivement [160,161].

I.3.2.1.2 Modifications par génie génétique

Une percée majeure dans le domaine de la modification de protéines a été le développement de méthodes de mutagenèse dirigée au niveau du gène codant pour la protéine, en utilisant des oligonucléotides synthétiques [162]. La première modification d'une enzyme, une tyrosyl ARNt synthétase, a été réalisée en 1982 [163]. Une cystéine a été

remplacée par une sérine par mutagenèse dirigée, altérant ainsi les caractéristiques de liaison substrat-protéine. La PCR ("Polymerase Chain Reaction" ou amplification en chaîne par polymérase) a permis aux chercheurs d'amplifier de petites quantités d'ADN [164]. Des variations de la méthodologie décrite à l'origine ont eu un grand impact comme outils pour l'introduction de mutations spécifiques ou aléatoires dans la séquence d'ADN [165].

Ces outils génétiques sont utilisés en routine aujourd'hui pour la conception et la production de mutants des protéines par ingénierie. L'arsenal des outils disponibles est en perpétuelle augmentation, avec la mise au point de méthodes de synthèse de peptides et de petites protéines artificielles [166], l'introduction d'acides aminés non-naturels [167,168] (qui peuvent ensuite subir des modifications chimiques) et la création de mutations distribuées aléatoirement le long de la séquence du gène.

L'ingénierie des protéines peut impliquer l'une et/ou l'autre des approches suivantes. Dans la première, les informations structurales et mécanistiques sont utilisées pour identifier les changements nécessaires à l'obtention d'une protéine possédant les propriétés souhaitées. Dans une seconde approche, l'emploi de méthodes aléatoires permet d'effectuer des recherches de nouvelles fonctions sans avoir une connaissance très précise des modifications requises. Dans certains cas, l'approche aléatoire peut être combinée avec la conception rationnelle pour réduire la taille de la séquence d'ADN sur laquelle l'évolution est effectuée.

1.3.2.2 La conception rationnelle ("rational design") de mutants de protéines

Une première approche d'ingénierie des protéines est de concevoir, d'une manière rationnelle des mutants, avec de nouvelles propriétés, en réalisant des substitutions, des insertions et/ou des délétions d'acides aminés seuls ou de séquences plus importantes, voire de domaines entiers.

Des connaissances sur la protéine étudiée sont nécessaires pour déterminer les modifications nécessaires à effectuer. La croissance rapide du nombre de structures tridimensionnelles résolues par cristallographie des rayons X ou par RMN [169], la protéine étant seule ou en complexe avec d'autres molécules, a contribué énormément à la compréhension des relations structure/fonction des protéines. Les informations structurales

peuvent permettre d'identifier des résidus ou régions importants impliqués dans la catalyse, les liaisons au substrat ou au ligand, et de faire des prédictions pour modifier la protéine dans le but, par exemple, d'augmenter sa stabilité, sa solubilité et/ou altérer d'autres propriétés, notamment sa spécificité de substrat.

Dans le cas où des données structurales ne sont pas disponibles, la méthode appelée "scanning mutagenesis" peut être une option. Les zones qui peuvent être importantes sont analysées par substitution d'un acide aminé à la fois, suivie par une analyse fonctionnelle des mutants exprimés pour explorer l'effet de la substitution. L'alanine est un des acides aminés le plus souvent utilisé comme substitut; on parle "d'alanine scanning". En effet, l'alanine est considérée comme un bon résidu de substitution car il ne modifie pas l'orientation de la chaîne peptidique, contrairement à la glycine ou la proline, et ne possède pas de caractéristiques stériques ou électrostatiques particulières [170]. Cette méthode peut également être utilisée pour analyser la contribution de chaque acide aminé potentiel impliqué dans une fonction spécifique.

Différentes approches de conception rationnelle sont disponibles à l'heure actuelle :

- la mutagenèse dirigée : une ou plusieurs substitutions, identifiées par des méthodes très diverses (modélisation moléculaire par exemple), sont réalisées;
- les extensions ou les troncatures : cela comprend la construction de protéines de fusion ou de mutants possédant des domaines délétés;
- la conception *de novo* de protéines.

Les altérations de la séquence peptidique sont principalement ciblées sur des positions au niveau ou proches des régions directement reliées à la fonction étudiée, tel que le site actif des enzymes. Toutefois, des effets dramatiques sur la fonctionnalité peuvent aussi avoir lieu grâce à des modifications au niveau de régions éloignées de la partie active de la protéine. Mais de telles substitutions seraient très difficiles à prédire par une simple approche rationnelle. C'est à ce niveau qu'intervient l'évolution dirigée des protéines.

I.3.2.3 L'évolution dirigée des protéines

L'évolution dirigée réunit un ensemble de méthodes qui tentent de mimer au laboratoire les processus de l'évolution naturelle, dans un laps de temps très court. Les développements récents de la biologie moléculaire permettent de reproduire efficacement dans un tube à essai les deux moteurs de l'évolution que sont la mutation et la recombinaison. Combinées à des stratégies de criblage de plus en plus puissantes et sophistiquées, ces méthodes aboutissent en quelques cycles à faire évoluer une protéine dans la direction définie par l'expérimentateur.

Les premiers champs d'application de ces méthodes concernent les biocatalyseurs et les anticorps, domaines où la diversité naturelle peut sembler immense, mais pour lesquels l'évolution naturelle a sélectionné des fonctions selon une logique biologique, qui ne répond pas forcément aux critères des biotechnologies.

Pour réaliser une telle évolution dirigée des protéines, il faut donc :

- une stratégie efficace d'évolution;
- un système d'expression dans lequel l'enzyme modifiée sera exprimée sous forme fonctionnelle; il peut s'agir de l'organisme d'origine, d'un organisme recombinant ou même d'un système *in vitro*; en général les bactéries ou les levures sont utilisées;
- un crible ou une méthode de sélection adaptés à la propriété recherchée.

Les sources de diversité à la disposition du chercheur sont nombreuses. L'étape commune à toutes les stratégies employées est la création d'une banque de gènes à partir desquels seront produites les protéines. Selon le projet et la classe de protéines étudiés, de telles banques peuvent être générées à partir de diverses sources naturelles et/ou la variabilité génétique peut être générée par l'homme *in vivo* ou *in vitro* par des techniques d'ingénierie génétique très sophistiquées. Parmi les méthodes de mutagenèse aléatoire, on peut citer la mutagenèse chimique, l'irradiation UV, la recombinaison homologue, la mutagenèse aléatoire par cassette. L'utilisation de souches mutatrices inactivées au niveau de leur système de réparation de l'ADN permet d'obtenir des plasmides comportant des mutations. La PCR mutagène, ou sujette à erreurs ("error prone PCR"), repose sur l'infidélité d'ADN polymérase (notamment la polymérase Taq); les erreurs sont introduites pendant l'amplification de l'ADN et peuvent varier en nombre selon la concentration en ions magnésium ou manganèse dans le

milieu ou par abaissement de la concentration d'un nucléotide. Le réarrangement d'ADN ou redistribution de l'ADN ("DNA shuffling"), développé en 1994 par Stemmer et coll. [171], consiste à fragmenter de multiples copies d'un gène (avec plus ou moins de mutations introduites) puis à les réassembler aléatoirement.

Dans le **Tableau 9**, sont présentés des exemples d'enzymes dont une propriété a été modifiée par évolution dirigée.

Enzyme	Propriété Cible	Approche	Références
kanamycine nucléotidyltransférase	Thermostabilité	souche mutatrice + sélection	[172]
subtilisine E	Activité dans les solvants organiques	PCR mutagène + criblage	[173]
β -lactamase	Activité vis-à-vis de nouveaux substrats	redistribution d'ADN + sélection	[174]
subtilisine BPN'	Stabilité en l'absence d'ions calcium	troncature (boucle) + cassette + criblage	[175]
β -galactosidase	Activité vis-à-vis de nouveaux substrats	redistribution d'ADN + criblage	[176]
lipase	Enantiosélectivité	PCR mutagène + criblage	[177]
subtilisine E	Thermostabilité	PCR mutagène + redistribution d'ADN + criblage	[178]
amylosaccharase	Activité vis-à-vis du saccharose (hydrolyse ou polymérisation)	PCR mutagène + redistribution d'ADN + sélection et criblage	[179]
subtilisine BPN'	Activité à 10°C	mutagénèse chimique + criblage	[180]
thymidine kinase	Spécificité de substrat	redistribution d'ADN + criblage	[181]

Tableau 9 : Quelques exemples d'enzymes modifiées obtenues par évolution dirigée.

Ces méthodes de mutagenèse aléatoire peuvent générer un très grand nombre de mutants qui doivent ensuite être étudiés pour la nouvelle propriété recherchée. L'identification des mutants avec la propriété cible peut être effectuée selon deux approches : le criblage (examen de chaque mutant séparément pour une propriété précise) et/ou la sélection (seuls les mutants possédant la nouvelle propriété survivent). Pour cela, de nombreuses méthodes d'analyse sont disponibles pour tester les fonctions des protéines produites, souvent adaptées à chaque protéine. Ensuite, l'étape suivante est la détermination des mutations introduites dans la séquence protéique.

L'amplification et le séquençage de l'ADN étant faciles à mettre en œuvre, de nouvelles méthodes ont été développées afin de lier physiquement chaque protéine produite à sa séquence d'ADN correspondante, autrement dit des techniques où le phénotype et le génotype sont liés [182]. On peut citer par exemple des techniques telles que le "phage display" (présentation de phage) ou le "cell surface display" (présentation à la surface des cellules). Cela permet ainsi d'obtenir rapidement la séquence d'acides aminés du mutant de la protéine étudiée possédant la propriété cible.

1.3.3 La modification de la spécificité de substrat des enzymes

L'ingénierie des protéines a été d'une grande aide dans l'élucidation des relations structure/activité par la modification de la spécificité de substrat des enzymes. Il s'agit de la préférence qu'une enzyme manifeste pour une molécule par rapport à d'autres.

Dans le cas de l'approche rationnelle, une compréhension détaillée du mécanisme catalytique et une connaissance de la structure et des déterminants de la spécificité de substrat d'une enzyme particulière sont utilisées comme base pour altérer sa spécificité d'une manière prévisible.

Plusieurs exemples d'ingénierie des protéines dans le but de modifier la spécificité de substrat d'enzymes actives sur les sucres peuvent être cités. La plupart sont des améliorations de l'activité de l'enzyme vis-à-vis de substrats pauvrement acceptés par les enzymes natives.

Après avoir modélisé le substrat naturel dans le site actif de l'enzyme, Deacon et coll. [183] ont ainsi augmenté l'activité d'une galactose oxydase vis-à-vis du D-fructose, en mutant un seul acide aminé, impliqué dans des interactions de type liaisons hydrogènes avec le groupe hydroxyle en 4 du substrat. D'autre part, les travaux de Sun et coll. [184] ont permis de

modifier une galactose oxydase en introduisant une activité supplémentaire vis-à-vis du D-glucose par mutagenèse au niveau d'acides aminés remarquables lors d'études de modélisation. Thorson et coll. [185] ont réussi à modifier une nucléotidyltransférase par une approche rationnelle pour augmenter le répertoire de sucres phosphatés en position anomère dont elle pouvait catalyser la conversion en nucléotide-sucre.

Des exemples d'ingénierie des glycoside hydrolases peuvent également être cités. Andrews et coll. [186] ont obtenu après mutagenèse dirigée un double mutant de la xylanase Xyl10A de *Pseudomonas cellulosa* avec une plus grande activité d'hydrolyse sur des substrats dérivés du glucose que sur des substrats équivalents dérivés du xylose. En 2000, Kaper et coll. [187] ont augmenté l'activité hydrolytique d'une β -glucosidase de *Pyrococcus furiosus* vis-à-vis d'un dérivé β -D-galactose phosphorylé en 6. De même, Corbett et coll. [188] ont modifié par mutagenèse dirigée la spécificité de substrat d'une β -glycosidase de *Sulfolobus solfataricus*.

Lorsque la conception rationnelle ne donne pas les résultats attendus ou n'est tout simplement pas envisageable, la modification de la spécificité de substrat peut être réalisée par évolution dirigée. La très grande majorité des travaux de mutagenèse aléatoire ont pour conséquence une modification des résidus constituant le site de liaison au substrat des enzymes. Toutefois, certains travaux ont démontré que des mutations en d'autres sites permettent aussi d'obtenir des enzymes dont la gamme de substrats est élargie. Il semble donc nécessaire de ne pas limiter la modification des enzymes au niveau de leur site actif.

Faire évoluer une enzyme vers une nouvelle fonction reste toutefois une aventure risquée. En effet, l'impact des modifications sur l'activité de l'enzyme et sur son repliement est rarement prévisible, et on ne peut que rarement prédire l'étendue de l'exploration des séquences à réaliser pour créer la nouvelle fonction et également la fréquence des solutions. L'évolution dirigée est envisageable lorsqu'une nouvelle fonction, par exemple une activité vis-à-vis d'un substrat qui n'est pas du tout accepté par l'enzyme native, peut être obtenue en réalisant une ou deux mutations d'acides aminés. Toutefois, si la nouvelle fonction nécessite le remplacement simultané de plusieurs résidus d'acides aminés proches dans la séquence peptidique, alors il ne paraît pas vraisemblable d'obtenir cela dans une bibliothèque de mutants obtenus aléatoirement.

Dans la famille des glycoside hydrolases, Zhang et coll. [176] ont utilisé le "DNA shuffling" pour modifier la spécificité de la β -galactosidase d'*Escherichia coli* en vue d'une

meilleure activité vis-à-vis d'un dérivé du fucose. En 2004, Kim et coll. [189] ont publié leurs travaux d'évolution dirigée sur la glycosynthase issue de la β -galactosidase d'*Agrobacterium* sp., qui ont permis d'augmenter l'activité catalytique de l'enzyme et d'étendre le répertoire des substrats acceptés comme par exemple le D-xylose.

Finalement, une combinaison des approches rationnelle et aléatoire semble une excellente solution. En effet, la fréquence avec laquelle des mutations bénéfiques ont lieu au niveau ou en dehors du site actif et le succès relatif de la conception rationnelle d'une nouvelle spécificité de substrat laissent penser que la mutagenèse aléatoire limitée à certaines régions de l'enzyme (sites de liaison du substrat ou du co-facteur par exemple) serait susceptible de produire des mutants intéressants.

OBJECTIFS

OBJECTIFS

Nous avons décrit précédemment (cf. paragraphe **I.2.4**) le potentiel de la glycosynthase Cel7B E197A d'*Humicola insolens* pour la synthèse d'oligosaccharides complexes. Pour la préparation d'autres biomolécules tels que les glycoconjugués, la glycosynthase Cel7B E197A pourrait constituer un nouvel outil dans le contrôle de la glycosylation des protéines recombinantes.

Une application de ces travaux de thèse serait de contribuer au développement d'une méthode alternative aux méthodes de remodelage de la glycosylation des protéines déjà existantes (cf. paragraphe **I.1.5**). En effet, le contrôle de la glycosylation des protéines recombinantes, notamment d'intérêt thérapeutique, apparaît comme un moyen d'augmenter le temps de demi-vie de celles-ci dans l'organisme et donc leur efficacité en tant que médicaments.

La stratégie envisagée pour le remodelage de la glycosylation des protéines recombinantes s'intéresse plus particulièrement aux *N*-glycoprotéines produites dans les plantes transgéniques, ce système d'expression possédant un très grand potentiel (cf. paragraphe I.1.4.1.5). Cependant, un défaut de glycosylation majeur présent chez ces glycoprotéines recombinantes reste l'absence de dérivés sialylés à l'extrémité non-réductrice des *N*-glycanes, des résidus dont l'importance a été abordée dans le paragraphe **I.1.5**.

Nous nous proposons ici d'utiliser cette glycosynthase pour greffer un thio-oligosaccharide porteur d'un résidu sialylé sur des résidus *N*-acétylglucosaminyles accepteurs. Ces résidus présentent en effet une structure proche de ceux naturellement acceptés dans le site actif de l'enzyme. D'autre part, ils peuvent être accessibles ou rendus accessibles plus facilement en position terminale des glycanes d'une *N*-glycoprotéine. Une représentation schématique de l'approche envisagée est présentée sur la **Figure 20**.

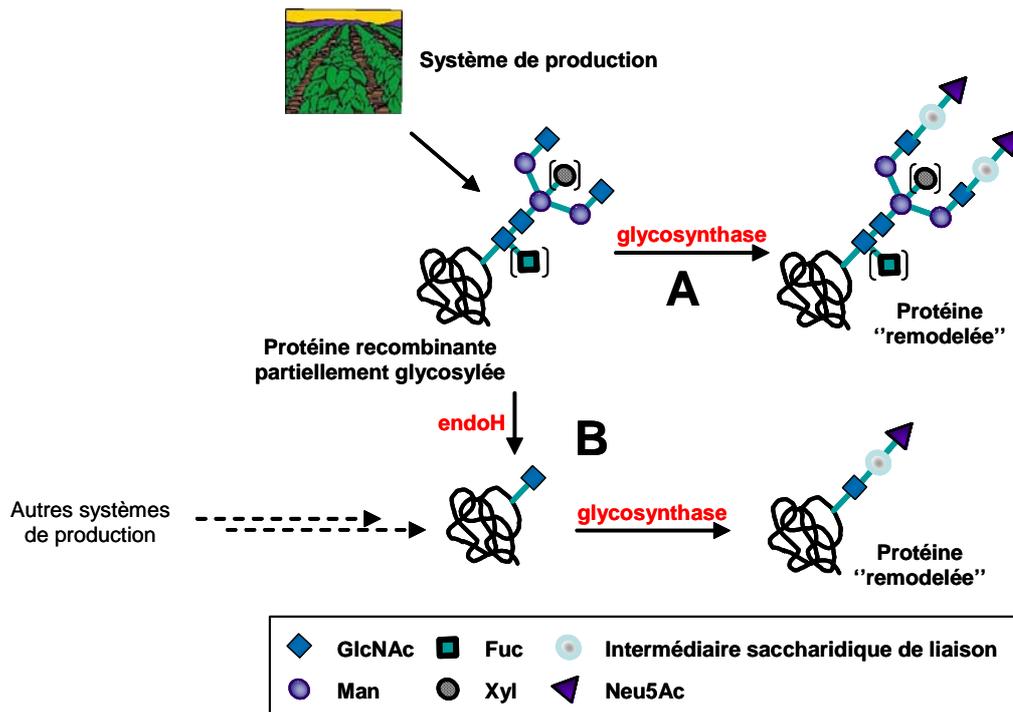


Figure 20 : Approche envisagée pour le remodelage de la *N*-glycosylation des protéines recombinantes produites chez les plantes.

Comme le montre la **Figure 20**, les résidus *N*-acétylglucosaminyles pourraient être soit présents au départ aux extrémités non-réductrices des glycanes à la surface de la protéine exprimée soit rendus accessibles *via* l'action d'une endo-*N*-acétylglucosaminidase, qui hydrolyse sélectivement le disaccharide GlcNAc(β 1-4)GlcNAc branché sur la protéine. Ensuite, l'utilisation de la glycosynthase Cel7B E197A d'*H. insolens* permettrait de greffer des oligosaccharides, porteurs de résidus sialylés.

Les oligosaccharides envisagés sont des thio-oligosaccharides où l'un des oxygènes d'une liaison glycosidique est substitué par un atome de soufre. La présence de celui-ci confère à la liaison thioglycosidique une résistance à l'hydrolyse enzymatique [190]. Les thio-oligosaccharides se comportent d'ailleurs généralement comme des inhibiteurs compétitifs des glycoside hydrolases. Cette liaison est également moins sensible à l'hydrolyse acide, et facilite donc la manipulation et la purification de ces composés.

Des études conformationnelles menées grâce à des données de cristallisation et de RMN indiquent que la conformation des thio-oligosaccharides est pratiquement identique à celle de leurs homologues oxygénés. Ces dérivés thioglycosides présentent une plus grande

flexibilité en solution, mais leur conformation permet tout de même leur reconnaissance par des glycoside hydrolases, des lectines ou des récepteurs protéiques.

Ainsi, la présence d'une liaison thioglycosidique au sein des glycanes d'une *N*-glycoprotéine permettrait une meilleure résistance face à l'hydrolyse par les sialidases [191] (notamment présentes dans le sérum), augmentant ainsi la durée de vie *in vivo* d'une telle glycoprotéine et donc son efficacité s'il s'agit d'une glycoprotéine d'intérêt thérapeutique.

Toutefois, l'approche envisagée nécessite une modification de la glycosynthase. En effet, l'enzyme Cel7B E197A d'*H. insolens* présente l'inconvénient de ne pas permettre le couplage d'oligosaccharides sur des résidus *N*-acétylglucosaminyles. Dans le cadre de cette thèse, il s'agit donc de modifier la spécificité de substrat de cette glycosynthase. Pour cela nous avons choisi d'adopter une approche rationnelle pour l'ingénierie de cette enzyme, la structure tridimensionnelle de l'enzyme native étant disponible pour des études de modélisation moléculaire.

Chapitre II

**ETUDES PRELIMINAIRES
SUR LA GLYCOSYNTHESE CEL7B E197A**

II ETUDES PRELIMINAIRES SUR LA GLYCOSYNTHESE CEL7B E197A

Le but de notre approche est de greffer sur une protéine partiellement glycosylée, portant des résidus *N*-acétylglucosaminyles aux extrémités non-réductrices des *N*-glycanes, un oligosaccharide porteur d'un acide *N*-acétylneuraminique. Cette étude préliminaire consiste à déterminer l'activité de la glycosynthase Cel7B E197A d'*Humicola insolens* vis-à-vis de substrats non-naturels. Tout d'abord, un glycopeptide modèle sera testé en tant qu'accepteur. Puis, un trisaccharide de type sialyllactose sera utilisé en tant que substrat donneur potentiel pour la glycosynthase.

II.1 Activité de la glycosynthase vis-à-vis d'un glycopeptide accepteur

Afin de tester l'activité de la glycosynthase vis-à-vis d'un accepteur glycopeptidique, nous avons envisagé la synthèse du *N*-glycopeptide **14** (Figure 21) porteur d'un résidu monosaccharidique tel que le *N*- β -D-glucoside, les β -D-glucosides étant connus pour être acceptés dans le sous-site +1 de l'enzyme. Le substrat donneur utilisé sera un fluorure de lactosyle **16**, lui aussi connu pour être un bon donneur pour la glycosynthase.

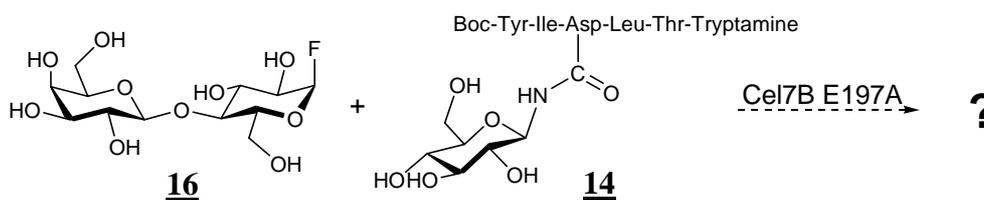


Figure 21 : Réaction de couplage avec la glycosynthase Cel7B E197A entre le donneur **16** et l'accepteur **14**.

Le but est de déterminer si l'enzyme Cel7B E197A d'*Humicola insolens* catalyse le transfert du lactose sur un monosaccharide porté par une chaîne peptidique. Cette dernière pourrait créer un encombrement stérique empêchant le substrat accepteur de se positionner

correctement dans le site actif de l'enzyme et ainsi ne pas permettre le greffage du donneur avec la formation d'une liaison β 1-4 entre les deux résidus D-glucosyles.

II.1.1 Préparation des synthons

II.1.1.1 Synthèse du glucopeptide 14

La synthèse du glucopeptide 14 a été envisagée selon une approche convergente [192] (**Figure 22**), où le monosaccharide est incorporé dans un peptide préformé présentant un acide aspartique dans sa séquence.

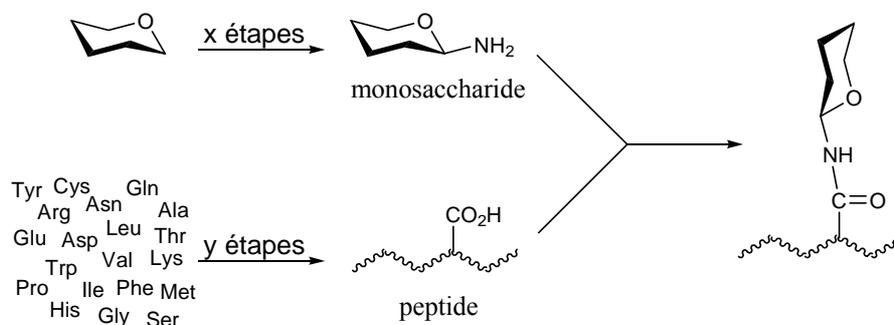


Figure 22 : Principe de l'approche convergente.

Le peptide cible 14 est un pentapeptide protégé à son extrémité N-terminale par un groupement *tert*-butyloxycarbonyle (Boc) et à son extrémité C-terminale par la tryptamine *via* une liaison peptidique classique (le greffage de la tryptamine avait été envisagé dans l'idée de préparer un substrat bifluorescent glycopeptidique pour l'étude de l'activité d'enzymes telles que les *endo-N*-acétylglucosaminidases (ENGases) ou les peptide *N*-glycanases (PNGases)).

Le pentapeptide est obtenu par une série de couplages/déprotections de l'extrémité C-terminale vers l'extrémité N-terminale (**Schéma 2**). Les réactions de couplage sont effectuées selon une méthodologie classique [193] en présence de deux réactifs de couplage, la dicyclohexylcarbodiimide (DCC) et le *N*-hydroxybenzotriazole (HOBT), et de triéthylamine.

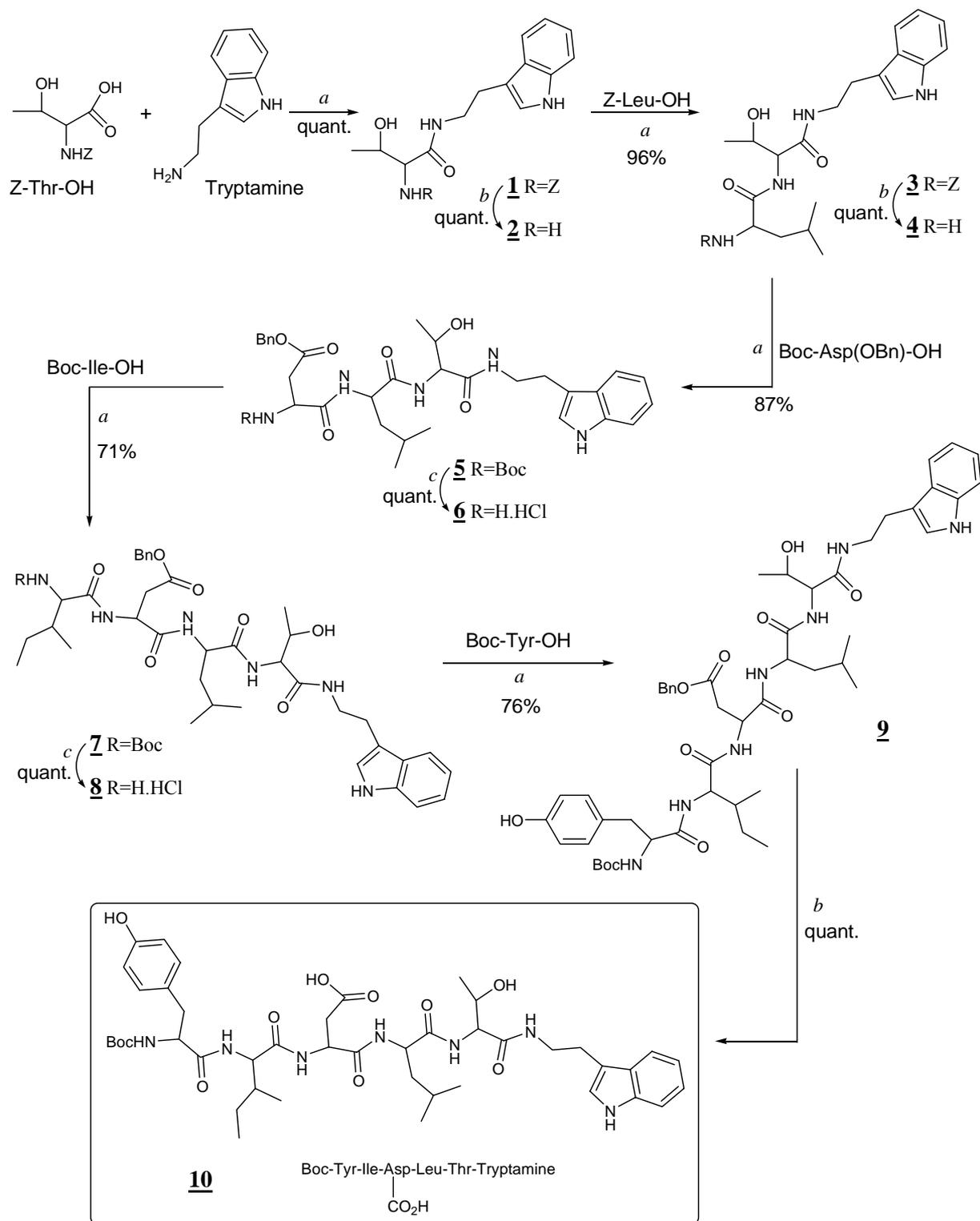


Schéma 2 : Préparation du pentapeptide 10.

Conditions : (a) Et₃N, HOBT, DCC, DMF; (b) H₂, Pd(OH)₂/C, MeOH; (c) HCl 4M, dioxane.
 Abréviations : Bn, benzyle; Z, benzyloxycarbonyle; Boc, *tert*-butyloxycarbonyle.

Le solvant utilisé est le diméthylformamide (DMF). La DCC permet d'activer la fonction acide d'un acide aminé; le HOBt permet, quant à lui, de limiter la racémisation [194]. La présence d'une base, la triéthylamine, dans le milieu permet de s'assurer de la déprotonation de la fonction amine. Les couplages sont réalisés avec d'excellents rendements; on note tout de même une diminution de la réactivité du peptide au fur et à mesure de son élongation.

L'extrémité N-terminale des différents acides aminés est dans un premier temps protégée par un groupement benzyloxycarbonyle (Z), facilement éliminé par hydrogénolyse en présence de palladium sur charbon dans le méthanol avec un rendement quantitatif (c'est le cas des composés **1** à **4**). Après l'introduction de l'acide aspartique dont la fonction acide de la chaîne latérale est protégée sous la forme d'un ester benzylique (Bn), l'extrémité N-terminale est alors protégée par un groupement Boc orthogonal au groupe Bn et facilement éliminé en présence d'acide chlorhydrique dans le dioxane (c'est le cas des composés **5** à **9**) [193].

Finalement, la chaîne latérale de l'acide aspartique du peptide **9** est déprotégée par hydrogénolyse en présence d'une quantité catalytique de palladium sur charbon avec un rendement quantitatif pour donner le synthon **10**.

La glucosylamine **13** est obtenue en 3 étapes à partir du D-glucose per-*O*-acétylé commercial (**Schéma 3**). La position anomère de ce dernier est dans un premier temps fonctionnalisée par un azoture selon une méthode de synthèse classique : l'intermédiaire bromure d' α -glucosyle, formé par action de l'acide bromhydrique dans l'acide acétique, est directement converti en azoture de β -glucosyle *via* une substitution nucléophile de l'halogénure par traitement avec l'azoture de sodium dans le DMF [195]. Le composé **11** obtenu est ensuite désacétylé en présence de méthylate de sodium dans le méthanol [196]. Enfin, le groupement azoture est réduit par hydrogénation catalytique en présence de palladium sur charbon dans le méthanol [197] aboutissant à la formation de la D-glucosylamine **13** de configuration β .

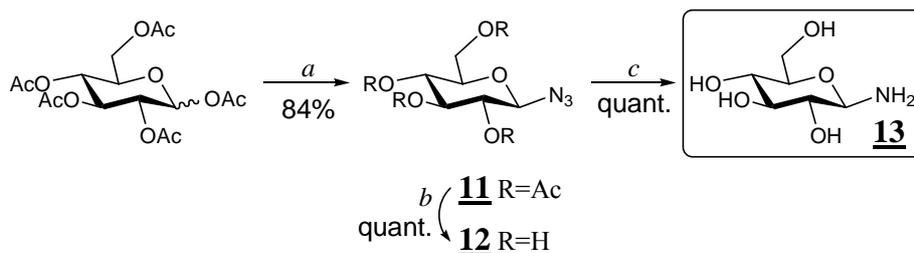


Schéma 3 : Préparation de la β-D-glucosylamine 13.

Conditions : (a) i. HBr/AcOH, CH₂Cl₂; ii. NaN₃, DMF; (b) MeONa, MeOH; (c) H₂, Pd(OH)₂/C, MeOH.

Le caractère nucléophile du groupe NH₂ peut être diminué avec la protection des hydroxyles du monosaccharide par des acétates [198]. L'utilisation d'un produit dont les hydroxyles sont libres permet à la glucosylamine d'être plus nucléophile, et donc plus réactive lors du couplage peptidique. Cela évite par ailleurs la migration d'acétate, qui formerait un groupement NHAc inactif.

D'autre part, la séquence peptidique choisie contient la séquence consensus présente au niveau des sites de N-glycosylation des protéines, le séquen Asn-X-Ser/Thr, X pouvant être n'importe quel acide aminé sauf la proline. Dans notre cas, nous avons choisi un résidu leucine pour X.

D'ailleurs, le choix de l'acide aminé adjacent (chaîne latérale notée R sur la **Figure 23**) à l'acide aspartique du côté C-terminal [199] revêt une importance particulière au moment du couplage entre le peptide **10** et le monosaccharide aminé **13**. En effet, des réactions parasites peuvent avoir lieu au cours de la synthèse convergente lors du couplage, notamment la formation (par cyclisation intramoléculaire) d'une aspartimide (**Figure 23**) [200].

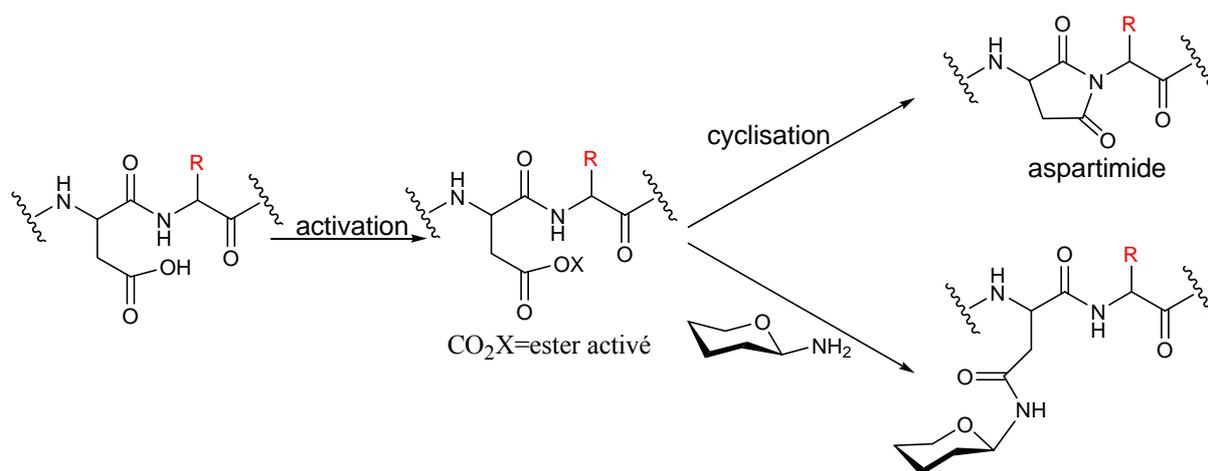


Figure 23 : Formation de l'aspartimide pendant la synthèse convergente d'un *N*-glycopeptide.

Cette réaction parasite a lieu au moment de l'activation de la fonction acide de la chaîne latérale de l'acide aspartique. Des études ont été menées avec des peptides de séquences différentes et ont ainsi montré que des résidus tels que la sérine, la thréonine, l'alanine ou la glycine favorisaient la formation de l'aspartimide [199,201].

Le couplage entre le peptide **10** et la glucosylamine **13** a finalement été réalisé (Schéma 4).

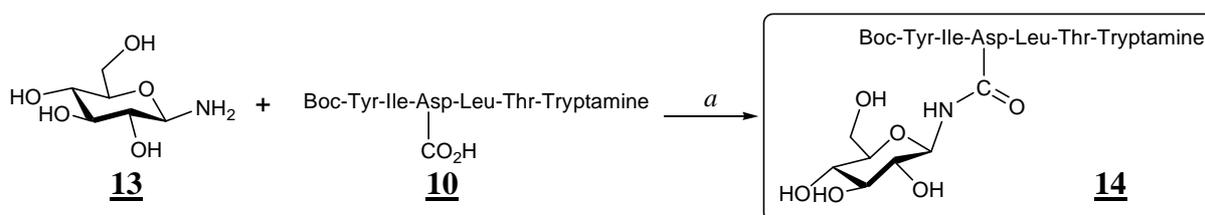


Schéma 4 : Préparation du glycopeptide **14**.

Conditions : (a) HOBt, BOP, DMF.

Les conditions utilisées pour ce couplage [198,202] sont l'association de l'HOBt avec un sel de phosphonium dérivé de l'HOBt, le BOP, dans le DMF. L'absence de base dans le milieu permet d'éviter à nouveau la formation de l'aspartimide (favorisée en milieu basique); c'est la glucosylamine, introduite en excès, qui joue le rôle à la fois de nucléophile et de base ici. Le rendement en produit isolé reste médiocre (de l'ordre de 20%) du fait de problèmes rencontrés lors de la purification du glycopeptide **14**.

L'anométrie de la liaison *N*-glycopeptidique a été vérifiée par une analyse RMN du produit. Le spectre RMN ^1H présente ainsi un doublet caractéristique à 4.91 ppm correspondant au proton H-1 de l'unité monosaccharidique. La constante de couplage $J_{1,2}$ de 9.0 Hz est caractéristique d'une configuration β au niveau de la position anomère.

II.1.1.2 Préparation du fluorure d' α -lactosyle

L'obtention du dérivé fluoré **16**, substrat donneur de la glycosynthase Cel7B E197A d'*Humicola insolens* est présentée sur le **Schéma 5**.

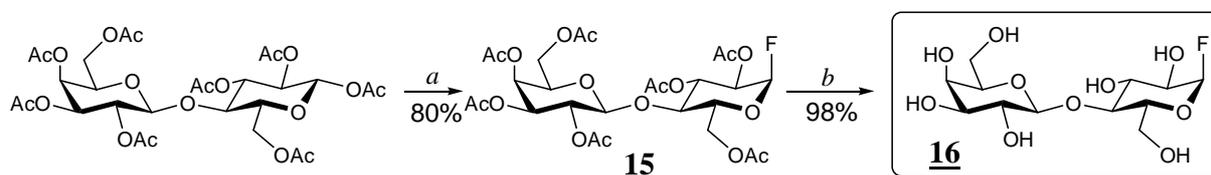


Schéma 5 : Préparation du fluorure d' α -lactosyle **16.**

Conditions : (a) i. HF/pyridine (7:3, v/v); ii. AcOH.NH₂NH₂, DMF; (b) MeONa, MeOH.

L'action du fluorure d'hydrogène [**203**] sur le lactose per-*O*-acétylé disponible au laboratoire (majoritairement de configuration β) conduit au fluorure acétylé **15** [**204**]. Afin de faciliter la purification, le lactose per-*O*-acétylé d'anométrie α n'ayant pas réagi est converti en l'hémiacétal correspondant par traitement du brut réactionnel avec l'acétate d'hydrazine dans le DMF [**205**]. Le fluorure **15** est ensuite déprotégé dans des conditions classiques [**196**] pour conduire au synthon **16**.

II.1.2 Test de couplage enzymatique

Le couplage du fluorure de lactosyle **16** sur le glucopeptide **14** a été testé en présence de la glycosynthase Cel7B E197A d'*Humicola insolens* (**Schéma 6**).

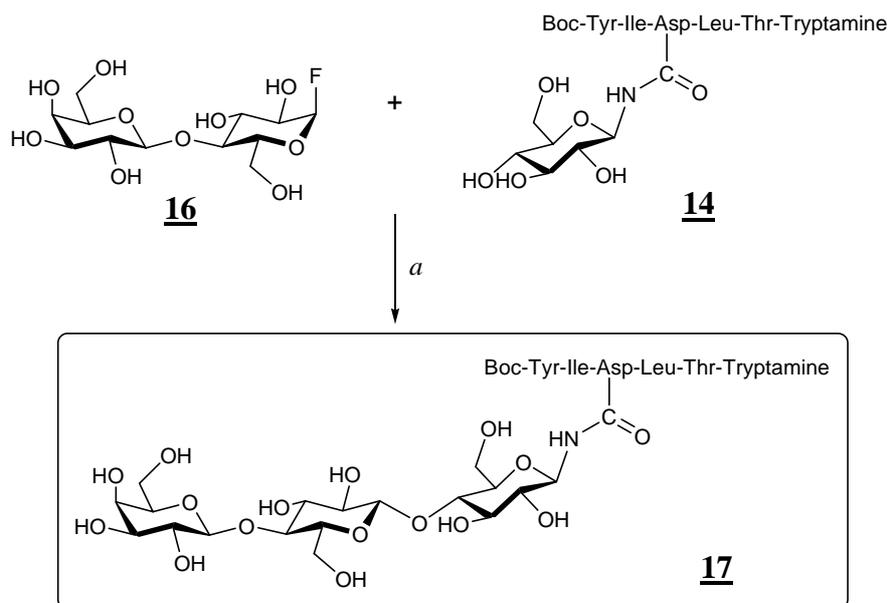


Schéma 6 : Couplage enzymatique entre le donneur **16 et l'accepteur **14**.**

Conditions : (a) Cel7B E197A *H. insolens*, tampon phosphate de sodium 50 mM pH 7/MeCN (4:1, v/v).

L'analyse par spectrométrie de masse réalisée sur le brut réactionnel après lyophilisation a permis la détection d'un composé de masse $m/z = 1373$ correspondant au produit **17** sous la forme $[M+Na]^+$. La glycosynthase peut donc utiliser un substrat de type glucopeptidique en tant qu'accepteur dans la réaction de glycosylation enzymatique.

II.2 Activité de la glycosynthase vis-à-vis d'un fluorure de 6''-S-sialyl-6''-thiolactosyle donneur

Après avoir testé un substrat glucopeptidique en tant qu'accepteur potentiel de la glycosynthase, il s'agit maintenant d'étudier l'activité de l'enzyme vis-à-vis d'un donneur de type sialyllactose (**Figure 24**).

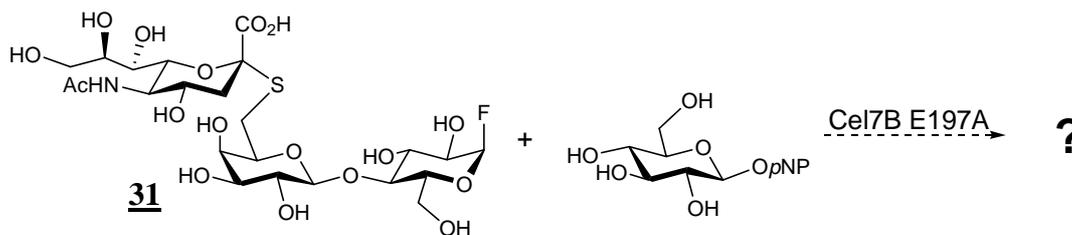


Figure 24 : Réaction de couplage entre le donneur **31 et l'accepteur β-D-glucopyranoside de *p*-nitrophényle.**

Le fluorure de 6^{II}-*S*-sialyl-6^{II}-thiolactosyle **31** est un thio-oligosaccharide, dont un atome d'oxygène inter-osidique est substitué par un atome de soufre. Il possède donc une liaison thioglycosidique stable vis-à-vis de l'hydrolyse par des glycoside hydrolases.

Le substrat accepteur utilisé dans la réaction de couplage avec la glycosynthase Cel7B E197A est un β-D-glucopyranoside de *p*-nitrophényle commercial.

II.2.1 Synthèse du fluorure de 6^{II}-*S*-sialyl-6^{II}-thiolactosyle **31**

La stratégie adoptée pour l'obtention de l'α-fluorure **31** est présentée dans le **Schéma 7**. L'étape clé est constituée par un couplage chimique pour la formation d'une liaison thioglycosidique α2-6.

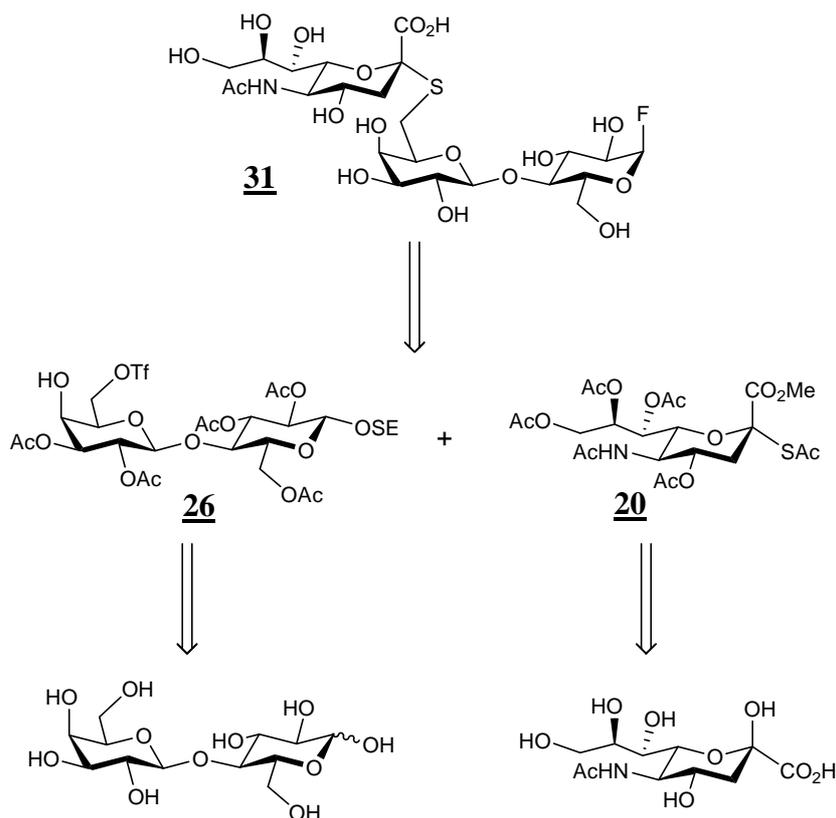


Schéma 7 : Rétrosynthèse du composé 31.

Abréviation : SE, $(\text{CH}_2)_2\text{SiMe}_3$.

La synthèse du composé **31** repose sur la thioglycosylation régio- et stéréospécifique de la position 6^{II} d'un précurseur lactoside **26** par un dérivé thioglycoside **20** de l'acide *N*-acétylneuraminique. Le triflate accepteur **26** et le composé **20** sont eux-mêmes obtenus en plusieurs étapes à partir respectivement du lactose et de l'acide *N*-acétylneuraminique.

II.2.1.1 Préparation du précurseur sialyle 20

Le précurseur **20** est obtenu en trois étapes à partir de l'acide *N*-acétylneuraminique commercial (**Schéma 8**).

Dans une première étape, la préparation de l'ester méthylique **18** est réalisée en présence d'Amberlite IR-120 (H^+) dans le méthanol [206]. Ensuite, le composé **18** est per-*O*-acétylé puis transformé en chlorure "one pot" avec le chlorure d'acétyle dans l'acide acétique, l'acide chlorhydrique nécessaire étant généré *in situ* pendant la réaction [207].

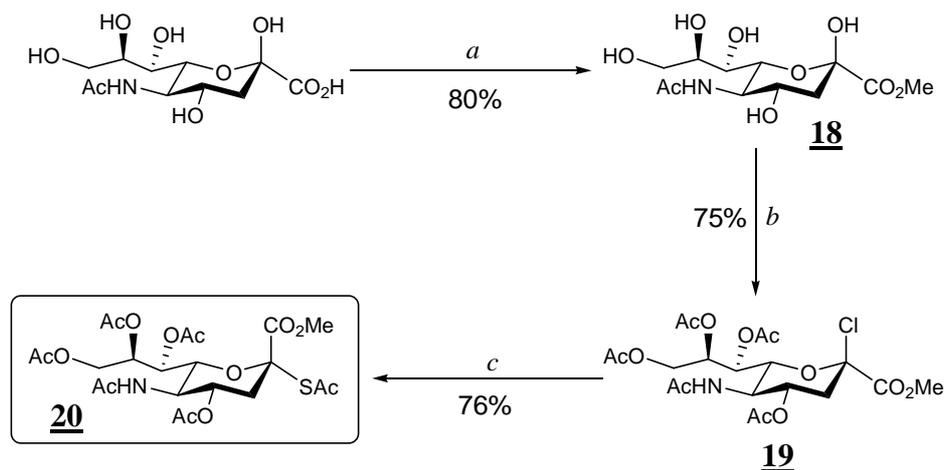


Schéma 8 : Préparation du précurseur sialyle 20.

Conditions : (a) Amberlite IR-120 (H^+), MeOH; (b) AcCl/AcOH;
(c) KSAc, TBAHS, EtOAc/ Na_2CO_3 1M (1:1, v/v).

Finalement, l'atome de chlore du composé 19 est déplacé par le thioacétate de potassium dans un milieu hydro-organique, en présence d'un catalyseur de transfert de phase, l'hydrogénosulfate de tétrabutylammonium (TBAHS) [208] pour donner le composé 20 attendu.

II.2.1.2 Préparation du précurseur lactoside 25

La synthèse du lactoside partiellement protégé 25 est effectuée en 5 étapes à partir du lactose commercial (Schéma 9). Le but est de préparer un dérivé lactoside dont les hydroxyles en 4^{II} et 6^{II} sont dans un premier temps protégés sélectivement puis, après acétylation des hydroxyles libres restants, déprotégés.

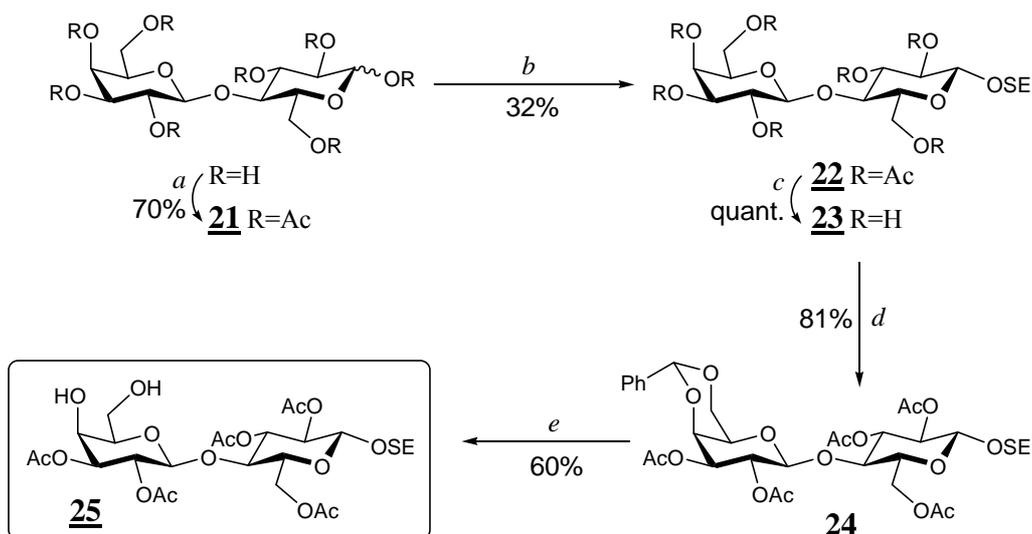


Schéma 9 : Préparation du précurseur lactoside **25**.

Conditions : (a) KOAc, Ac₂O; (b) i. HBr/AcOH, CH₂Cl₂; ii. HO(CH₂)₂SiMe₃, AgClO₄, AgCO₃, CH₂Cl₂; (c) MeONa, MeOH; (d) i. α,α -diméthoxytoluène, *p*TsOH, MeCN; ii. DMAP, Ac₂O, pyridine; (e) AcOH 80%.
Abréviation : SE, (CH₂)₂SiMe₃.

L'obtention du lactose per-*O*-acétylé **21** sous la forme d'un mélange d'anomères est réalisée en présence d'acétate de potassium dans l'anhydride acétique à partir du lactose commercial [209]. Après avoir introduit le brome en position anomère, la glycosylation sur le 2-(triméthylsilyl)éthanol [210] est effectuée en présence de perchlorate d'argent et de carbonate d'argent dans le dichlorométhane [211] et permet d'obtenir le β -lactoside **22**. Après désacétylation du composé **22** par le méthylate de sodium dans le méthanol [196,211], les hydroxyles 4^{II} et 6^{II} du composé **23** sont protégés par un *O*-benzylidène. L'acétalation est effectuée par réaction avec l' α,α -diméthoxytoluène en présence d'acide *p*-toluène sulfonique dans l'acétonitrile [212]. Le benzylidène intermédiaire est acétylé *in situ* par un mélange d'anhydride acétique et de pyridine en présence de diméthylaminopyridine pour aboutir à la formation du composé **24**. Finalement, le benzylidène est hydrolysé à chaud en présence d'acide acétique à 80% dans l'eau [212]. Le précurseur lactosyle **25** présente alors deux hydroxyles libres et sera sélectivement fonctionnalisé.

II.2.1.3 Thioglycosylation

La voie de synthèse du trisaccharide fluoré **31** repose sur l'étape clé de thioglycosylation afin de former une liaison α 2-6 (Schéma 10).

La réaction de couplage suppose au préalable l'activation du composé lactosyle **25** dont l'hydroxyle primaire libre 6^{II} (plus réactif que l'hydroxyle 4^{II}) est transformé en triflate par action de l'anhydride triflique (Tf₂O) dans un mélange pyridine/dichlorométhane à -20°C pour donner le composé **26** utilisé directement après lavage sans autre purification [212]. Le déplacement du triflate **26** par le thiolate, généré *in situ* en présence de diéthylamine à partir du composé **20**, est réalisé à froid dans le DMF [212]. Le produit de thioglycosylation obtenu est alors acétylé *in situ* pour finalement obtenir le composé **27**.

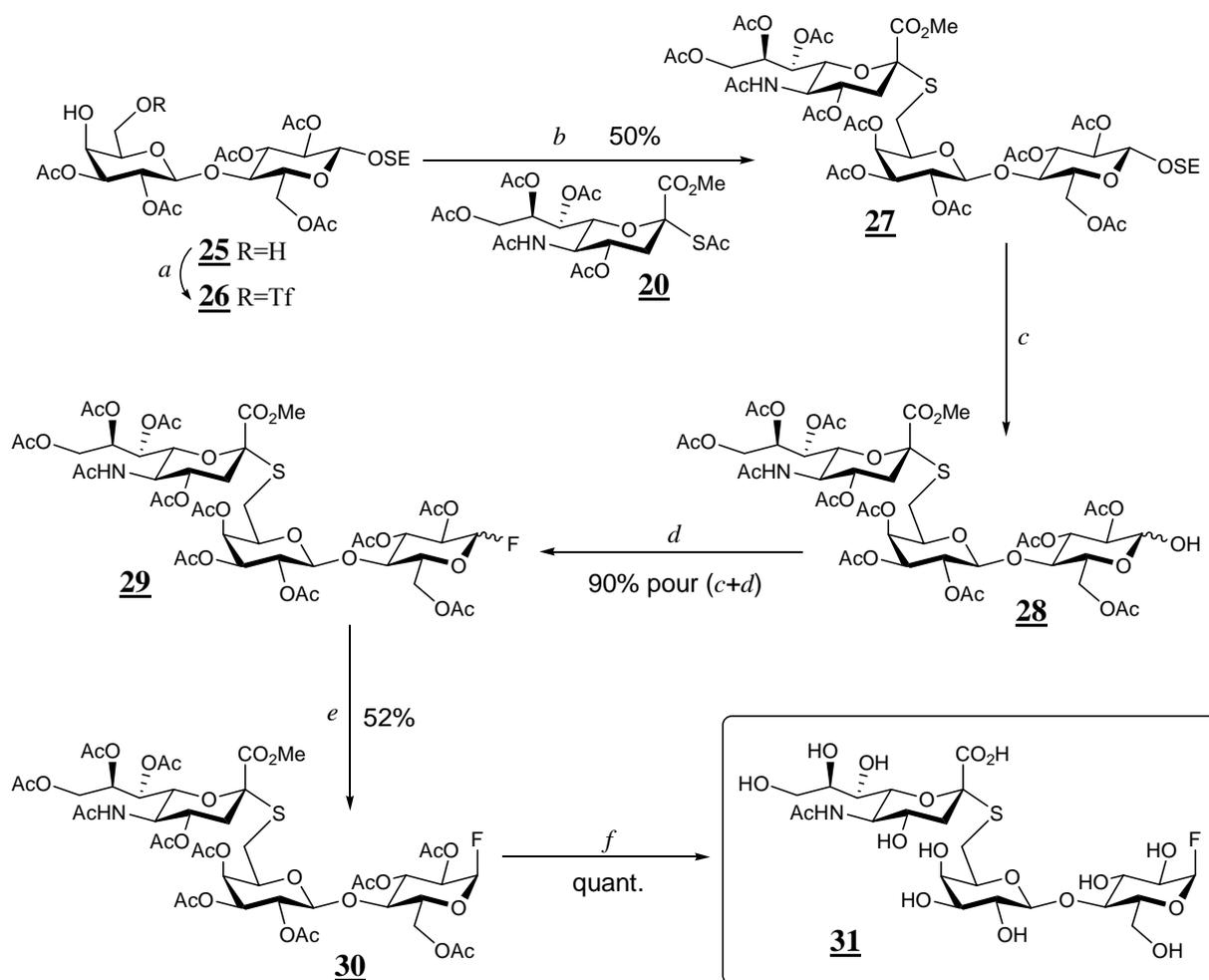


Schéma 10 : Thioglycosylation entre le donneur **20** et l'accepteur **25**.

Conditions : (a) Tf₂O, pyridine, CH₂Cl₂; (b) i. Et₂NH, DMF; ii. Ac₂O, pyridine; (c) TFA, CH₂Cl₂;

(d) DAST, CH₂Cl₂; (e) HF/pyridine (7:3, v/v); (f) i. MeONa, MeOH; ii. NaOH 0.1M.

Abréviation : SE, (CH₂)₂SiMe₃.

La fonctionnalisation de la position anomère du résidu D-glucosyle de l'extrémité réductrice du trisaccharide **27** nécessite quelques étapes supplémentaires pour l'obtention du fluorure **31**.

La déprotection sélective de la position anomère, en présence d'acide trifluoroacétique (TFA) dans le dichlorométhane [212], conduit à l'hémiacétal **28** sous la forme d'un mélange d'anomères. Puis la formation du fluorure **29**, toujours en mélange d'anomères, est réalisée en présence de trifluorure de diéthylaminosulfure (DAST) à froid dans le dichlorométhane [213]. L'anomérisation du β -fluorure en α -fluorure est ensuite réalisée à basse température (-50°C puis -10°C) par le fluorure d'hydrogène dans la pyridine [214] pour conduire au seul fluorure d'anomérisation α **30**. Finalement, le trisaccharide **31** est obtenu après désacétylation [196] et déprotection de la fonction acide de l'acide *N*-acétylneuraminique par les actions successives de méthylate de sodium dans le méthanol et d'une solution aqueuse d'hydroxyde de sodium [215].

II.2.2 Test de couplage enzymatique

Le couplage du fluorure de 6^{II}-*S*-sialyl-6^{II}-thiolactosyle **31** sur le β -D-glucopyranoside de *p*-nitrophényle a été testé en présence de la glycosynthase Cel7B E197A d'*Humicola insolens* (Schéma 11).

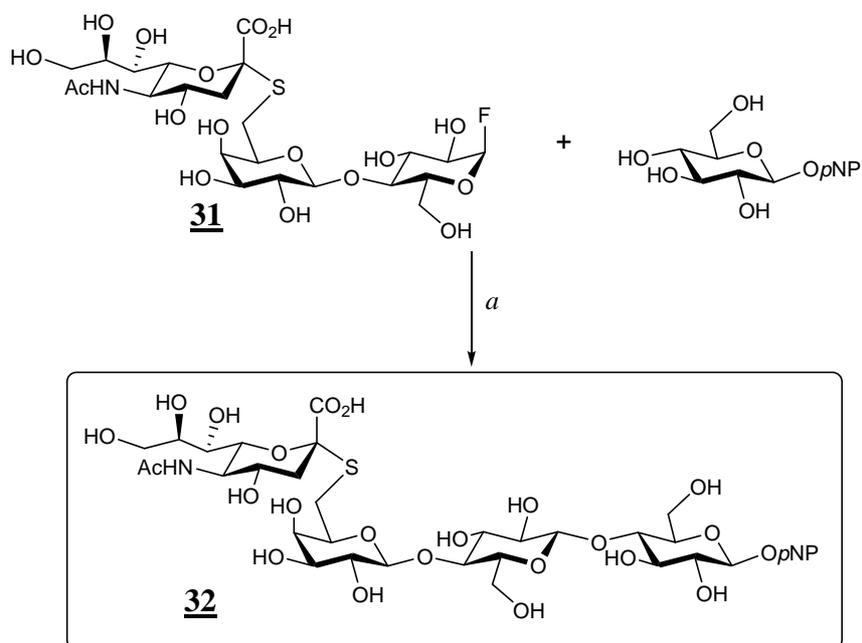


Schéma 11 : Couplage enzymatique entre le donneur **31** et le β -D-glucopyranoside de *p*-nitrophényle.

Conditions : (a) Cel7B E197A *H. insolens*, tampon phosphate de sodium 50 mM pH 7.

L'analyse qualitative par chromatographie sur couche mince de la réaction de couplage montre la formation d'un produit UV-visible (grâce au groupement *p*-nitrophényle) et révélé à l'acide sulfurique à chaud (test caractéristique des sucres), en accord avec le produit de couplage **32** attendu.

La mesure de masse à basse résolution réalisée sur le brut réactionnel après lyophilisation a permis de confirmer la formation d'un composé dont l'ion $m/z = 931 [M-H]^-$ concorde avec la structure tétrasaccharidique du composé **32** attendu. Ce test permet ainsi de valider à nouveau l'activité de la glycosynthase vis-à-vis d'un substrat substitué en 6^{II}, comme l'avaient montré les travaux de Fort et coll. [130].

L'absence d'interactions entre la protéine et l'oligosaccharide dans le sous-site -2 en position 6^{II} de l'unité glycosidique a déjà été exploitée (cf. paragraphe **1.2.4**) et l'enzyme mutée Cel7B E197A d'*Humicola insolens* a été utilisée pour la préparation de β 1,4-glucanes substitués.

II.3 Conclusion

Au cours de ce chapitre, nous avons synthétisé des substrats potentiels puis nous les avons testé en tant que donneurs et accepteurs dans des réactions de couplage enzymatique catalysées par la glycosynthase Cel7B E197A d'*Humicola insolens*.

Ces couplages enzymatiques ont tout d'abord permis de confirmer la capacité de l'enzyme mutée à utiliser des substrats donneurs et accepteurs différents de ses substrats naturels.

Nous avons pu ainsi mettre en évidence la possibilité d'utiliser la glycosynthase pour greffer un oligosaccharide complexe sur un substrat glycopeptidique modèle, validant par la même occasion le concept envisagé pour le remodelage de la glycosylation des protéines recombinantes. Toutefois, cela suppose que la glycosynthase utilisée permette de greffer des structures glycosidiques sur un résidu *N*-acétylglucosaminyle terminal.

Cette glycosynthase ne catalysant pas de réaction de glycosylation sur un résidu *N*-acétylglucosaminyle, il restait donc à modifier la spécificité de substrat de cette enzyme, ce que nous allons décrire dans les chapitres suivants.

Chapitre III

**INGENIERIE DE LA GLYCOSYNTHASE
CEL7B E197A D'*HUMICOLA INSOLENS***

III INGENIERIE DE LA GLYCOSYNTHASE CEL7B E197A d'*H. INSOLENS*

Le chapitre précédent a permis de montrer qu'il était possible de greffer un oligosaccharide sur un accepteur glucopeptidique en utilisant la glycosynthase Cel7B E197A d'*Humicola insolens*. Toutefois, la spécificité de substrat au niveau du sous-site accepteur reste un facteur limitant son utilisation. En particulier, le fait que l'enzyme ne permette pas de glycosyler un résidu *N*-acétylglucosaminyle empêche son utilisation dans le remodelage de la glycosylation des protéines par un transfert de dérivés sialylés sur un résidu *N*-acétylglucosaminyle terminal.

Dans ce cadre et dans un souci d'accroître le potentiel synthétique de la glycosynthase Cel7B E197A d'*H. insolens*, nous avons décidé de faire évoluer l'enzyme en suivant une stratégie de conception rationnelle, qui était possible grâce à la connaissance de la structure tridimensionnelle de la protéine. Par des études de modélisation moléculaire, il s'agissait d'étudier la spécificité de substrat de l'enzyme dans son sous-site accepteur +1 et en particulier, de déterminer les interactions d'un groupement *N*-acétyl, remplaçant l'hydroxyle en 2 de l'unité saccharidique, avec la protéine, pour proposer des mutations ponctuelles permettant la fixation d'un résidu *N*-acétylglucosaminyle.

III.1 Etude de la spécificité de l'enzyme

L'ensemble de l'étude a été réalisée en Janvier 2002 et dirigée en collaboration avec Pr Pedro M. Coutinho, alors à l'Instituto Superior Técnico de l'Université Technique de Lisbonne (Portugal).

Comme écrit au paragraphe **I.2.4**, la glycosynthase Cel7B E197A est issue de l'endoglucanase Cel7B d'*Humicola insolens* dont le résidu catalytique nucléophile Glu-197 a

été muté en alanine. Cette cellulase appartient à la famille GH7 des glycoside hydrolases dans la classification CAZy^f.

III.1.1 La famille GH7

La famille GH7 regroupe 120 protéines (CAZy version Octobre 2004), divisées en deux sous-familles mono-spécifiques, une contenant des endoglucanases et l'autre contenant des cellobiohydrolases (Pedro M. Coutinho, communication personnelle). La grande majorité de ces protéines sont d'origine fongique. Des structures tridimensionnelles sont disponibles pour 6 protéines différentes et sont répertoriées dans le **Tableau 10**. Plusieurs accessions à la "Protein Data Bank" (PDB)^g existent pour une même protéine, seule ou en complexe avec un substrat ou inhibiteur.

Il est à noter que nous avons choisi de conserver tout au long de ce manuscrit le nom de *Trichoderma reesei*, traditionnellement utilisé dans l'industrie. Mais l'organisme peut aussi être désigné sous le nom de son anamorphe *Hypocrea jecorina* (comme reporté dans la classification CAZy et dans la banque de données Taxonomy du NCBI).

^f Coutinho, P.M. & Henrissat, B. (1999) Carbohydrate-Active Enzymes server : <http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html>

^g Site Internet : <http://www.pdb.org/>

Organisme	Protéine	PDB	Détails^a
<i>Fusarium oxysporum</i>	<i>endo</i> - β 1,4-glucanase	1OVW	en complexe avec un thiocellopentaose (dans les sous-sites de -2 à +1);
		2OVW	en complexe avec le cellobiose (-2,-1);
		3OVW	structure native;
		4OVW	en complexe avec un époxybutyl cellobiose (-2,-1).
<i>Humicola insolens</i>	<i>endo</i> - β 1,4-glucanase (Cel7B)	1A39	mutant S37W P39W;
		1DYM	mutant E197A;
		1OJI	
		1OJJ	mutants E197A ou E197S;
		1OJK	
		2A39	structure native.
<i>Trichoderma reesei</i>	<i>endo</i> - β 1,4-glucanase (Cel7B)	1EG1	délétion du domaine (372-436).
<i>Trichoderma reesei</i>	cellobiohydrolase (Cel7A)	1CEL	en complexe avec le (2-iodobenzyl)thiogluco-side (+1,+2);
		1DY4	en complexe avec le (<i>S</i>)-propranolol (-1,+1);
		1EGN	mutant E223S A224H L225V T226A D262G;
		1EK7	n.d.
		1Q2B	mutant D241C D249C;
		1Q2E	en complexe avec le cellotétraose (+ délétion d'une boucle 245-252);
		2CEL	mutant E212Q;
		3CEL	mutant E212Q en complexe avec le cellobiose (+1,+2);
		4CEL	mutant D214N;
		5CEL	mutant E212Q en complexe avec 2 cellotétraoses (-7,-4 et -2,+2);
		6CEL	mutant E212Q en complexe avec le cellopentaose (-6,-2) et le cellotétraose (+1,+4);
		7CEL	mutant E217Q en complexe avec le cellohexaose (-7,-2) et le cellobiose (+1,+2);
8CEL	modèle théorique en complexe avec un cellononaose (-7,+2).		
<i>Phanerochaete chrysosporium</i>	cellobiohydrolase (Cel7D)	1GPI	domaine catalytique;
		1H46	domaine catalytique en complexe avec le (<i>R</i>)-propranolol.
<i>Talaromyces emersonii</i>	cellobiohydrolase	1Q9H	n.d.

Tableau 10 : Structures PDB disponibles à ce jour pour la famille GH7 des glycoside hydrolases.

^a n.d. : non disponible.

III.1.2 Etude préliminaire qualitative

Dans un premier temps, nous avons voulu observer de manière qualitative l'environnement du substituant *N*-acétyle dans le sous-site +1 du site actif de l'enzyme. Pour cela, nous avons utilisé comme référence le cellononaose modélisé en complexe dans la cellobiohydrolase Cel7A de *Trichoderma reesei* dont la structure est accessible sous le fichier PDB 8CEL.

Comme le montre la **Figure 25**, nous avons superposé une unité *N*-acétylglucosaminyle sur le résidu glucosyle présent dans le sous-site +1 de l'enzyme.

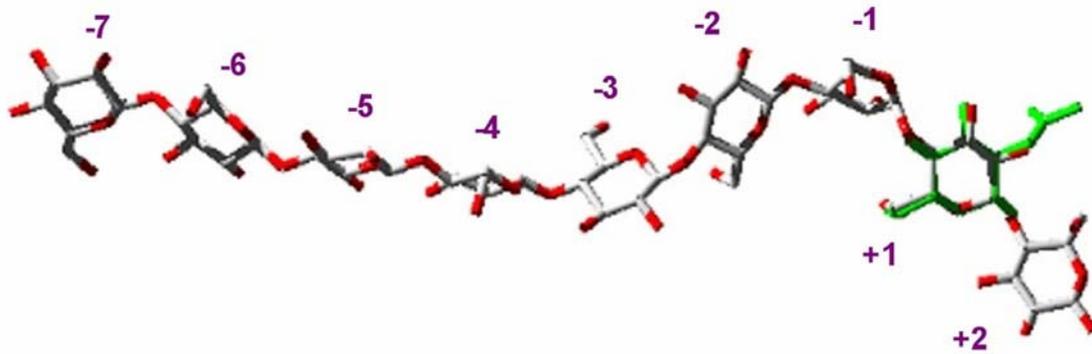


Figure 25 : Superposition d'un cellononaose et d'un résidu *N*-acétylglucosaminyle dans le sous-site +1.

En vert, le résidu *N*-acétylglucosaminyle.

Ce résidu *N*-acétylglucosaminyle a dans un deuxième temps été superposé dans le site actif de l'endoglucanase Cel7B d'*Humicola insolens* dont la structure tridimensionnelle est disponible sous le numéro d'accèsion PDB 2A39 (**Figure 26**).

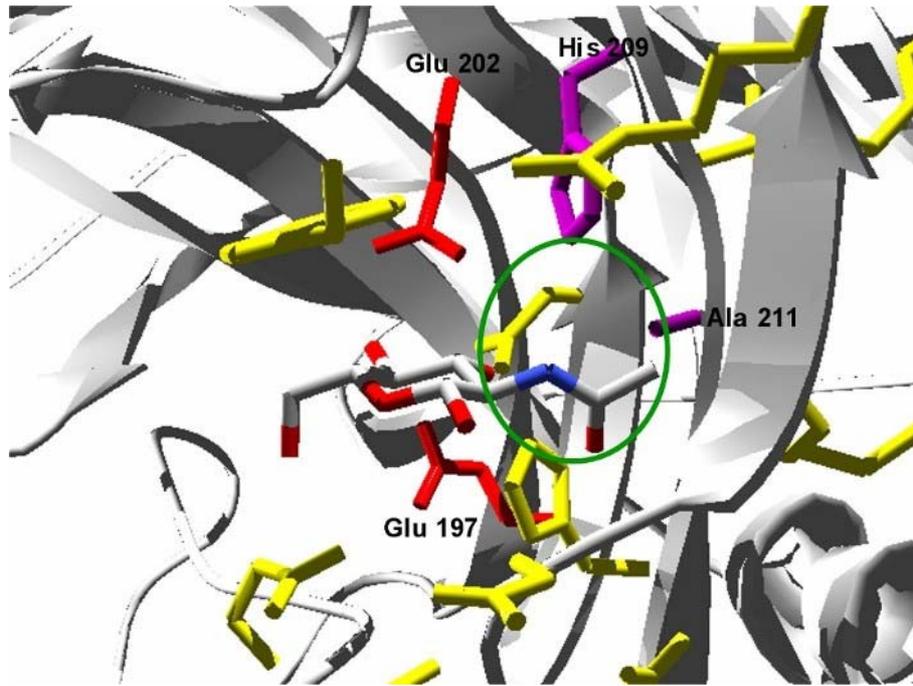


Figure 26 : Superposition d'un résidu *N*-acétylglucosaminyle dans le sous-site +1 de Cel7B d'*H. insolens*. En rouge, les deux résidus catalytiques; en violet, deux résidus susceptibles de limiter l'introduction d'une chaîne *N*-acétyl dans le site actif de l'enzyme.

Cette première superposition, présentée dans la **Figure 26**, permet de visualiser les interactions possibles entre l'enzyme et la chaîne latérale *N*-acétyl. Ainsi, des résidus tels que l'histidine en position 209 ou l'alanine en 211 de l'enzyme Cel7B peuvent gêner l'introduction d'une *N*-acétylglucosamine dans le sous-site +1 de l'enzyme. En effet, parmi les résidus mis en évidence, l'histidine est un acide aminé relativement volumineux.

III.1.3 Alignements de séquences

Afin d'avoir une idée de la nature et de la variabilité des acides aminés dans les positions 209 et 211 (dans la numérotation de Cel7B d'*H. insolens* qui sera toujours utilisée dans ce manuscrit pour les comparaisons entre les séquences), dans l'ensemble de la famille GH7, nous avons effectué des alignements de séquences de protéines appartenant à cette famille.

III.1.3.1 Principe

Les séquences de la famille GH7 sont alignées grâce au programme ClustalW [216]. Il s'agit d'un programme d'alignement multiple global qui utilise une approche heuristique d'alignement. ClustalW commence par générer tous les alignements par paire et construit pour chaque alignement une matrice de similarité. Chaque matrice est convertie en distances, lesquelles vont constituer un tableau ou arbre des relations évolutives. L'alignement va être construit en partant des deux séquences les plus proches puis va remonter vers les séquences les plus distantes. Cette méthode s'avère très efficace et rapide dans le cas d'un grand nombre de séquences relativement proches, ce qui est le cas des glycoside hydrolases appartenant à une même famille.

III.1.3.2 Résultats

La **Figure 27** montre une partie de l'alignement multiple des séquences peptidiques obtenu par le programme ClustalW à partir d'une cinquantaine de séquences d'enzymes de la famille GH7 (issues de CAZy, version Janvier 2002). Entre parenthèses, pour chaque enzyme est indiqué le numéro d'accession SwissProt. Sur la **Figure 27** n'est représentée que la partie de la séquence comprenant les résidus catalytiques.

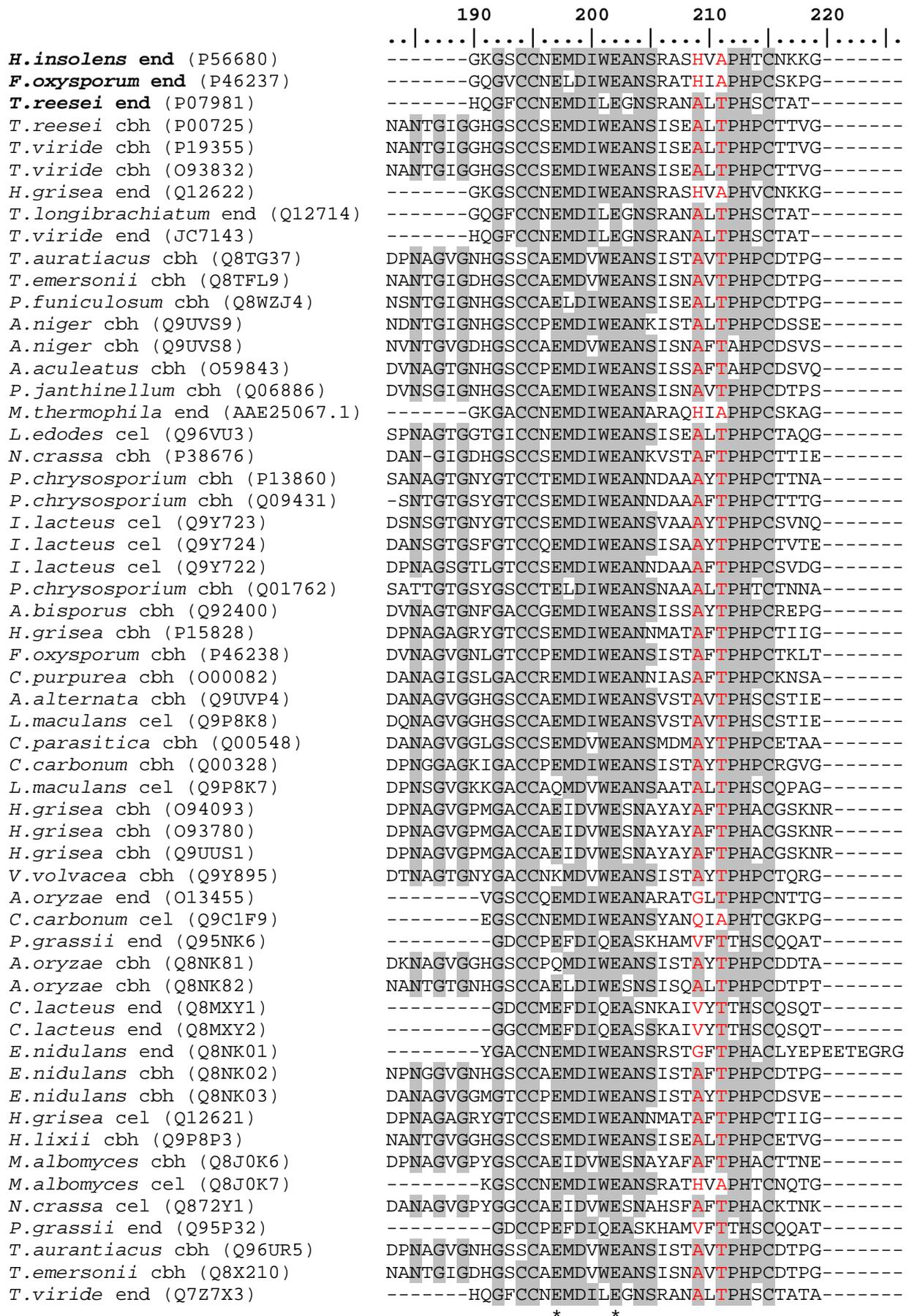


Figure 27 : Alignement de séquences d'enzymes de la famille GH7 des glycoside hydrolases.

En gris, les résidus identiques tout au long de l'alignement; marqués d'un astérisque, les deux résidus catalytiques; en rouge, les résidus 209 et 211 mis en évidence. La numérotation adoptée correspond à la numérotation de Cel7B d'*H. insolens*. Abréviations : end, endoglucanase; cbh, cellobiohydrolase; cel, cellulase.

Cet alignement a pu ainsi montrer que la séquence peptidique varie peu à proximité des résidus catalytiques 197 et 202. Le résidu 209, préalablement mis en évidence lors de l'étude qualitative, est dans la majeure partie des protéines, une histidine ou une alanine. De même, au niveau du résidu 211, on remarque la présence d'une alanine ou d'une thréonine.

En étudiant plus précisément cet alignement de séquences, on peut noter que le couple de résidus 209/211 se présente soit sous la forme d'une paire histidine/alanine (H/A) soit sous la forme d'une paire alanine/thréonine (A/T). Le couple A/T est toutefois prépondérant dans les enzymes de la famille GH7. Dans le cas de l'enzyme Cel7B d'*Humicola insolens* (première séquence de l'alignement présentée en **Figure 27**), sa séquence peptidique présente la paire H/A.

Après avoir mis en évidence ces deux acides aminés, nous avons effectué des études d'arrimage moléculaire substrat-enzyme pour étudier le comportement d'un oligosaccharide porteur d'une unité *N*-acétylglucosaminyle dans le sous-site +1 d'une enzyme mutée au niveau de ces deux résidus.

III.1.4 Etudes d'arrimage moléculaire

L'arrimage moléculaire, autrement appelé "docking", est une technique de prédiction du positionnement d'un ligand dans le site de fixation d'une macromolécule. Il est le plus souvent accompagné d'un champ de force simplifié permettant de prédire l'énergie d'interaction entre le ligand et la protéine, afin de pouvoir classer les différents résultats. Des études de docking peuvent ainsi permettre de confirmer ou prédire le positionnement d'oligosaccharides dans le centre actif d'enzymes. Des aspects structuraux de la sélectivité de la glucoamylase ont pu notamment être compris, suite à l'étude de l'interaction de di- et trisaccharides dans son centre actif [217-219].

Le positionnement du ligand dans la poche de fixation est un problème particulièrement complexe, vu le nombre de degrés de liberté mis en jeu : 3 pour la translation, 3 pour la rotation, plus la liberté conformationnelle des deux structures impliquées. Il est donc nécessaire de recourir à des approximations pour garder un compromis acceptable entre le temps de calcul et l'exhaustivité de l'exploration. Parmi les simplifications classiques, on peut citer la description purement atomique des sites d'interaction de la protéine

ainsi que l'élimination de toute flexibilité de celle-ci. Le développement de nouveaux algorithmes de recherche de l'espace conformationnel et l'augmentation de la puissance de calcul permet actuellement une exploration efficace de l'espace conformationnel des ligands pendant le docking, tout en utilisant des moyens de calcul modestes.

Dans notre cas, le ligand et la macromolécule sont, respectivement, le substrat et l'enzyme. Le site d'interaction est ici constitué par le site catalytique de l'enzyme.

Les études de docking réalisées dans cette thèse ont été effectuées grâce au programme AutoDock version 3.05 [220] dont les grands principes sont exposés dans le paragraphe suivant.

III.1.4.1 Principe d'utilisation d'AutoDock

Le programme AutoDock a été développé dans le but de fournir une procédure automatique pour la prédiction des interactions de ligands avec des biomacromolécules cibles. Il s'agit d'une suite d'exécutables et de scripts permettant d'obtenir des solutions de docking d'un ligand flexible dans une cible rigide. Ce programme est particulièrement adapté à l'étude de petits ligands de nature flexible en interaction avec des protéines.

Le principe d'utilisation de ce programme repose sur les étapes suivantes.

➤ *La préparation de la macromolécule*

Au fichier PDB original de la protéine, sont ajoutés les hydrogènes avec WHATIF [221] et des charges selon AMBER [222]. Les hydrogènes polaires sont différenciés des hydrogènes non-polaires pour rendre compte des ponts hydrogènes, en utilisant des paramètres Lennard-Jones différents selon le type d'hydrogène considéré.

➤ *La description de la macromolécule sous la forme de grilles*

Pour faciliter le positionnement du ligand dans la macromolécule, celle-ci est généralement décrite par un modèle constitué d'une grille tridimensionnelle de points associés

à l'énergie d'interaction d'une charge sonde et d'atomes sondes. Ce modèle définit l'étendue et les propriétés de la macromolécule. Les points sont ensuite utilisés pour calculer par interpolation l'énergie d'interaction intermoléculaire, permettant ensuite la correction itérative du positionnement des atomes du ligand.

Une zone (boîte) est donc définie à la périphérie de la macromolécule dans laquelle des solutions de docking pour un ligand donné sont recherchées. Le module AutoGrid prépare alors une matrice de points régulièrement espacés dans le volume de cette boîte. Pour chaque type d'atome présent dans le ligand, AutoGrid calcule ensuite l'énergie d'interaction (regroupant les énergies de type van der Waals et de type pont hydrogène) entre un atome sonde placé en un point de la matrice et l'ensemble de la macromolécule. Ainsi, si le ligand est constitué d'atomes de carbone, d'oxygène, d'hydrogène et d'azote, quatre grilles seront créées, une pour chaque type d'atome. AutoGrid calcule également le potentiel électrostatique correspondant à l'interaction d'une charge sonde de 1 eV en un point de la matrice avec l'ensemble de la macromolécule, conduisant à une grille supplémentaire.

➤ *La préparation du ligand*

Après l'addition des hydrogènes, il est nécessaire d'attribuer des charges partielles à la molécule, basées sur les charges définies dans PIM et celles calculées par MOPAC. Puis, afin de permettre la flexibilité du ligand, les liaisons au niveau desquelles des rotations seront appliquées doivent être désignées. Le module AutoTors permet ainsi de définir la partie rigide et la partie flexible de la molécule.

➤ *Les méthodes d'exploration de l'espace conformationnel du ligand*

AutoDock fait ensuite varier les différents degrés de liberté du ligand (axes de translation, de rotation et angles dièdres) à l'intérieur de la boîte. AutoDock propose plusieurs méthodes de recherche conformationnelle. La méthode originelle de ce programme est la technique Monte-Carlo appliquée avec recuit simulé. Il s'agit d'une des méthodes stochastiques la plus couramment utilisée. Avec celle-ci, la conformation et le positionnement du ligand sont explorés en même temps par perturbation aléatoire des variables définissant les degrés de liberté du complexe ligand-macromolécule.

Des algorithmes génétiques peuvent également être utilisés : ce sont des méthodes d'optimisation basées sur un hasard dirigé. Dans le programme AutoDock, deux types d'algorithmes génétiques ont été codés : l'algorithme génétique traditionnel (noté GA) et un algorithme génétique hybride – combinaison d'un optimiseur global et d'un optimiseur local – dit Lamarckien (noté LGA).

➤ *L'évaluation des résultats ("scoring")*

Les grilles préalablement préparées sont utilisées pour évaluer l'énergie d'un conformère du ligand à une position donnée dans la macromolécule. Cette méthode permet de calculer très rapidement l'énergie d'interaction ligand-macromolécule car les potentiels sont précalculés une fois pour toutes.

Les énergies calculées ainsi que les distances de la solution par rapport à l'origine permettent de trier et d'analyser les différents résultats obtenus. Les meilleures solutions sont finalement visualisées.

III.1.4.1.1 Oligosaccharides modélisés

Dans un premier temps, et afin de valider la méthode de docking utilisée, nous avons utilisé des cello-oligosaccharides préparés à partir d'un cellononaose modèle (**Figure 28**) issu du fichier PDB 8CEL.

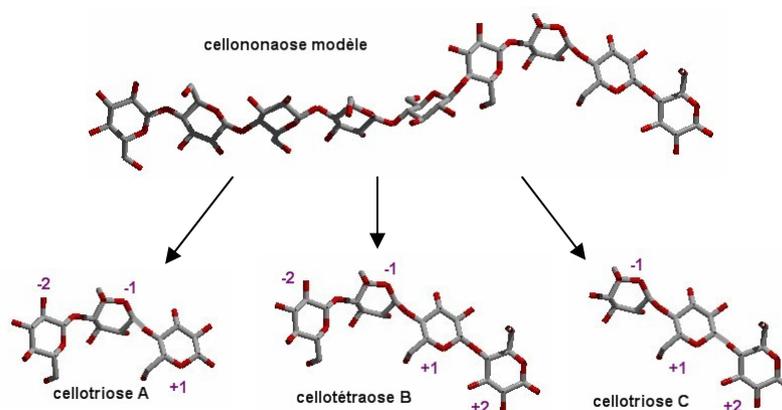


Figure 28 : Cello-oligosaccharides modélisés à partir du cellononaose.

De la même manière, à partir du cellononaose, deux oligosaccharides contenant un résidu *N*-acétylglucosaminyle ont été préparés puis modélisés dans le site actif de l'enzyme. Toutefois, par rapport aux cello-oligosaccharides utilisés dans un premier temps, il est nécessaire de prendre en compte les différentes conformations possibles du groupement *N*-acétyl porté par le cycle (**Figure 29**).

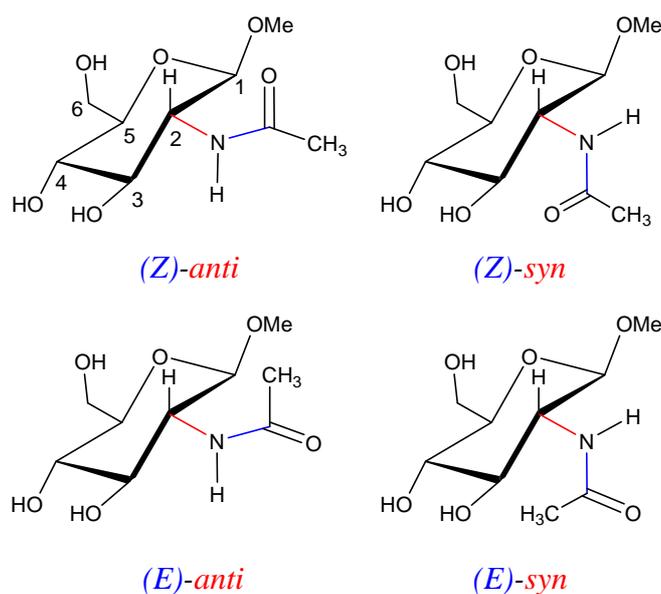


Figure 29 : Les différentes conformations* du groupe *N*-acétyl dans le *N*-acétylglucosaminide de méthyle [223].

* Dans la nomenclature utilisée, les conformations (*Z*) et (*E*) désignent les orientations décalées autour de la liaison C-N (relatives aux substituants prioritaires selon Cahn-Ingold-Prelog, l'oxygène du carbonyle et le résidu glucosyle) tandis que les conformations *syn* et *anti* désignent les orientations relatives des hydrogènes autour de la liaison C₂-N.

Deux types de trisaccharides contenant une unité *N*-acétylglucosamine dans le sous-site +1 de l'enzyme avec le groupe *N*-acétyl dans ses différentes conformations ont été modélisés. Les différentes structures sont résumées dans le **Tableau 11**.

Type de trisaccharide [sous-sites occupés]	Conformations du groupe N-acétyle	Notations
Glc(β1,4)GlcNAc(β1,4)Glc [-1,+2]	(Z)- <i>anti</i>	trisaccharide I
	(Z)- <i>syn</i>	trisaccharide II
	(E)- <i>anti</i>	trisaccharide III
	(E)- <i>syn</i>	trisaccharide IV
Glc(β1,4)Glc(β1,4)GlcNAc [-2,+1]	(Z)- <i>anti</i>	trisaccharide V
	(Z)- <i>syn</i>	trisaccharide VI
	(E)- <i>anti</i>	trisaccharide VII
	(E)- <i>syn</i>	trisaccharide VIII

Tableau 11 : Les différents trisaccharides modélisés dans le site actif de l'enzyme.

Dans notre étude, le cycle d'un monosaccharide constitue la partie rigide de la molécule tandis que le reste des liaisons de la molécule est flexible.

III.1.4.1.2 Protéines utilisées

Une première protéine utilisée est l'enzyme Cel7B d'*Humicola insolens*. La deuxième protéine utilisée lors du docking est dérivée de la précédente, obtenue par mutation au niveau de l'histidine 209 en alanine.

Les fichiers sources utilisés décrivant la protéine sont dérivés du fichier PDB 2A39 de la protéine native Cel7B d'*H. insolens*.

III.1.4.2 Les expériences de docking réalisées

Nous nous sommes donc intéressés au comportement de différentes structures oligosaccharidiques positionnées à l'intérieur du site actif de l'enzyme avec le logiciel AutoDock en prenant en compte leur flexibilité.

Les différentes expériences réalisées sont réunies dans le **Tableau 12**.

Entrée	Ligand	Macromolécule
1	cellotriose A	
2	cellotétraose B	2A39
3	cellotriose C	
4	trisaccharide I	
5	trisaccharide II	2A39
6	trisaccharide III	portant
7	trisaccharide IV	une mutation
8	trisaccharide V	supplémentaire
9	trisaccharide VI	H209A
10	trisaccharide VII	
11	trisaccharide VIII	

Tableau 12 : Les différentes expériences de docking réalisées.

Dans un premier temps, les expériences de docking réalisées avec les cello-oligosaccharides dans l'enzyme native (Entrées **1**, **2** et **3**) ont permis de tester les paramètres de docking employés. La méthode ainsi validée, le comportement des trisaccharides présentant une unité *N*-acétylglucosaminyle (Entrées **4** à **11**) dans l'enzyme mutée au niveau de l'histidine 209 a été analysé.

Dans chaque cas, la recherche dans l'espace conformationnel du ligand a été effectuée *via* l'algorithme génétique de type lamarckien (LGA).

III.1.4.3 Analyses des solutions de docking

III.1.4.3.1 Docking des cellotrioses

Afin de valider les paramètres de docking utilisés, nous avons donc réalisé différentes expériences dont les résultats obtenus sont exposés dans le **Tableau 13**.

Entrée	Substrat	Cluster [nombre de structures]	Energie totale ^{b,c} (kcal/mol)	Déviatoin RMS ^{d,e} (Å)
1	cellotriose A	1 [5/100]	-169.7 (-164.7)	1.36
2	cellotriose C	1 [2/100]	-158.6 (-154.5)	1.57
3		2 [3/100]	-155.0 (-152.0)	1.76

Tableau 13 : Docking des cellotrioses A et C dans le site actif de l'endoglucanase Cel7B d'*H. insolens*^a.

^a Les structures arrimées sont regroupées en clusters dont les meilleurs sont montrés;

^b Les énergies sont calculées par AutoDock;

^c La valeur d'énergie pour la structure optimale de chaque cluster est à gauche et la valeur moyenne pour le cluster est entre parenthèses;

^d Déviation de la moyenne quadratique par rapport à celle de la structure de référence;

^e Correspondant à la valeur d'énergie la plus basse du cluster.

Ce tableau montre l'énergie totale du substrat arrimé dans le site actif de l'enzyme. Seuls les meilleurs clusters obtenus sont présentés. Nous obtenons des énergies relativement équivalentes pour les deux types de cellotrioses modélisés, avec des coefficients de déviation RMS corrects. L'ensemble des résultats de docking obtenus pour le cellotriose A positionné dans les sous-sites -2 à +1 reste meilleur que ceux obtenus pour le cellotriose C positionné dans les sous-sites -1 à +2. Ceci laisse penser à une déstabilisation plus grande quand le sous-site +2 est occupé, ou encore à une stabilisation plus grande quand le sous-site -2 est occupé.

Les résultats correspondant au cellotétraose B ne sont pas montrés car le positionnement de l'oligosaccharide ne s'effectuait pas dans le site actif de l'enzyme mais en dehors de la crevasse.

La **Figure 30** montre une des solutions de docking obtenues selon différentes vues. On remarque bien l'arrimage du cellotriose A dans la crevasse de Cel7B d'*H. insolens*.

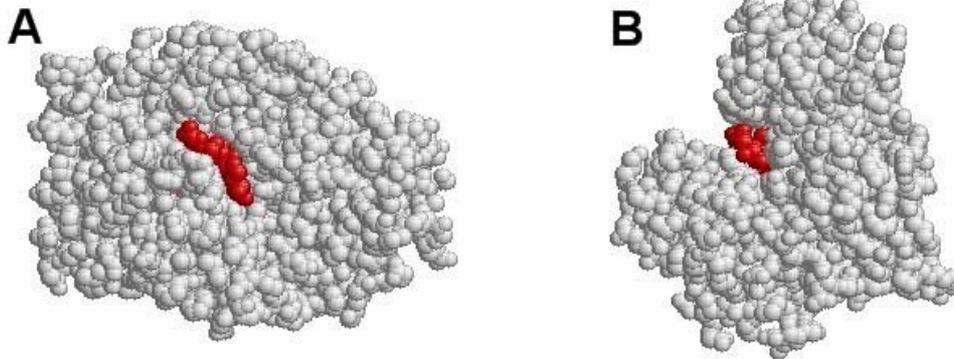


Figure 30 : Vues (A) et (B) d'une solution de docking du cellotriose A dans le site actif de l'enzyme Cel7B d'*H. insolens*.

En rouge, le cellotriose A.

Les résultats obtenus lors du docking des cellotrioses A et C laissent penser que les paramètres utilisés sont corrects et que l'étude du comportement des trisaccharides I à VIII dans le site actif de l'enzyme mutée Cel7B H209A d'*H. insolens* peut être réalisée.

III.1.4.3.2 Docking des trisaccharides

Les résultats obtenus au cours des différentes expériences de docking des trisaccharides I à VIII sont présentés dans le **Tableau 14**. Seuls les clusters les plus significatifs sont listés.

Entrée	Substrat	Cluster [nombre de structures]	Energie totale ^{b,c} (kcal/mol)	Déviaton RMS ^{d,e} (Å)
1	trisaccharide I	1 [1/100]	-145.4 (-145.4)	2.66
2		11 [1/100]	-130.2 (-130.2)	3.54
3	trisaccharide II	1 [1/100]	-151.3 (-151.3)	2.25
4		6 [1/100]	-136.4 (-136.4)	2.08
5	trisaccharide III	1 [1/100]	-150.0 (-150.0)	2.60
6		3 [1/100]	-142.6 (-142.6)	1.72
7	trisaccharide IV	1 [1/100]	-149.8 (-149.8)	2.01
8		3 [1/100]	-144.1 (-144.1)	1.94
9		4 [2/100]	-142.3 (-140.4)	1.83
10	trisaccharide V	1 [2/100]	-160.9 (-158.3)	1.55
11	trisaccharide VI	1 [1/100]	-139.6 (-139.6)	2.58
12	trisaccharide VII	1 [3/100]	-160.1 (-153.2)	1.53
13	trisaccharide VIII	1 [1/100]	-168.8 (-168.8)	1.47
14		5 [4/100]	-129.8 (-128.1)	3.43

Tableau 14 : Docking de trisaccharides I-VIII dans le site actif de l'endoglucanase mutée Cel7B H209A d'*H. insolens*^a.

^a Les structures arimées sont regroupées en clusters dont les meilleurs sont montrés;

^b Les énergies sont calculées par AutoDock;

^c La valeur d'énergie pour la structure optimale de chaque cluster est à gauche et la valeur moyenne pour le cluster est entre parenthèses;

^d Déviation de la moyenne quadratique par rapport à celle de la structure de référence;

^e Correspondant à la valeur d'énergie la plus basse du cluster.

Dans le cas des trisaccharides de type Glc(β 1,4)GlcNAc(β 1,4)Glc (trisaccharides I à IV) occupant les sous-sites -1 à +2 du site de catalyse, les meilleures énergies obtenues ainsi que les meilleurs coefficients de déviation RMS sont relativement élevés. Les oligosaccharides ne sont donc pas correctement positionnés dans le site actif de l'enzyme, voire même sortis de la crevasse. Ce résultat est d'ailleurs confirmé par la visualisation en trois dimensions.

Les solutions de docking obtenues pour les trisaccharides de type Glc(β 1,4)Glc(β 1,4)GlcNAc (trisaccharides V à VIII) occupant les sous-sites -2 à +1 sont meilleures que les précédentes, comme c'était le cas pour le cellotriose A dans le paragraphe précédent.

La **Figure 31** montre une des solutions de docking obtenues pour le trisaccharide VII dans le site actif de l'enzyme mutée Cel7B H209 d'*H. insolens*.

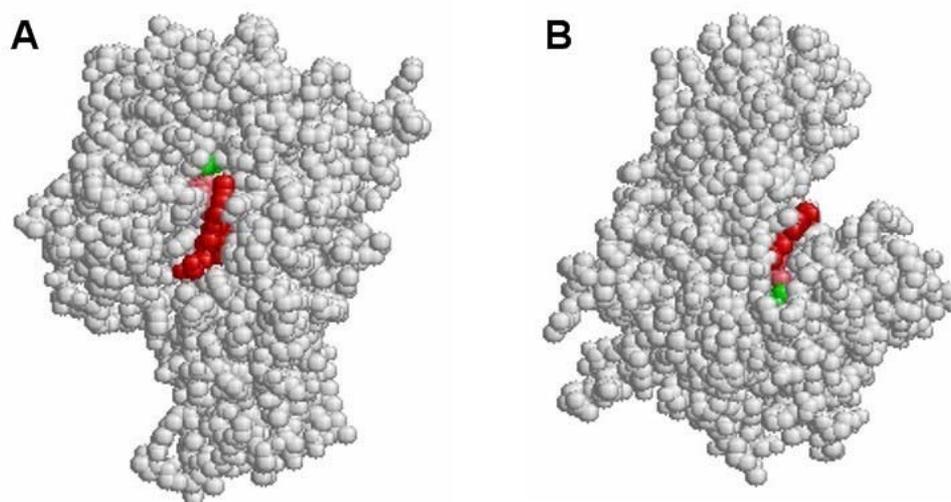


Figure 31 : Vues (A) et (B) d'une solution de docking du trisaccharide VII dans le site actif de l'enzyme Cel7B H209A d'*H. insolens*.

En rouge, le trisaccharide VII; en rose, les résidus catalytiques de l'enzyme; en vert, le résidu 209 de l'enzyme.

On peut ainsi observer l'arrimage du trisaccharide dans le site actif de l'enzyme, laissant penser que la mutation est compatible avec la fixation de l'oligosaccharide comme substrat pour l'hydrolyse ou comme produit d'une réaction de couplage enzymatique.

Cette même solution de docking a été dans un premier temps placée dans le site actif de l'enzyme mutée dans laquelle elle a été positionnée par AutoDock. Dans un deuxième temps, cette solution a été superposée dans le site actif de l'enzyme native Cel7B d'*H. insolens*, ne présentant pas la mutation au niveau de l'histidine 209. La **Figure 32** montre les agrandissements de la zone autour du substrat trisaccharidique dans la protéine mutée H209A (**Figure 32A**) et dans la protéine native (**Figure 32B**).

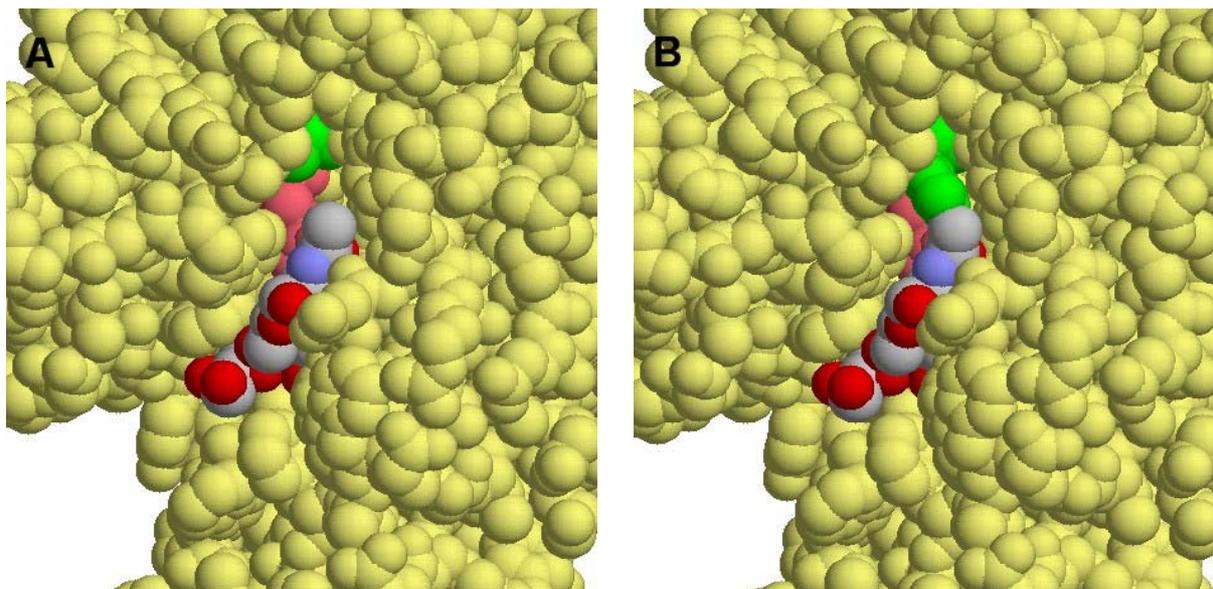


Figure 32 : (A) Agrandissement de la solution de docking du trisaccharide VII dans le site actif de l'enzyme mutée Cel7B H209A d'*H. insolens*; (B) Superposition de la même solution de docking dans le site actif de l'enzyme native Cel7B d'*H. insolens*.

En couleurs CPK, le trisaccharide VII; en rose, les résidus catalytiques; en vert, le résidu 209 (une alanine en (A) ou une histidine en (B)).

La comparaison entre les deux complexes montre la possibilité d'un clash stérique entre la chaîne latérale de l'histidine 209 et le groupement *N*-acétyle (**Figure 32B**) qui n'existe pas lorsqu'une alanine dont la chaîne latérale est moins encombrante remplace l'histidine. La mutation de l'histidine en alanine devrait créer une poche suffisante pour permettre la fixation de la chaîne *N*-acétyle dans le sous-site +1 du site actif de l'enzyme.

III.1.5 Résumé

Les différents résultats obtenus laissent penser qu'une mutation au niveau du résidu 209 de l'endoglucanase Cel7B d'*Humicola insolens* pourrait modifier la spécificité de substrat de l'enzyme et permettre ainsi la fixation d'un résidu *N*-acétyle dans le sous-site +1 du site catalytique par la création d'une poche. Cette dernière peut être obtenue par la mutation de l'histidine 209 présente dans l'enzyme native en un acide aminé dont la chaîne latérale est moins encombrante (**Figure 33**).

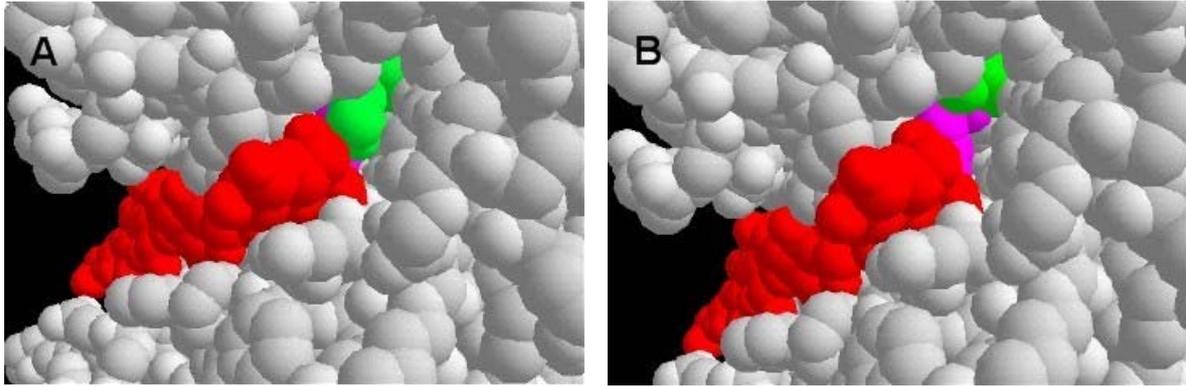


Figure 33 : Cellononaose dans le site actif (A) de l'enzyme Cel7B d'*H. insolens* et (B) du mutant Cel7B H209A d'*H. insolens*.

En rouge, le cellononaose; en violet, les résidus catalytiques de l'enzyme; en vert, le résidu 209 de l'enzyme (une histidine en (A) ou une alanine en (B)).

Plusieurs études dans la littérature montrent que la création d'une poche suffisamment grande par mutagenèse dirigée peut permettre la fixation de la chaîne *N*-acétyle, plus encombrante qu'un simple hydroxyle. On peut citer les travaux de Ramakrishnan et coll. en 2002 [224] sur la β 1,4-galactosyltransférase I, dont l'activité *N*-acétylgalactosaminyltransférase a pu être augmentée par la mutation du résidu Tyr-289 impliqué dans la formation d'un pont hydrogène avec la chaîne *N*-acétyle.

Les travaux de Marcus et coll. en 2003 [225] ont également montré qu'il était possible de modifier la spécificité de donneur de l' α 1,3-galactosyltransférase (GTB) participant à la formation de l'antigène du groupe B vis-à-vis du donneur UDP-GalNac par la mutation du résidu proline en position 234 en une sérine.

Ainsi, dans un premier temps, un mutant intéressant à produire serait porteur d'une mutation de l'histidine 209 en alanine. La mutation de ce résidu est déjà présente au sein de la famille GH7, laissant penser que cela ne modifierait pas sensiblement l'activité de l'enzyme.

Dans un deuxième temps, la création d'une poche encore plus grande pour la fixation de la chaîne peut être réalisée par la mutation de ce même résidu 209 en glycine, la chaîne latérale étant ainsi réduite à son minimum.

Finalement, l'alignement de séquences des enzymes de la famille GH7 a pu mettre en évidence un couple d'acides aminés : les résidus 209 et 211. Il peut paraître donc nécessaire de muter ses deux résidus, l'histidine 209 en alanine et l'alanine 211 en thréonine.

III.2 Etude de l'hydrolyse du tétrasaccharide Gal(β 1,4)Glc(β 1,4)GlcNAc(β 1,4)GlcNAc 34

Avant de débiter la mutagenèse dirigée au niveau des acides aminés précédemment mis en relief, nous avons voulu valider les études de modélisation en comparant l'activité d'endoglucanases de la famille GH7 vis-à-vis du substrat tétrasaccharidique Gal(β 1,4)Glc(β 1,4)GlcNAc(β 1,4)GlcNAc. La structure de cet oligosaccharide est conçue de telle manière qu'il ne peut *a priori* se positionner que d'une seule façon dans le centre actif de l'enzyme (**Figure 34**).

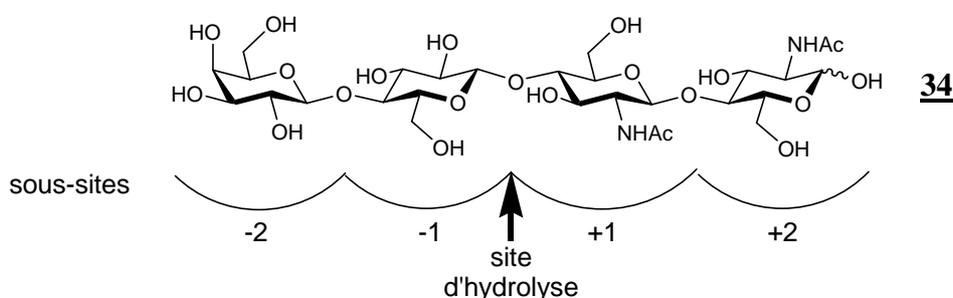


Figure 34 : Positionnement du tétrasaccharidique **34** dans le site actif de l'endoglucanase.

Ce substrat permet ainsi de positionner une unité *N*-acétylglucosaminyle dans le sous-site +1 de l'enzyme et donc d'étudier l'activité d'hydrolyse d'endoglucanases vis-à-vis de ce substrat. Dans le cadre de notre étude, nous avons cherché à comparer l'activité de deux enzymes en particulier. D'un côté, l'endoglucanase Cel7B d'*Humicola insolens* présente le couple histidine/alanine en position 209/211 tandis que d'un autre côté, l'endoglucanase Cel7B de *Trichoderma reesei* possède la paire alanine/thréonine. Il était donc intéressant d'étudier l'activité d'hydrolyse de cette dernière vis-à-vis du composé **34**, afin de confirmer nos études de docking.

III.2.1 Synthèse du substrat **34**

La stratégie envisagée pour l'obtention du tétrasaccharide **34** est présentée sur le **Schéma 12**.

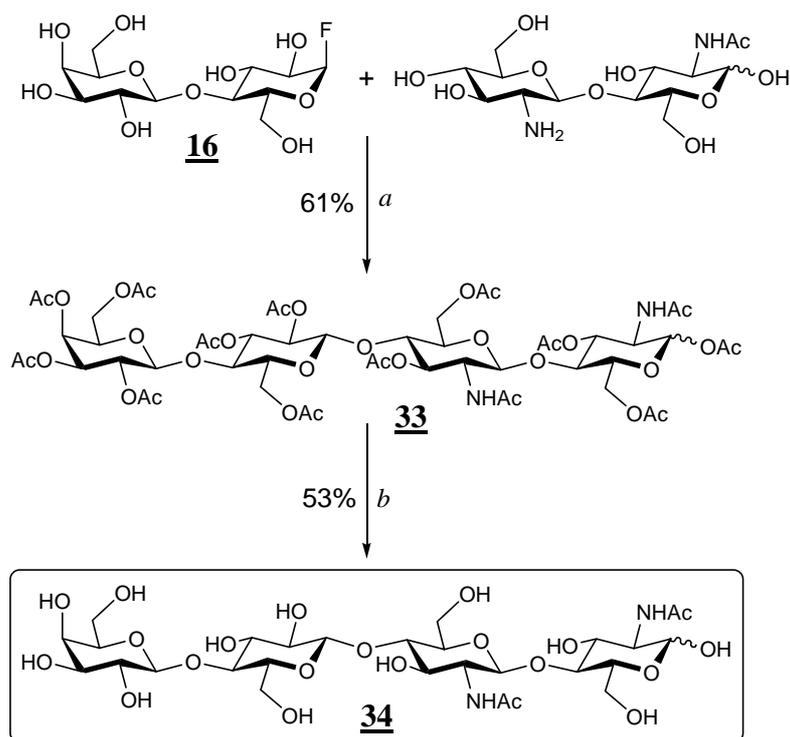


Schéma 12 : Synthèse du composé **34**.

Conditions : (a) i. Cel7B E197A *H. insolens*, tampon carbonate-bicarbonate de sodium 50 mM pH 9.2;
ii. DMAP, Ac₂O, pyridine; (b) MeONa, MeOH.

Il s'agit dans une première étape d'une réaction de couplage enzymatique entre le donneur de lactosyle **16** et le *N*¹-acétylchitobiose disponible au laboratoire. Il est effectué en présence de la glycosynthase Cel7B E197A d'*H. insolens*, le *N*¹-acétylchitobiose pouvant se fixer dans les sous-sites accepteurs (cf. paragraphe **1.2.4**). Après lyophilisation, le brut réactionnel est *N*- et *O*-acétylé dans un mélange d'anhydride acétique et de pyridine. Après purification, le produit **33** est *O*-désacétylé par le méthylate de sodium dans le méthanol [196] pour conduire à la formation du produit **34**.

III.2.2 Hydrolyse du substrat

L'hydrolyse du substrat **34** a été réalisée par deux endoglucanases : l'endoglucanase Cel7B d'*Humicola insolens* et l'endoglucanase Cel7B de *Trichoderma reesei* (Schéma 13).

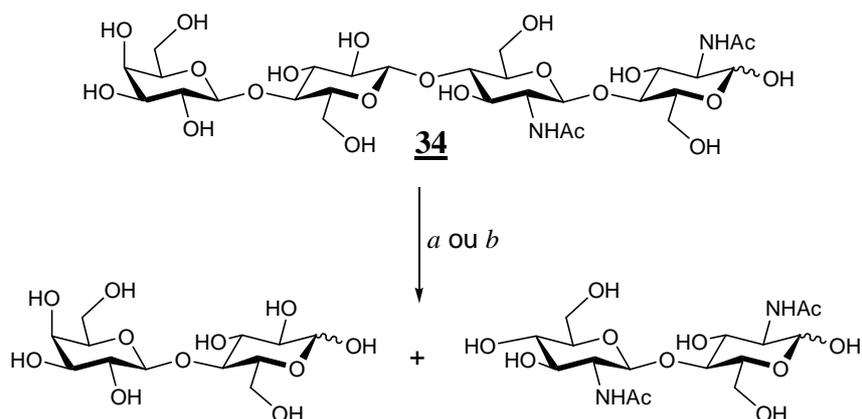


Schéma 13 : Hydrolyse enzymatique du tétrasaccharide 34.

Conditions : (a) Cel7B *H. insolens*, tampon phosphate de sodium 50 mM pH 7;
(b) Cel7B *T. reesei*, tampon acétate de sodium 25 mM pH 4.6.

Le gel SDS-PAGE des deux enzymes est présenté dans la **Figure 35**. La réaction d'hydrolyse a été effectuée au pH optimum de chaque enzyme.

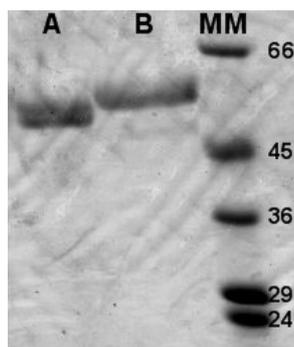


Figure 35 : Analyse des endoglucanases Cel7B d'*H. insolens* et de *T. reesei* sur gel de polyacrylamide à 10% en conditions dénaturantes.

Coloration des protéines au bleu de Coomassie.

Puits **A** : Cel7B d'*H. insolens*; puits **B** : Cel7B de *T. reesei*;
puits **MM** : marqueur de masses moléculaires (indiquées en kDa).

Une étude par HPLC a permis de quantifier l'activité hydrolytique des deux enzymes vis-à-vis de ce tétrasaccharide 34. Les différents chromatogrammes du milieu réactionnel à différents temps de réaction mettent en évidence la disparition du tétrasaccharide 34 et l'apparition de deux pics correspondant aux deux disaccharides formés au cours de l'hydrolyse (**Figure 36**).

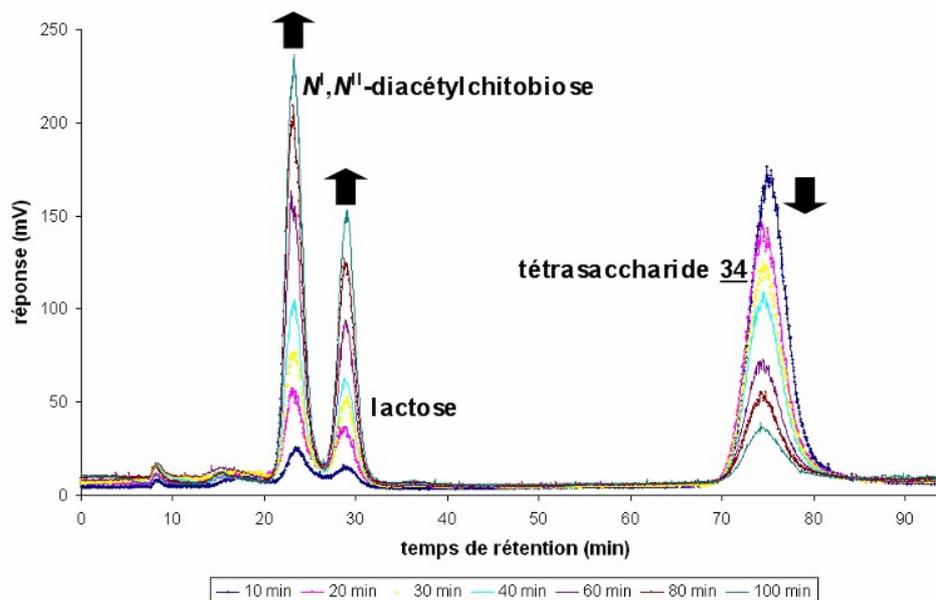


Figure 36 : Analyse par HPLC de la réaction d'hydrolyse du tétrasaccharide **34** par une endoglucanase.

Les mesures de l'aire du pic correspondant au tétrasaccharide **34** ou de l'aire du pic correspondant au produit d'hydrolyse N^I, N^{II} -diacétylchitobiose au cours du temps permettent de déterminer l'activité hydrolytique pour les deux enzymes vis-à-vis de ce substrat.

En se plaçant en condition de vitesse initiale d'hydrolyse, nous avons calculé l'activité spécifique de chacune des enzymes, exprimée en micromoles de substrat **34** hydrolysé par minute et par milligramme d'enzyme (Figure 37).

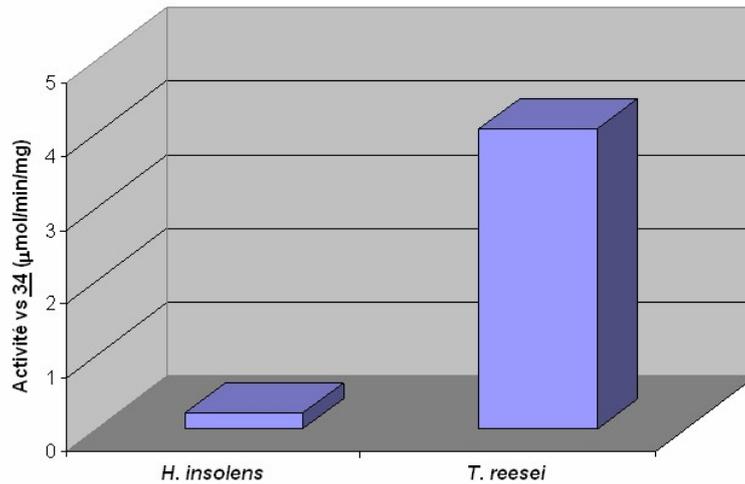


Figure 37 : Mesure de l'activité hydrolytique des endoglucanases Cel7B vis-à-vis du tétrasaccharide 34.

La Figure 37 met ainsi en évidence une plus grande activité hydrolytique de la cellulase Cel7B de *Trichoderma reesei* vis-à-vis du substrat 34, presque vingt fois plus importante que celle de la cellulase Cel7B de *Humicola insolens*.

Nous avons comparé l'activité d'hydrolyse de ces deux enzymes vis-à-vis du cellotétraose. La Figure 38 montre que l'enzyme Cel7B de *T. reesei* est environ trois fois plus active sur ce substrat que l'enzyme Cel7B d'*H. insolens*.

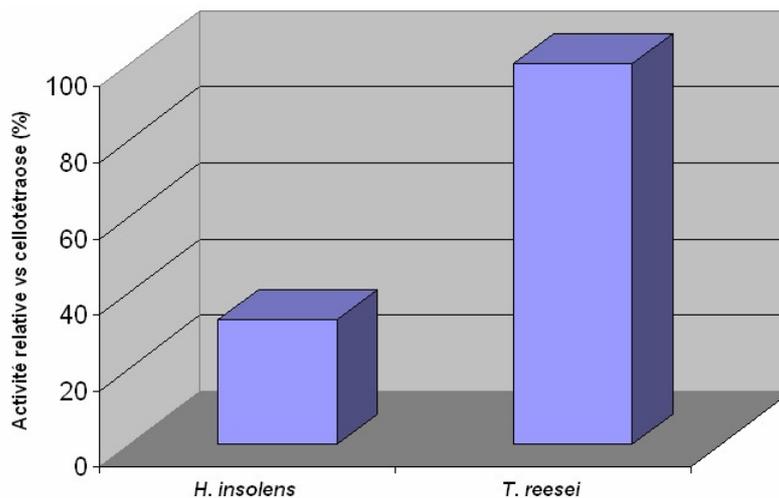


Figure 38 : Activité relative d'hydrolyse entre les deux endoglucanases Cel7B vis-à-vis du cellotétraose.

Les différences d'activités observées sur le tétrasaccharide 34 sont donc significatives.

III.3 Etude de l'activité de transglycosylation de Cel7B de *Trichoderma reesei*

Compte-tenu des résultats obtenus lors de l'hydrolyse du tétrasaccharide **34** par l'enzyme Cel7B de *Trichoderma reesei*, nous avons voulu tester l'activité de transglycosylation de cette dernière pour le couplage entre un fluorure de β -lactosyle **36** et le N^I, N^{II} -diacétylchitobiose pour la formation du tétrasaccharide **34** (Schéma 14).

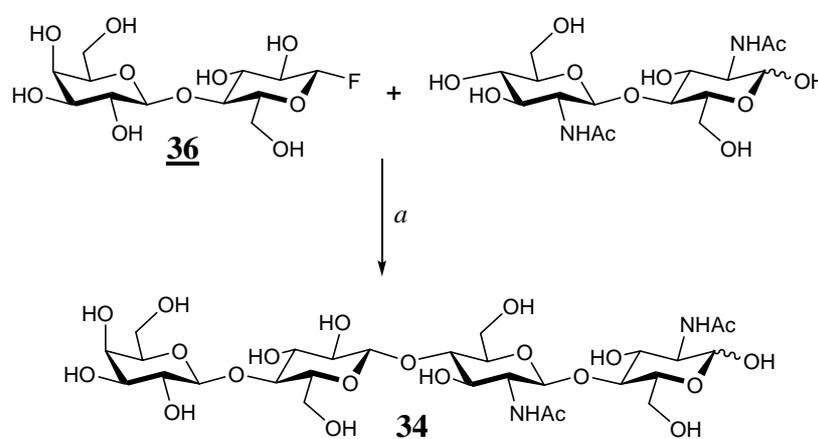


Schéma 14 : Couplage enzymatique entre le donneur **36** et le N^I, N^{II} -diacétylchitobiose.

Conditions : (a) Cel7B *T. reesei*, tampon acétate de sodium 25 mM pH 4.6/MeCN (2:3, v/v).

Comme le montre le Schéma 15, le donneur de glycosyle **36** est obtenu en deux étapes à partir du lactose per-*O*-acétylé **21**, précédemment préparé (Schéma 9).

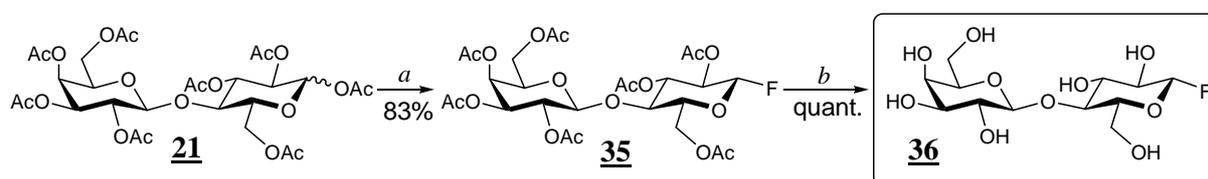


Schéma 15 : Préparation du fluorure de β -lactosyle **36**.

Conditions : (a) i. HBr/AcOH, CH_2Cl_2 ; ii. AgF, CH_3CN ; (b) MeONa, MeOH.

Le lactose peracétylé **21** est modifié en β -fluorure **35** par une voie de synthèse classique [226] : l'intermédiaire bromure d' α -glycosyle formé est directement converti en

fluorure de β -glycosyle *via* une substitution nucléophile d'halogénure par traitement avec le fluorure d'argent. Etant donné l'instabilité des fluorures de β -glycosyle libres, le disaccharide **36** est désacétylé selon la procédure classique [196] juste avant les tests enzymatiques pour conduire au donneur de lactosyle **36**.

Le couplage du donneur **36** avec le N^I, N^{II} -diacétylchitobiose a donc été testé en présence de Cel7B de *Trichoderma reesei* dans un milieu hydro-organique (mélange d'acétonitrile et d'eau) pour minimiser la réaction d'hydrolyse, concurrente de la réaction de transglycosylation. L'analyse par spectrométrie de masse du milieu réactionnel après concentration et lyophilisation a montré la formation majoritaire du tétrasaccharide bis-lactose (issu d'un couplage C1:C1 entre deux unités lactosyles). Cependant, nous avons également observé la présence dans le milieu du produit **34** avec un pic à $m/z = 771$ correspondant à l'ion $[M+Na]^+$. Le N^I, N^{II} -diacétylchitobiose pourrait donc être utilisé en tant qu'accepteur par l'enzyme Cel7B de *T. reesei* et la mutation de l'acide aminé nucléophile pourrait en faire une glycosynthase susceptible d'être utilisée dans le remodelage de la glycosylation des protéines.

Les différents résultats obtenus avec l'enzyme Cel7B de *Trichoderma reesei*, avec d'une part son activité d'hydrolyse sur le tétrasaccharide **34** et d'autre part son activité de transglycosylation conduisant à la synthèse de ce même tétrasaccharide **34**, permettent ainsi de valider les observations effectuées au cours des études de docking et de confirmer la pertinence des mutations suggérées précédemment.

III.4 Mutagenèse dirigée

III.4.1 Préparation des mutants de Cel7B E197A d'*H. insolens*

Compte-tenu des résultats obtenus précédemment, la préparation de trois mutants de Cel7B E197A d'*Humicola insolens* a été retenue (**Tableau 15**).

Mutations réalisées sur Cel7B E197A d'*H. insolens*

H209A

H209G

H209A A211T

Tableau 15 : Liste des différentes mutations réalisées sur Cel7B d'*H. insolens*.

Dans un premier temps, nous avons tenté de muter uniquement le résidu 209 qui est une histidine dans l'enzyme native Cel7B d'*H. insolens*. Cette histidine a été substituée soit par une alanine, soit par une glycine, avec dans l'idée de créer une poche plus importante en supprimant totalement la chaîne latérale.

Dans un deuxième temps, nous avons réalisé une mutation supplémentaire au niveau du résidu 211 dans le but d'obtenir le triple mutant Cel7B E197A H209A A211T.

Les plasmides contenant les gènes codant pour Cel7B (noté pHW704) et pour Cel7B E197A (noté pCE 10) d'*H. insolens* nous ont été fournis par le Dr Jesper Vind (Novozymes, Danemark). Il s'agit de plasmides du type pCR4-TOPO.

Après sous-clonage du fragment *EcoRV/SacI* du gène de Cel7B d'*H. insolens*, contenant les codons des résidus 209 et 211 de l'enzyme, dans le plasmide pBlueScript, les différentes mutations ont été introduites par PCR. Le séquençage des plasmides obtenus après mutagenèse dirigée a permis de vérifier la présence des mutations envisagées ainsi que la non-introduction de mutations indésirables.

Les fragments *EcoRV/SacI* présentant les différentes mutations désirées ont ensuite été replacés dans un vecteur approprié.

III.4.2 Préparation de la glycosynthase de Cel7B de *T. reesei*

Etant donné les résultats obtenus pour l'hydrolyse du tétrasaccharide **34** par la cellulase de *Trichoderma reesei* ainsi que de son activité de transglycosylation, il a été envisagé de préparer le mutant non-nucléophile de l'enzyme, c'est-à-dire la glycosynthase Cel7B E196A.

Le plasmide noté pMP311 contenant l'ADNc de Cel7B de *Trichoderma reesei* nous a été donné par le Dr Anu Koivula (VTT, Finlande). Il s'agit en fait d'un vecteur navette type

pMA91 [227] utilisé pour le passage entre la bactérie *Escherichia coli* (organisme de choix pour les différents clonages) et la levure *Saccharomyces cerevisiae* (organisme choisi pour l'expression de la protéine). Les gènes hétérologues clonés dans ce plasmide sont exprimés sous le contrôle du promoteur fort de la phosphoglycérate kinase de levure.

Une première stratégie envisagée pour la préparation du mutant glycosynthase Cel7B E196A de *T. reesei* consistait à sous-cloner un fragment du gène dans un plasmide de type pBlueScript, dans le but d'éviter de séquencer la totalité du gène après mutagenèse dirigée et de travailler avec un plasmide de 11.3 kb. Toutefois, le gène de Cel7B de *T. reesei* présente l'inconvénient d'avoir peu de sites de restriction utilisables pour les clonages et peu d'informations sont disponibles sur le vecteur pMA91.

Après différents essais de sous-clonages, cette stratégie a été abandonnée.

Dans une deuxième stratégie, nous avons décidé de tenter de réaliser la mutagenèse dirigée sur le plasmide pMP311 complet.

Les amorces dessinées ont permis de muter le résidu nucléophile de Cel7B de *T. reesei* en alanine et dans un même temps d'introduire un site de restriction *NcoI* sans modifier la séquence d'acides aminés, pour faciliter le criblage des plasmides obtenus après mutagenèse dirigée effectuée par PCR.

Le séquençage a permis de montrer que la mutation du nucléophile avait été réalisée et qu'aucune autre mutation supplémentaire indésirable n'avait été introduite au cours de la PCR dans le gène de Cel7B de *T. reesei*.

III.5 Conclusion

Les différentes études entreprises dans ce chapitre nous ont permis d'envisager la modification de la spécificité de la glycosynthase Cel7B E197A d'*Humicola insolens* par mutagenèse dirigée. En effet, les études de docking ont montré que la mutation de l'acide aminé en position 209 de l'enzyme pouvait créer une poche suffisante pour la fixation d'une chaîne *N*-acétyle dans le site actif de l'enzyme, dirigée vers l'intérieur de la crevasse.

Ces résultats ont été confirmés par l'étude de la réaction d'hydrolyse du tétrasaccharide **34** par deux enzymes présentant des acides aminés différents au niveau des résidus mis en évidence. Les travaux de mutagenèse dirigée ont donc été réalisés par PCR.

Le chapitre suivant est consacré à la production des différents mutants d'*Humicola insolens* et de *Trichoderma reesei* obtenus, suivie de la caractérisation de leur activité enzymatique vis-à-vis de substrats naturels ou non-naturels.

Chapitre IV

**PRODUCTION ET CARACTERISATION
DES ENZYMES MUTANTES**

IV PRODUCTION ET CARACTERISATION DES ENZYMES MUTANTES

IV.1 Production des enzymes

IV.1.1 Tentatives de surexpression de Cel7B d'*H. insolens* chez *E. coli* et *B. subtilis*

Des tentatives précédentes d'expression de la protéine Cel7B d'*H. insolens* chez *E. coli* ont été réalisées par Novozymes (Danemark) et n'ont pas permis de l'obtenir dans des quantités jugées suffisantes du point de vue d'un industriel. En effet, cette protéine contient neuf ponts disulfures et forme des corps d'inclusion lorsqu'elle est produite chez la bactérie car ce système d'expression ne possède pas un environnement propice à la formation des ponts disulfures [228].

Récemment, des moyens d'expression ont été développés chez *E. coli* pour permettre leur formation, par exemple par un adressage de la protéine dans le périplasme de la cellule, qui est un milieu plus oxydant et donc plus propice à la formation de ces ponts disulfures [228,229].

Puisque nos études préliminaires ne nécessitaient pas de disposer de quantités importantes d'enzyme, nous avons tenté d'exprimer Cel7B d'*H. insolens* chez deux bactéries disponibles au laboratoire : *E. coli* et *B. subtilis*. Nous avons dans un premier temps travaillé avec l'enzyme sauvage car son activité est plus facilement détectable que celle de la glycosynthase. La détection peut en effet être réalisée par des mesures d'activités cellulases classiques, par exemple *via* l'hydrolyse de la CMC mise en évidence par le dosage des sucres réducteurs (test au DNS) ou encore l'hydrolyse d'un hexasaccharide fluorescent suivie par la méthode du FRET [146].

IV.1.1.1 Expression de *Cel7B* chez *Escherichia coli*

Le plasmide noté pHW704 contenant le gène codant pour *Cel7B* d'*Humicola insolens* nous a été donné par le Dr Jesper Vind (Novozymes, Danemark). Après l'insertion d'un site *NotI* en 3', le gène a pu être cloné dans deux vecteurs pET différents : le plasmide pET-20b(+) et le plasmide pET-32a(+). Le système pET est le système le plus performant développé jusqu'à maintenant pour l'expression de protéines recombinantes chez *E. coli*. Il est basé sur la sélectivité et la puissance de l'enzyme ARN polymérase du bactériophage T7 (**Figure 39**).

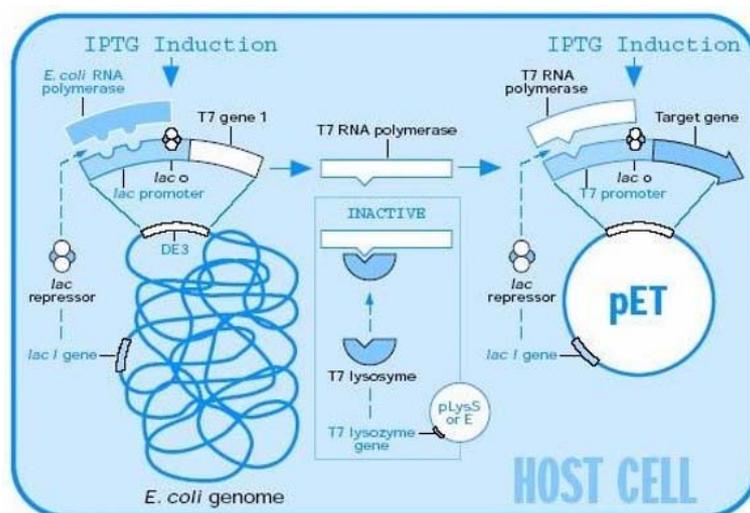


Figure 39 : Les éléments de contrôle du système pET.

Les plasmides utilisés ici permettent notamment de fusionner la protéine exprimée à une étiquette C-terminale polyhistidine (His-Tag), facilitant ainsi sa purification sur une colonne d'affinité utilisant le nickel. Dans le cas du plasmide pET-20b(+), le vecteur porte en amont du gène codant pour la protéine d'intérêt une séquence *pelB* codant pour un peptide signal qui permet d'adresser la protéine dans le périplasme de la cellule, connu pour être un milieu plus propice à la formation des ponts disulfures que le cytoplasme.

Le plasmide pET-32a(+) possède quant à lui une séquence codant pour une protéine de fusion, la thiorédoxine, permettant d'augmenter la solubilité de la protéine d'intérêt en favorisant la formation des ponts disulfures dans le cytoplasme bactérien.

L'IPTG permet d'induire la transcription du promoteur T7 présent dans les vecteurs pET, promoteur sous le contrôle duquel se trouve l'expression de la protéine cible.

Les différentes souches d'*E. coli* utilisées contiennent le prophage λ DE3 dont l'un des gènes code pour l'ARN polymérase du bactériophage T7. Ce gène est sous le contrôle du promoteur lacUV5 inductible par l'IPTG. Ainsi, quand ces souches sont transformées par un vecteur d'expression (comme ceux de la famille pET) qui contient un promoteur fort (celui du bactériophage T7), la synthèse d'ARNm est accrue et la synthèse de la protéine d'intérêt peut atteindre une proportion de 40% à 50% des protéines totales.

Le **Tableau 16** résume les différents systèmes utilisés pour exprimer l'enzyme Cel7B d'*H. insolens* chez la bactérie *Escherichia coli*.

Parmi les différentes souches d'*E. coli* utilisées, la souche Origami (DE3) est caractérisée par la présence dans son génome de mutations au niveau des gènes *trxB* et *gor* codant pour la thiorédoxine réductase et la glutathione réductase respectivement, ce qui minimise la réduction des ponts disulfures dans le cytoplasme.

La souche BL21 (DE3) est quant à elle la plus utilisée pour l'expression des protéines chez *E. coli*. Elle présente des mutations au niveau de gènes codant pour des protéases. La souche BL21 (DE3) pLysS possède en plus le plasmide pLys qui code pour une protéine bifonctionnelle, le lysozyme. Cette protéine se lie d'une part, à l'ARN polymérase et inhibe la transcription. Ainsi, le plasmide pLys stabilise la présence du plasmide d'expression dans la bactérie car la quantité de protéine produite est plus faible, ce qui est intéressant dans les cas où la protéine est toxique pour la bactérie. D'autre part, le lysozyme détruit également la paroi d'*E. coli* en hydrolysant le peptidoglycane, favorisant ainsi la lyse bactérienne.

Entrée	Souche d' <i>E. coli</i>	Vecteur Conditions testées	Résultats	
			Détection de Cel7B sur gel SDS-PAGE 10%	Détection de l'activité cellulasique
1	Origami (DE3)	pET-20b(+)/Cel7B - concentrations en IPTG : 0.5 mM, 1 mM; - températures d'induction : 37°C, 25°C; - durée de l'induction : 2 h, 5 h, 16 h.	non	non
2	BL21 (DE3) pLysS	pET-20b(+)/Cel7 B - concentrations en IPTG : 0.5 mM, 1 mM; - température d'induction : 25°C; - durée de l'induction : 3 h, 6 h, 16 h.	oui	oui
3	BL21 (DE3)	pET-20b(+)/Cel7B - concentrations en IPTG : 0.01 mM, 0.1 mM, 1 mM; - températures d'induction : 37°C, 25°C; - durée de l'induction 2 h, 5 h, 16 h.	non	non
4	BL21 (DE3)	pET-20b(+)/Cel7B mature (sans peptide signal) - concentration en IPTG : 0.05 mM; - température d'induction : 37°C; - durée de l'induction : 2.5 h, 16 h.	non	non
5	BL21 (DE3)	pET-32a(+)/Cel7B - concentrations en IPTG : 0.01 mM, 0.1 mM, 1 mM; - température d'induction : 25°C; - durée de l'induction : 2 h, 6 h, 16 h.	non	non

Tableau 16 : Etude d'expression de Cel7B d'*H. insolens* chez *E. coli*.

Plusieurs études d'expression de la protéine chez *E. coli* ont été réalisées à différentes températures, différents temps d'induction et à différentes concentrations en IPTG dans le milieu.

La délétion du peptide signal présent en amont du gène de Cel7B d'*H. insolens* a également été effectuée par PCR, cette séquence pouvant éventuellement gêner le bon repliement de la protéine.

Lors de chaque étude, la fraction périplasmique (Entrées **1** à **4**, **Tableau 16**) ou l'extrait cellulaire total (Entrée **5**) ont été testés pour leur activité cellulastique et déposés sur gel de polyacrylamide à 10% en conditions dénaturantes. La seule combinaison qui nous a permis de détecter une activité cellulastique est l'utilisation de la souche BL21 (DE3) pLysS d'*E. coli* transformée avec le plasmide pET-20b(+) porteur du gène de Cel7B d'*H. insolens* (Entrée **2**), après 6 h d'induction à 25°C avec 0.5 mM d'IPTG. Toutefois, des problèmes dans la préparation des fractions périplasmiques ont été rencontrés du fait de la fragilisation de la paroi cellulaire par le lysozyme produit. En utilisant la même construction dans la souche BL21 (DE3) d'*E. coli*, l'activité cellulastique n'a pu être détectée.

IV.1.1.2 Expression de Cel7B chez *Bacillus subtilis*

La bactérie *Bacillus subtilis* (souche DB104A) constituait un système alternatif pour l'expression de protéines, qui était utilisé au laboratoire pour sa capacité à produire la CGTase de *Bacillus circulans* (souche 251) en grande quantité (jusqu'à 35 mg par litre de culture) [230]. Le plasmide utilisé pour exprimer la CGTase est le vecteur navette pDP66K dans lequel l'expression de l'enzyme est sous le contrôle du promoteur p32 inductible par l'érythromycine.

Après insertion des sites de restriction *NcoI* et *NotI* sur le plasmide pDP66K, le gène codant pour la CGTase a été remplacé par celui codant pour Cel7B mature d'*H. insolens*. La présence d'un peptide signal en amont du gène cloné permet de sécréter la protéine éventuellement produite dans le milieu de culture, favorisant ainsi sa détection et sa purification.

Cependant, ce système s'est révélé inefficace pour exprimer Cel7B d'*H. insolens*.

IV.1.1.3 Production des mutants chez le champignon *Aspergillus oryzae*

En collaboration avec Novozymes (Danemark), les différents mutants de Cel7B E197A d'*H. insolens* préparés par mutagenèse dirigée ont finalement été produits chez le champignon *Aspergillus oryzae*, sous le contrôle du promoteur de la Taka-amylase A [231].

Après fermentation, les protéines extracellulaires sont purifiées par chromatographie échangeuse d'ions et sont utilisées ensuite pour nos tests enzymatiques.

IV.1.2 Production de la glycosynthase Cel7B E196A de *T. reesei* chez *Saccharomyces cerevisiae*

Le plasmide pMA91 porteur du gène de Cel7B E196A de *Trichoderma reesei* a été utilisé par l'entreprise VTT en Finlande pour exprimer la glycosynthase potentielle chez la levure *Saccharomyces cerevisiae*.

Le mutant Cel7B E196A de *T. reesei* a été produit chez la levure sous le contrôle du promoteur du gène *pgk1* de la phosphoglycérate kinase de la levure et sécrété dans le milieu de culture, comme décrit précédemment [232].

La protéine est concentrée et utilisée sans autre purification pour nos tests enzymatiques.

IV.2 Caractérisation des enzymes

Comme le montre la **Figure 40**, les enzymes utilisées pour la caractérisation semblent relativement pures. Elles présentent sur gel SDS-PAGE une bande caractéristique à environ 55 kDa correspondant à la taille des protéines natives dont elles sont dérivées.

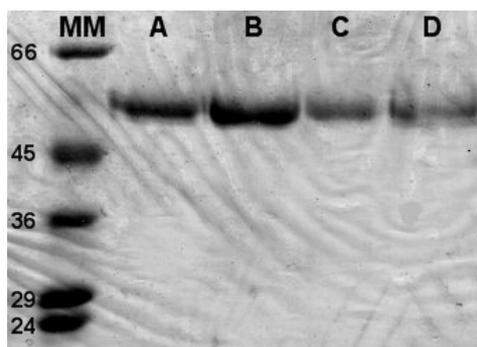


Figure 40 : Analyse des mutants de la glycosynthase d'*H. insolens* sur gel de polyacrylamide à 10% en conditions dénaturantes.

Coloration des protéines au bleu de Coomassie.

Puits **MM**, marqueur de masses moléculaires (indiquées en kDa); puits **A** : Cel7B E197A d'*H. insolens*;
puits **B** : Cel7B E197A H209A d'*H. insolens*; puits **C** : Cel7B E197A H209G d'*H. insolens*;
puits **D** : Cel7B E197A H209A A211T d'*H. insolens*.

Les différents mutants produits ont ensuite été testés pour leur capacité à transférer différents substrats accepteurs sur le fluorure d' α -lactosyle utilisé comme donneur dans chacune des réactions enzymatiques réalisées.

Les différentes réactions enzymatiques ont généralement été réalisées au pH optimum de chaque enzyme native correspondante, à pH 7 pour les enzymes dérivées de Cel7B d'*H. insolens* et à pH 4.6 pour les enzymes dérivées de Cel7B de *T. reesei*.

IV.2.1 Caractérisation des nouvelles glycosynthases avec un cellobiose comme accepteur

L'activité glycosynthase de chaque enzyme a été testée dans un premier temps de manière qualitative par chromatographie sur couche mince, puis quantifiée par HPLC dans le cas des glycosynthases dérivées de Cel7B d'*H. insolens*. La réaction suivie est le couplage enzymatique entre un donneur de lactosyle **16** et le cellobiose pour aboutir à la formation du tétrasaccharide **37** (Schéma 16).

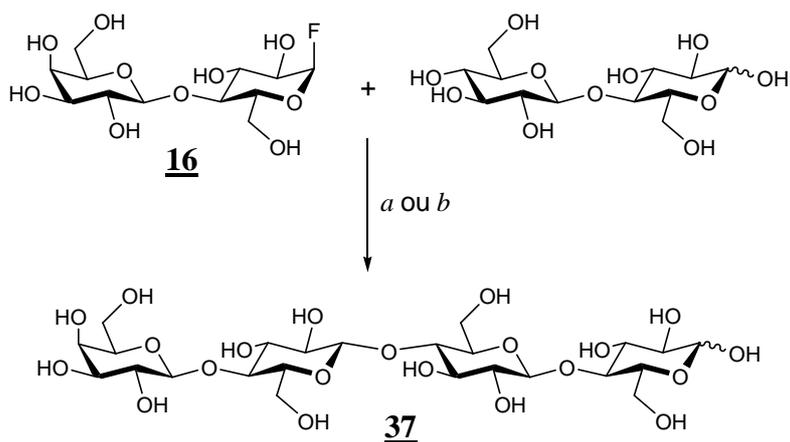


Schéma 16 : Couplage enzymatique entre le donneur 16 et le cellobiose.

Conditions : (a) mutant de Cel7B E197A *H. insolens*, tampon phosphate de sodium 50 mM pH 7;
(b) Cel7B E196A *T. reesei*, tampon acétate de sodium 25 mM pH 4.6.

➤ *Analyse qualitative*

L'analyse qualitative par suivi de la réaction sur chromatographie sur couche mince et par HPLC a tout d'abord montré sans ambiguïté que toutes les glycosynthases dérivées de Cel7B E197A d'*H. insolens* permettent l'obtention du produit de couplage 37, dans lequel la liaison formée entre le donneur et l'accepteur est du type β 1-4.

L'analyse qualitative par CCM a également révélé la formation de ce tétrasaccharide 37 avec le mutant Cel7B E196A de *T. reesei*, comme le montre la chromatographie sur couche mince de la **Figure 41**.

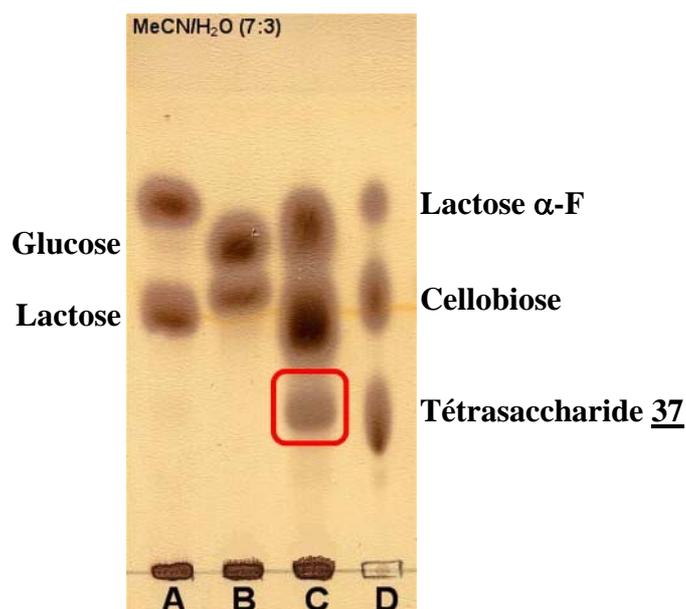


Figure 41 : Analyse par CCM du couplage enzymatique entre le donneur 16 et le cellobiose par Cel7B E196A de *T. reesei*, après 48 h d'incubation.

Ligne **A** : donneur en présence d'enzyme; ligne **B** : accepteur en présence d'enzyme;
ligne **C** : réaction entre le donneur et l'accepteur en présence d'enzyme;
ligne **D** : mélange du donneur, de l'accepteur et du tétrasaccharide 38.

L'analyse par spectrométrie de masse des différents milieux réactionnels obtenus après lyophilisation a permis de montrer la formation d'un produit dont la masse correspond au produit de couplage 37. Cela démontre que le mutant non-nucléophile E196A de Cel7B de *T. reesei* possède l'activité glycosynthase lorsque le fluorure d' α -lactosyle est utilisé comme donneur et le cellobiose comme accepteur.

Toutefois, nous avons décelé une activité β -glucosidase dans l'extrait enzymatique comme l'indique la formation de glucose en incubant l'enzyme en présence de cellobiose (Ligne **B**, **Figure 41**). Comme nous ne disposions que d'une faible quantité de solution protéique, nous n'avons pu tenter de la purifier et l'étude de cette nouvelle glycosynthase s'est limitée au test de couplage sur le N^I, N^{II} -diacétylchitobiose décrit dans le paragraphe suivant.

➤ *Quantification de l'activité glycosynthase des mutants de Cel7B E197A d'H. insolens par HPLC*

Cette caractérisation a été réalisée en mesurant la vitesse initiale de la réaction catalysée par les différentes enzymes, par suivi de la disparition du pic correspondant au cellobiose (**Figure 42**).

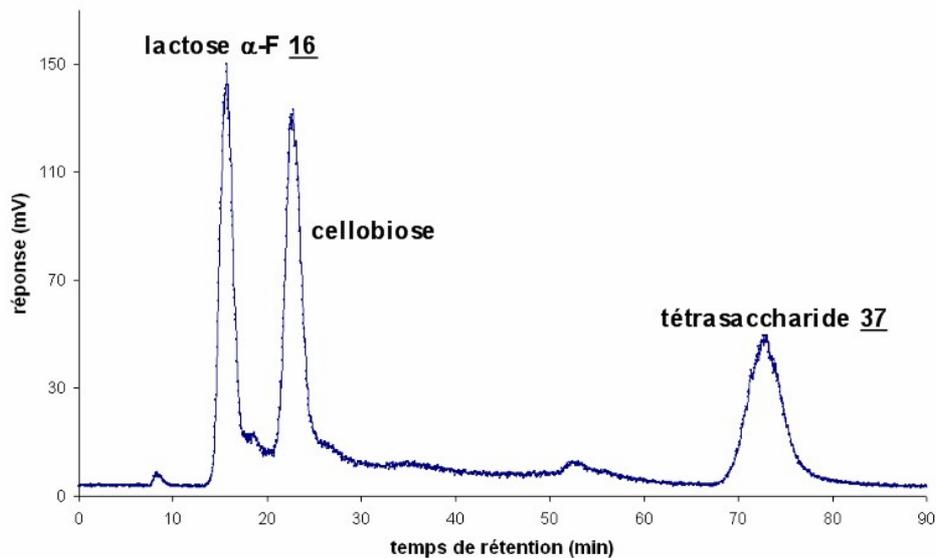


Figure 42 : Analyse par HPLC de la réaction de couplage entre le donneur 16 et le cellobiose catalysée par une glycosynthase.

Les valeurs des activités spécifiques de chaque mutant de la glycosynthase de Cel7B E197A d'*H. insolens* sont représentées sur la **Figure 43**. Elles sont exprimées en nanomoles de cellobiose consommé par minute et par milligramme de protéine.

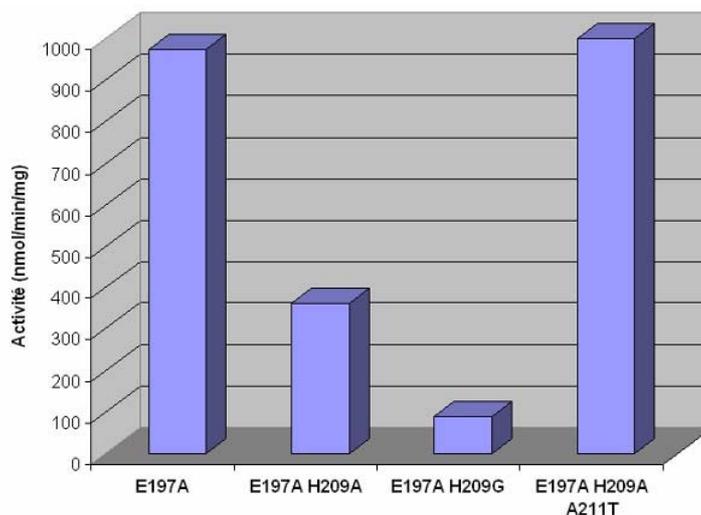


Figure 43 : Mesure de l'activité glycosynthase des différents mutants de Cel7B E197A d'*H. insolens*.

Le donneur utilisé est le fluorure d' α -lactosyle **16** et le cellobiose est utilisé comme accepteur.

En comparant l'activité spécifique de Cel7B E197A avec celle de Cel7B E197A H209A ou H209G, on peut dans un premier temps remarquer que la mutation de l'His-209 provoque une forte diminution de l'activité glycosynthase, même si celle-ci est encore présente. Le passage d'une chaîne latérale comportant un groupe méthyle (alanine) à une chaîne ne comportant qu'un atome d'hydrogène diminue encore plus fortement l'activité de l'enzyme. L'absence de chaîne latérale pourrait provoquer une modification de la structure de la protéine, conduisant à un mauvais positionnement de certains résidus essentiels pour la catalyse ou pour la fixation du substrat accepteur dans le site actif de l'enzyme.

Par contre, lorsqu'une mutation supplémentaire est insérée au niveau du résidu 211, on observe un rétablissement de l'activité glycosynthase de l'enzyme, proche de celle de la glycosynthase d'origine.

IV.2.2 Caractérisation de la spécificité de substrat

Après avoir montré que les différents mutants de type glycosynthase obtenus possédaient toujours une activité leur permettant de réaliser le couplage entre le fluorure d' α -lactosyle **16** et le cellobiose, nous avons ensuite étudié la spécificité de substrat de ces nouvelles glycosynthases. Dans le cas des mutants dérivés de la glycosynthase Cel7B E197A d'*H. insolens*, nous avons voulu savoir si la spécificité de substrat avait pu être modifiée par

les différentes mutations introduites, en utilisant une gamme variée d'accepteurs potentiels, qui n'étaient pas tous accepteurs pour Cel7B E197A.

Dans le cas de la nouvelle glycosynthase Cel7B E196A de *T. reesei*, nous avons étudié son comportement vis-à-vis d'un seul substrat accepteur, le N^I, N^{II} -diacétylchitobiose.

Les différents mono- et disaccharides utilisés pour caractériser la spécificité de substrat de ces enzymes sont répertoriés dans le **Tableau 17**, ainsi que les enzymes utilisées. Ainsi, nous avons d'une part testé en tant qu'accepteurs, des monosaccharides porteurs d'un groupement *p*-nitrophényle en position anomère (Entrées **1 à 4**).

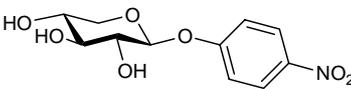
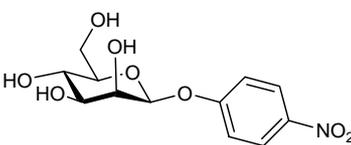
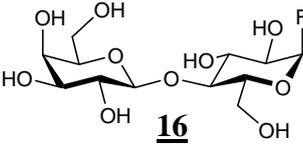
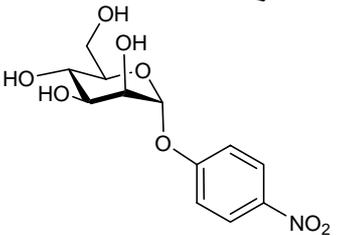
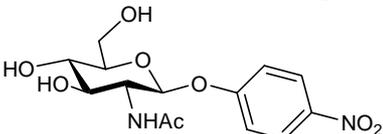
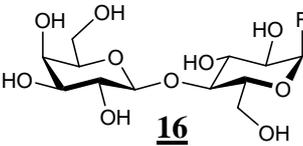
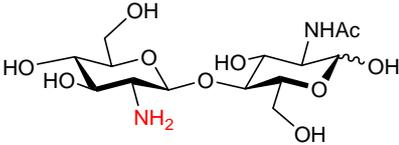
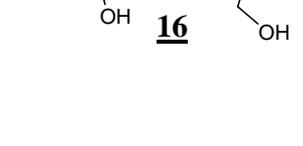
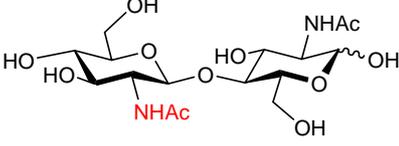
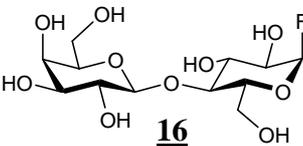
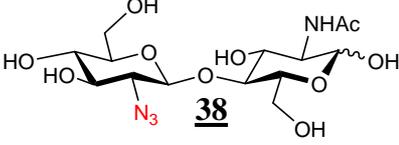
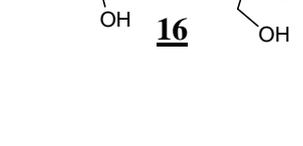
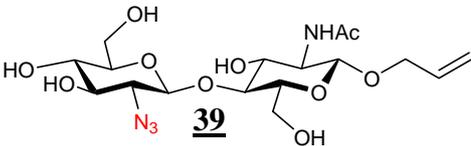
Entrée	Donneur	Accepteur	Mutants de Cel7B utilisés
1			E197A E197A H209A
2			E197A E197A H209A
3			E197A E197A H209A
4			E197A H209A E197A H209G
5			E197A H209A E197A H209G E197A H209A A211T
6			E197A H209A E197A H209G E197A H209A A211T
7			E197A E197A H209A E197A H209G E197A H209A A211T
8			E197A E197A H209A E197A H209G E197A H209A A211T

Tableau 17 : Les différentes réactions de couplage enzymatique réalisées avec les mutants de Cel7B E197A d'*H. insolens*.

Nous avons d'autre part testé également des disaccharides de type Glc(β 1-4)GlcNAc qui se différencient par leur substitution en position 2 de l'extrémité non-réductrice : un

groupement amine (Entrée 5), un groupe *N*-acétylé (Entrée 6) et un groupement azido branché sur un disaccharide substitué ou non en position anomère par un groupement allyle (Entrées 7-8). La structure du disaccharide **38** ne correspond pas directement à la spécificité de substrat recherchée, mais il nous a paru intéressant de tester ce composé substitué en position 2^{II}.

Le *N*^I-acétylchitobiose et le *N*^I,*N*^{II}-diacétylchitobiose sont disponibles au laboratoire, produits par ingénierie métabolique d'*Escherichia coli*. Les disaccharides **38** et **39**, quant à eux, sont obtenus à partir du *N*^I-acétylchitobiose selon le **Schéma 17**.

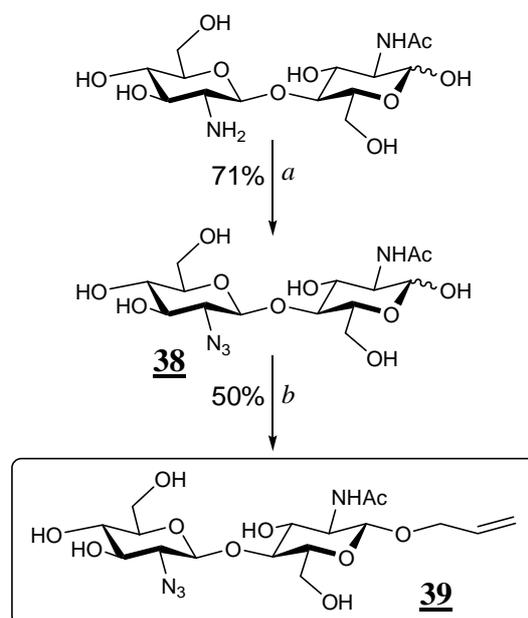


Schéma 17 : Préparation des disaccharides **38 et **39**.**

Conditions : (a) TfN_3^* , ZnCl_2 , Et_3N , $\text{H}_2\text{O}/\text{CH}_2\text{Cl}_2/\text{MeOH}$ (3:10:3, v/v/v); (b) NaH , AlBr , LiCl , DMF .

* Préparation : NaN_3 , Tf_2O , $\text{H}_2\text{O}/\text{CH}_2\text{Cl}_2$ (1:1, v/v).

Dans un premier temps, le *N*^I-acétylchitobiose est utilisé comme réactif dans une réaction de diazotransfert pour former l'azoture à partir de la fonction amine. Cette réaction a lieu en présence d'azoture triflique et de triéthylamine et est catalysée par un sel métallique tel que le chlorure de zinc, dans un mélange de solvants hydro-organique [233]. Compte-tenu du caractère explosif de TfN_3 , il est préparé *in situ* à partir d'azoture de sodium et d'anhydride triflique et est utilisé seulement en solution. Le composé **38** obtenu est alors fonctionnalisé en position anomère par un groupement allyle. La formation du disaccharide **39** est réalisée en présence d'hydrure de sodium, de bromure d'allyle et de chlorure de lithium dans le DMF [234].

IV.2.2.1 Activité vis-à-vis de glycosides de *p*-nitrophényle et de disaccharides substitués

➤ Cas des xylo- et mannopyranosides de *p*-nitrophényle et du N^I -acétylchitobiose

L'activité des différentes glycosynthases dérivées de Cel7B d'*H. insolens* vis-à-vis d'accepteurs tels que les β -D-glycosides de *p*-nitrophényle reste identique à l'activité originelle de la glycosynthase Cel7B E197A d'*H. insolens*. Ainsi, à l'opposé du β -D-mannopyranoside de *p*-nitrophényle (Entrée **2**, **Tableau 17**), l' α -D-mannopyranoside (Entrée **3**) n'est pas reconnu en tant qu'accepteur, confirmant à nouveau (a) l'activité de la glycosynthase vis-à-vis de β -D-glycosides et que (b) l'hydroxyle en position 2 n'est pas impliqué dans des interactions clés avec l'enzyme.

Le β -D-xylopyranoside de *p*-nitrophényle (Entrée **1**) est, quant à lui, toujours utilisé comme accepteur par les glycosynthases, l'hydroxyle en position 6 ne paraissant pas essentiel à la fixation de l'accepteur dans le sous-site +1 de l'enzyme.

Le N^I -acétylchitobiose reste utilisé par toutes les glycosynthases en tant qu'accepteur pour conduire à la formation du tétrasaccharide correspondant à son couplage avec le fluorure de lactosyle. Les chromatogrammes réalisés pour chacune des glycosynthases étaient identiques. Ces résultats ont été à chaque fois corrélés avec les analyses par spectrométrie de masse des milieux réactionnels après lyophilisation.

➤ Cas du N^I, N^{II} -diacétylchitobiose et du *N*-acétylglucosaminide de *p*-nitrophényle

Les différents tests effectués avec les glycosynthases d'*H. insolens* en présence de substrats accepteurs tels que la *N*-acétyl- β -D-glucosaminide de *p*-nitrophényle ou le N^I, N^{II} -diacétylchitobiose (Entrées **4** et **6**), et du fluorure d' α -lactosyle **16** comme donneur, n'ont pas permis de mettre en évidence la présence d'un produit de couplage. Ces résultats laissent donc penser que les mutations supplémentaires introduites sur la glycosynthase originelle ne permettent pas la fixation de la chaîne *N*-acétyle dans le sous-site +1 de l'enzyme.

Lorsque la glycosynthase Cel7B E196A de *T. reesei* est testée avec le N^I, N^{II} -diacétylchitobiose en tant qu'accepteur, les mêmes conclusions sont obtenues, à savoir que nous n'avons pas réussi à utiliser ce substrat en tant qu'accepteur dans le site actif de l'enzyme

et ainsi obtenir le tétrasaccharide issu du couplage entre le dérivé lactoside et ce substrat accepteur.

Ces derniers résultats ne concordent pas avec l'activité d'hydrolyse du tétrasaccharide **34** par l'endoglucanase de *T. reesei* que nous avons observée, et qui nous avait laissé penser que les mutations H209A et A211T au niveau de la glycosynthase d'*H. insolens* allaient permettre la fixation du substrat accepteur dans le site actif de l'enzyme.

Cette différence de comportement entre l'activité d'hydrolyse de l'enzyme native de *T. reesei* et l'activité de transglycosylation de la glycosynthase au niveau du sous-site +1 de l'enzyme pourrait être expliquée de la manière suivante. Lorsque l'on se place en condition d'hydrolyse avec l'enzyme native, la partie disaccharidique du côté non-réducteur (le lactosyle) du tétrasaccharide **34** se fixe dans les sous-sites donneurs de l'enzyme (-2,-1), forçant ensuite la fixation de la partie N^I, N^{II} -diacétylchitobiosyle dans les sous-sites accepteurs (+1,+2). Le tétrasaccharide **34** est alors positionné correctement dans les sous-sites catalytiques de l'enzyme et peut subir l'hydrolyse de la liaison glycosidique liant le lactosyle avec le N^I, N^{II} -diacétylchitobiose.

La fixation du N^I, N^{II} -diacétylchitobiose ne semble pas possible lorsque les deux disaccharides, substrats donneur et accepteur de la glycosynthase ne sont pas liés l'un à l'autre. Le fluorure d' α -lactosyle peut se positionner correctement dans les sous-sites donneurs du site actif de l'enzyme, mais le N^I, N^{II} -diacétylchitobiose, substrat non-naturel de l'enzyme, n'est pas positionné correctement et ne peut donc pas être utilisé en tant qu'accepteur.

➤ *Cas du 2^{II}-azidochitobiose **38***

L'analyse de la réaction de couplage entre le donneur de lactosyle **16** et le dérivé azido **38** du chitobiose, catalysée par la nouvelle glycosynthase Cel7B E197A H209A d'*H. insolens* (Entrée 7, **Tableau 17**) a révélé un comportement surprenant de l'enzyme. En effet, l'analyse par HPLC (**Figure 44**) met en évidence la présence de deux produits (pics 1 et 2) à la place d'un seul attendu si un seul produit de couplage entre le donneur et l'accepteur était formé.

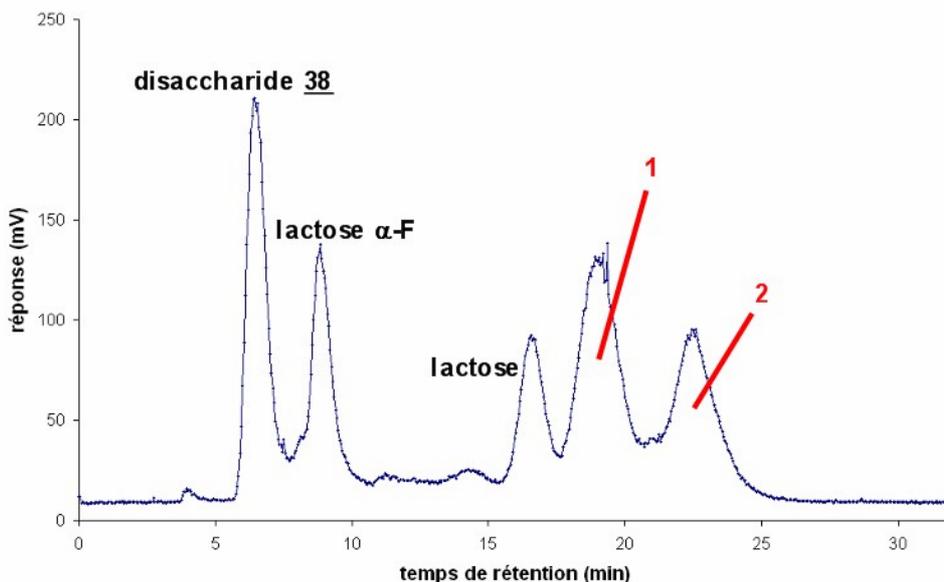


Figure 44 : Analyse par HPLC de la réaction de couplage enzymatique catalysée par la glycosynthase Cel7B E197A H209A d'*H. insolens* entre le donneur 16 et l'accepteur 38.

Afin d'identifier la nature des deux produits de couplage obtenus, nous avons réalisé une analyse par HPLC couplée à la spectrométrie de masse. Les résultats sont présentés dans le **Tableau 18**.

Pics sur le chromatogramme	Pics en masse
Pic 1	752.53 [2M+Na+NH ₄] ²⁺
	755.27 [M+Na] ⁺
Pic 2	752.52 [2M+Na+NH ₄] ²⁺
	755.25 [M+Na] ⁺

Tableau 18 : Résultats de l'analyse LC-MS du couplage enzymatique entre le donneur 16 et l'accepteur 38 par la glycosynthase Cel7B E197A H209A d'*H. insolens*.

Les deux produits de la réaction présentent la même masse moléculaire, correspondant à celle du produit de couplage attendu entre le donneur 16 et l'accepteur 38. Cette observation nous a laissé penser que nous étions en présence de deux régioisomères.

Des observations précédentes avaient montré que l'enzyme Cel7B E197A d'*H. insolens*, incubée uniquement en présence du fluorure d' α -lactosyle, pouvait former le

tétrasaccharide bis-lactose, résultat du couplage C1:C1 entre deux unités lactosyles. Pour s'affranchir d'un éventuel couplage C1:C1 possible entre le fluorure de lactosyle **16** et le 2^{II}-azidochitobiose **38**, qui pourrait expliquer la formation de deux produits de même masse, nous avons protégé la position anomère de l'accepteur **38**. L'accepteur allylique **39** dont la synthèse est décrite précédemment (Schéma 17) a ainsi été préparé.

IV.2.2.2 Activité vis-à-vis du N^I-acétyl-2^{II}-azidochitobioside d'allyle **39**

En incubant l'accepteur **39** en présence de fluorure d' α -lactosyle **16** avec Cel7B E197A H209A d'*H. insolens* (Schéma 18), nous avons à nouveau observé la formation de deux produits de couplage.

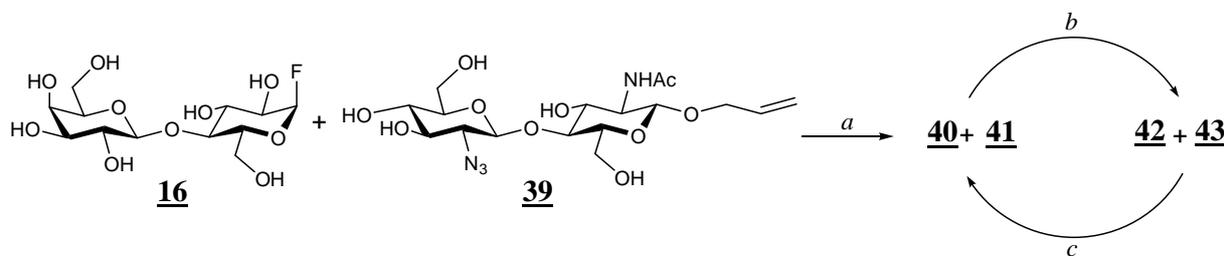


Schéma 18 : Couplage enzymatique entre le donneur **16** et l'accepteur **39**.

Conditions : (a) Cel7B E197A H209A *H. insolens*, tampon phosphate de sodium 50 mM pH 7;
(b) DMAP, Ac₂O, pyridine; (c) MeONa, MeOH.

Les deux produits **40** et **41** sont d'ailleurs clairement séparés par chromatographie sur couche mince (Figure 45). Ce résultat démontre que les régioisomères formés ne résultent pas d'un couplage de type C1:C1 entre le donneur et l'accepteur.

Afin de savoir si ce comportement de la glycosynthase pouvait être dû à la mutation de l'His-209 introduite, la même réaction a été réalisée avec la glycosynthase Cel7B E197A d'*H. insolens*, pour comparer son comportement vis-à-vis de l'accepteur **39**. La chromatographie sur couche mince comparant les deux réactions de couplage est présentée sur la Figure 45.

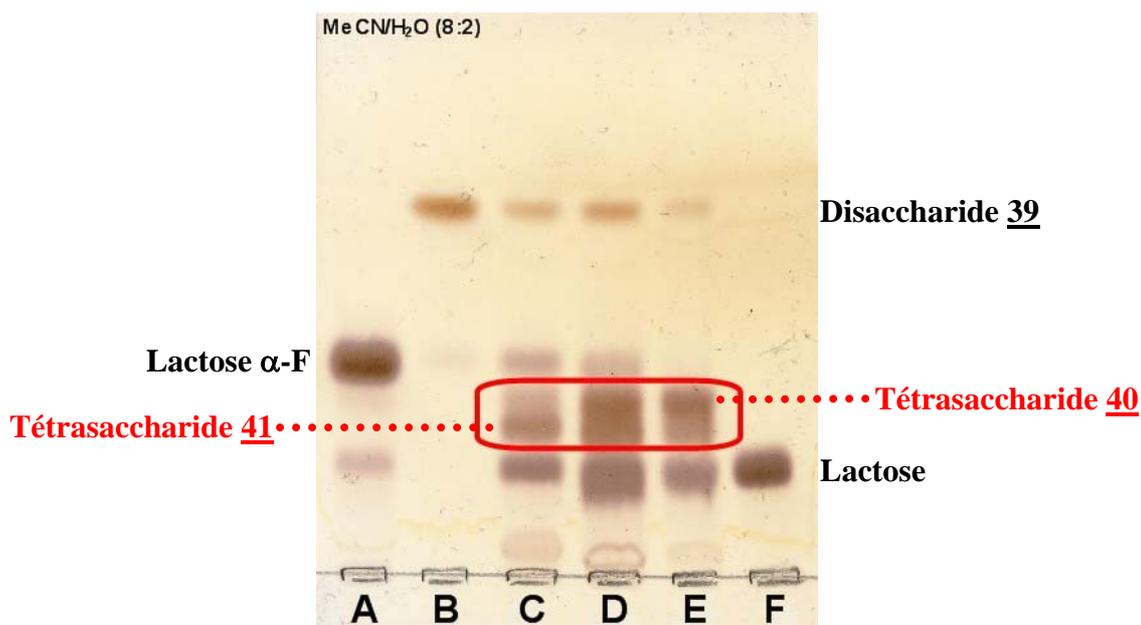


Figure 45 : Analyse par CCM du couplage entre le donneur de lactosyle 16 et l'accepteur 39.

Ligne **A** : donneur seul; ligne **B** : accepteur seul;
 ligne **C** : réaction de couplage en présence de Cel7B E197A d'*H. insolens*; ligne **D** : mélange C+E;
 ligne **E** : réaction en présence de Cel7B E197A H209A d'*H. insolens*; ligne **F** : lactose seul.

Nous pouvons de nouveau observer la formation des deux tétrasaccharides 40 et 41 dans la réaction catalysée par Cel7B E197A d'*H. insolens*. Toutefois, les proportions entre les deux composés sont totalement inversées entre les deux réactions de couplage. La glycosynthase Cel7B E197A forme préférentiellement le tétrasaccharide 41, tandis que le tétrasaccharide 40 est majoritairement obtenu avec le mutant Cel7B E197A H209A d'*H. insolens*.

IV.2.2.2.1 Caractérisations des deux tétrasaccharides 40 et 41 par RMN après acétylation

Afin de pouvoir caractériser par RMN chacun des deux tétrasaccharides formés dans les réactions de couplage catalysées par les deux glycosynthases, nous avons réalisé ces réactions à l'échelle préparative afin de pouvoir isoler en quantité suffisante chacun des deux produits.

Nous avons réalisé la réaction de couplage avec la glycosynthase Cel7B E197A d'*H. insolens* selon la même méthodologie que celle exposée dans le **Schéma 18**, ce qui nous a

permis d'obtenir majoritairement le composé **41**. Le couplage avec la glycosynthase Cel7B E197A H209A d'*H. insolens* nous a, quant à lui, permis d'obtenir majoritairement le composé **40**.

Une première purification par chromatographie rapide sur gel de silice du mélange réactionnel a été réalisée sur les produits libres, ce qui n'a pas permis de séparer correctement les deux tétrasaccharides. Après acétylation du mélange **40+41** contaminé par le lactose, nous n'avons pas pu trouver de système d'éluant nous permettant de séparer correctement les produits **42** et **43**; toutefois, le lactose a pu être éliminé. Finalement, après désacétylation du mélange de produits **42+43**, redonnant à nouveau les produits **40** et **41** en mélange, nous avons réalisé une purification sur colonne ouverte de silice, qui nous a permis de séparer les deux tétrasaccharides grâce à un gradient en acétonitrile/eau.

Les interprétations des spectres RMN ^1H et ^{13}C des deux produits de couplage sous leur forme per-*O*-acétylée ont permis de résoudre leur structure. Les analyses ont été effectuées au moyen d'expériences bidimensionnelles homonucléaire de type COSY et hétéronucléaire de type HMQC. Les interprétations des spectres ont été facilitées par le fait que les deux tétrasaccharides présentent chacun quatre unités monosaccharidiques différentes les unes des autres :

- une unité *N*-acétylglucosaminyle porteur en position anomère d'un groupe allyle;
- une unité *N*-azidoglucoaminyle;
- une unité glucosyle;
- une unité galactosyle.

Chaque système peut être résolu en utilisant un signal d'un proton facilement repérable dans l'ensemble du spectre RMN. La localisation des signaux caractéristiques du proton porté par le carbone lié au groupe *N*-acétyle d'une part, et d'autre part du proton porté par le carbone lié au groupement azido permet de remonter tout le système sans ambiguïté. Les signaux caractéristiques des protons liés à un carbone qui porte soit un groupement acétate, soit lié par une liaison glycosidique permettent de déterminer l'enchaînement des unités monosaccharidiques entre elles, notamment la régiochimie et la stéréochimie.

Les structures obtenues par l'interprétation des spectres des deux tétrasaccharides **42** et **43** sont données dans le **Tableau 19**.

Composés	Structures
40 (R=H) 42 (R=Ac)	
41 (R=H) 43 (R=Ac)	

Tableau 19 : Structures des deux régioisomères produits par les glycosynthases dérivées de Cel7B d'*H. insolens*.

➤ *Caractérisation du composé 42 (Figures 46A et 47A)*

Le signal H-2^I ($\delta = 3.93$ ppm) permet d'attribuer l'ensemble des signaux de l'unité I. Le déplacement chimique du signal H-4^I ($\delta = 3.79$ ppm) est caractéristique de la glycosylation de cette position. De même, le signal H-2^{II} ($\delta = 3.33$ ppm) permet d'attribuer l'ensemble des signaux de l'unité II. Le déplacement chimique du signal H-4^{II} ($\delta = 3.61$ ppm) est caractéristique de la glycosylation de cette position.

Les doublets H-1^I ($\delta = 4.52$ ppm, $J_{1,2} = 8.1$ Hz), H-1^{II} ($\delta = 4.28$ ppm, $J_{1,2} = 8.0$ Hz) et H-1^{III,IV} ($\delta = 4.41$ ppm, $J_{1,2} = 8.0$ Hz) sont caractéristiques d'une position anomère glycosylée de configuration β -D.

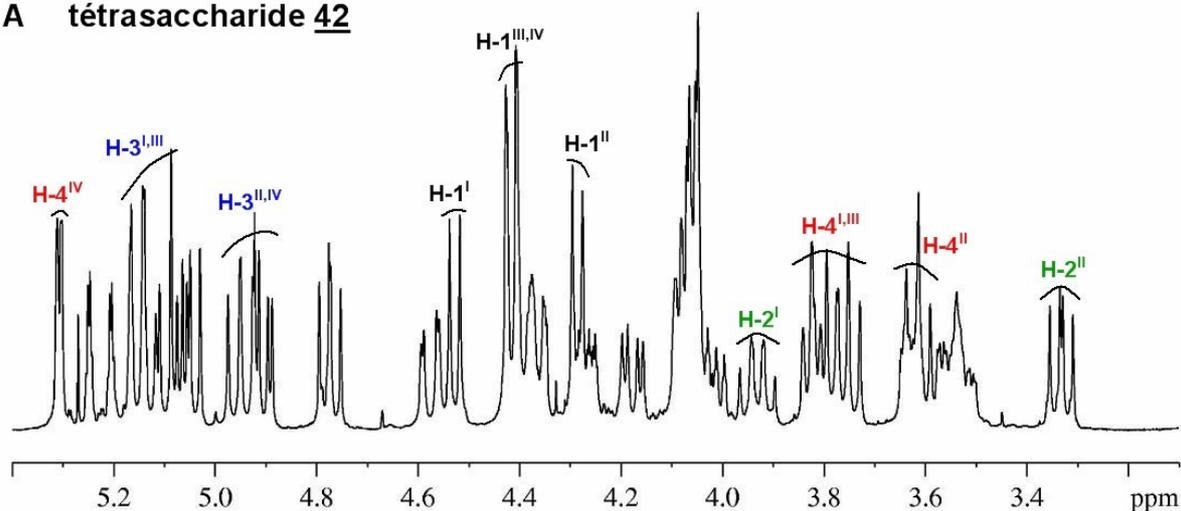
Le signal H-1^{III} ($\delta = 4.41$ ppm) permet d'attribuer l'ensemble des signaux de l'unité III. Le déplacement chimique du signal H-4^{III} ($\delta = 3.75$ ppm) est caractéristique de la glycosylation de cette position.

Les signaux H-1^{IV} ($\delta = 4.41$ ppm) et H-4^{IV} ($\delta = 5.30$ ppm) permettent d'attribuer l'ensemble des signaux de l'unité IV. Le signal H-4^{IV} ($\delta = 5.30$ ppm) est également caractéristique de l'acétylation de cette position.

Le spectre RMN ¹³C correspond également à la structure du composé **42**. Nous noterons les signaux des carbones anomères : C-1^I ($\delta = 99.6$ ppm), C-1^{II} ($\delta = 100.8$ ppm), C-1^{III} ($\delta = 100.4$ ppm) et C-1^{IV} ($\delta = 101.0$ ppm). Les signaux attribués aux C-4^{I,II,III} ($\delta = 75.8$ à

76.1 ppm) sont caractéristiques de la glycosylation de ces positions alors que le signal attribué au C-4^{IV} ($\delta = 66.5$ ppm) est caractéristique de l'acétylation de cette position. Nous notons également les signaux C-2^I ($\delta = 54.1$ ppm) et C-2^{II} ($\delta = 64.3$ ppm) caractéristiques d'une substitution de ces positions par un groupement azoté.

A tétrasaccharide 42



B tétrasaccharide 43

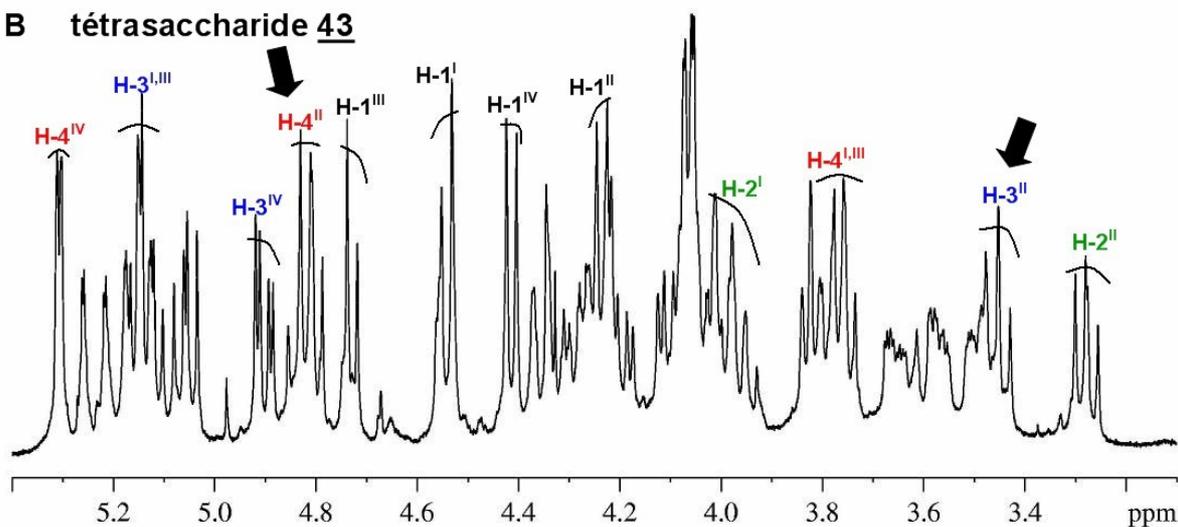


Figure 46 : Spectres RMN ¹H (A) du tétrasaccharide 42 et (B) du tétrasaccharide 43 (CDCl₃, 400 MHz, 298 K).

➤ **Caractérisation du composé 43 (Figures 46B et 47B)**

Le signal H-2^I ($\delta = 3.96$ ppm) permet d'attribuer l'ensemble des signaux de l'unité I. Le déplacement chimique du signal H-4^I ($\delta = 3.76$ ppm) est caractéristique de la glycosylation de

cette position. De même, le signal H-2^{II} ($\delta = 3.27$ ppm) permet d'attribuer l'ensemble des signaux de l'unité II. Le déplacement chimique du signal H-3^{II} ($\delta = 3.45$ ppm) est caractéristique de la glycosylation de cette position, tandis que le signal H-4^{II} ($\delta = 4.83$ ppm) est caractéristique de l'acétylation de cette position.

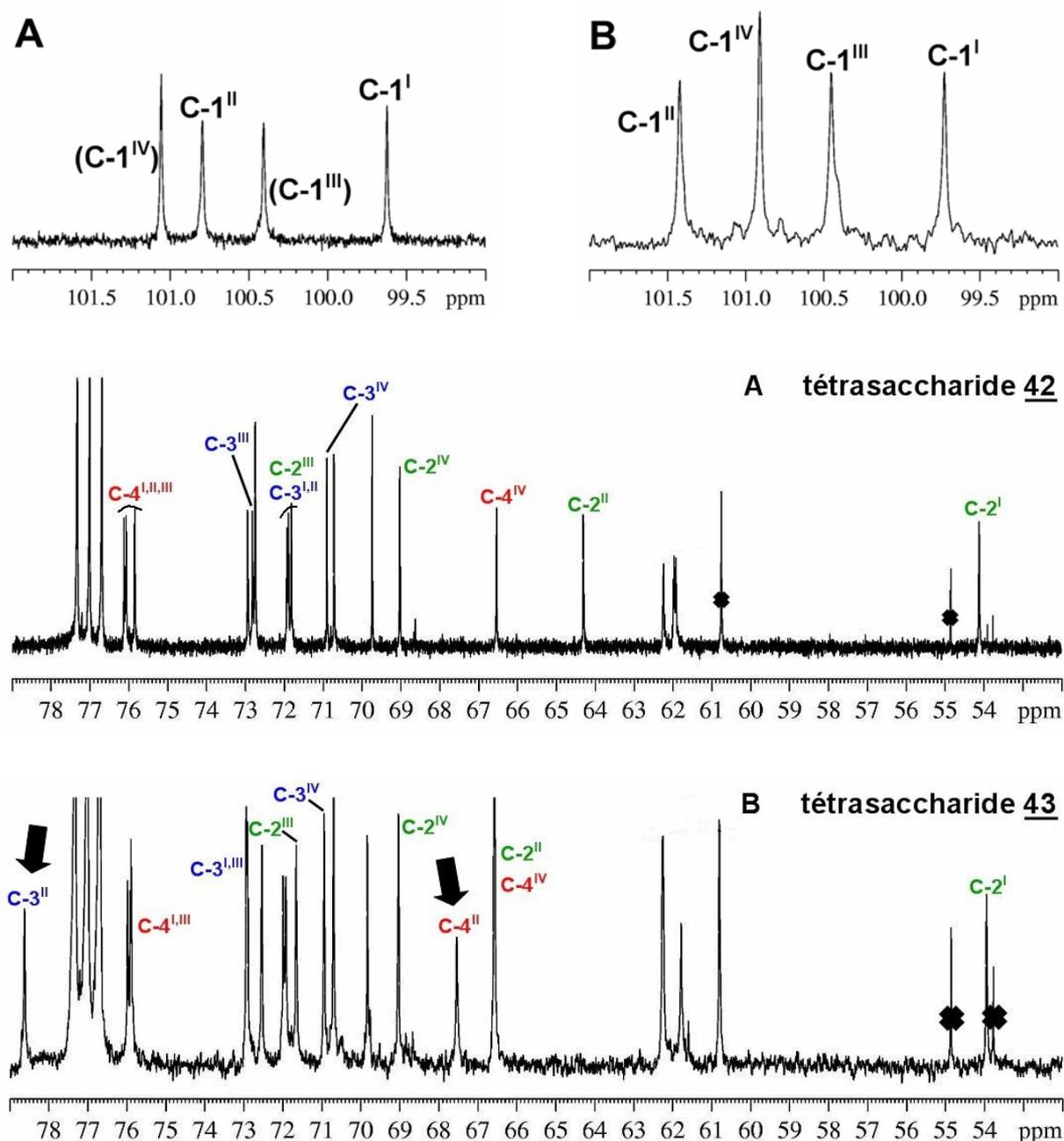


Figure 47 : Spectres RMN ¹³C (A) du tétrasaccharide 42 et (B) du tétrasaccharide 43 (CDCl₃, 400 MHz, 298 K).

Les doublets H-1^I ($\delta = 4.54$ ppm, $J_{1,2} = 8.1$ Hz), H-1^{II} ($\delta = 4.23$ ppm, $J_{1,2} = 8.1$ Hz), H-1^{III} ($\delta = 4.72$ ppm, $J_{1,2} = 7.9$ Hz) et H-1^{IV} ($\delta = 4.41$ ppm, $J_{1,2} = 7.9$ Hz) sont caractéristiques d'une position anomère glycosylée de configuration β -D.

Le signal H-1^{III} ($\delta = 4.72$ ppm) permet d'attribuer l'ensemble des signaux de l'unité III. Le déplacement chimique du signal H-4^{III} ($\delta = 3.77$ ppm) est caractéristique de la glycosylation de cette position.

Les signaux H-1^{IV} ($\delta = 4.41$ ppm) et H-4^{IV} ($\delta = 5.30$ ppm) permettent d'attribuer l'ensemble des signaux de l'unité IV. Le signal H-4^{IV} ($\delta = 5.30$ ppm) est également caractéristique de l'acétylation de cette position.

Le spectre RMN ¹³C est également en accord avec la structure du composé **43**. Nous noterons les signaux des carbones anomères : C-1^I ($\delta = 99.7$ ppm), C-1^{II} ($\delta = 101.4$ ppm), C-1^{III} ($\delta = 100.4$ ppm) et C-1^{IV} ($\delta = 100.9$ ppm). Les signaux attribués aux C-4^{I,III} ($\delta = 75.9$ à 76.0 ppm) sont caractéristiques de la glycosylation de ces positions, tout comme le signal attribué au C-3^{II} ($\delta = 78.6$ ppm). Les signaux attribués aux C-4^{IV} ($\delta = 66.6$ ppm) et C-4^{II} ($\delta = 67.5$ ppm) sont caractéristiques de l'acétylation de ces positions. Nous notons également les signaux C-2^I ($\delta = 53.9$ ppm) et C-2^{II} ($\delta = 66.5$ ppm) caractéristiques d'une substitution de ces positions par un groupement azoté.

➤ Conclusions

Les caractérisations par RMN ont permis de montrer que le tétrasaccharide **40** (**42** sous sa forme per-*O*-acétylée) était uniquement constitué d'une succession de monosaccharides liés en β 1-4 entre eux. Par contre, le tétrasaccharide **41** (**43** sous sa forme per-*O*-acétylée) présente une alternance de liaisons β 1-4 et β 1-3.

Les glycosynthases Cel7B E197A et Cel7B E197A H209A ne sont donc pas régiosélectives à 100% sur la position 4^{II} de l'accepteur, et peuvent également catalyser la formation d'une liaison β 1-3 entre le fluorure de lactosyle **16** et le 2^{II}-azidochitobiose **38**. Cette observation n'a pas de précédent sur l'ensemble des *endo*-glycosynthases décrites à ce jour. Par contre, il avait déjà été montré que certaines *exo*-glycoynthases présentaient une régiosélectivité assez large (**Tableau 5**); c'est notamment le cas de la glycosynthase dérivée de la β -glucosidase d'*Agrobacterium sp.* [**136**] (cf. paragraphe **1.2.3**).

IV.2.2.2.2 Comparaison du comportement des différents mutants

Le comportement des différentes glycosynthases d'*H. insolens* vis-à-vis de l'accepteur **39** a été comparé de façon qualitative par chromatographie sur couche mince (**Figure 48**).



Figure 48 : Analyse par CCM du couplage entre le donneur de lactosyle **16** et l'accepteur **39** après 24 h d'incubation.

Ligne A : lactose; ligne B : donneur;
ligne C : réaction de couplage en présence de Cel7B E197A d'*H. insolens*;
ligne D : réaction en présence de Cel7B E197A H209A d'*H. insolens*;
ligne E : réaction en présence de Cel7B E197A H209G d'*H. insolens*;
ligne F : réaction en présence de Cel7B E197A H209A A211T d'*H. insolens*; ligne G : accepteur.

Dans le cas du mutant Cel7B E197A H209A A211T d'*H. insolens*, seul le tétrasaccharide **41** a pu être décelé par chromatographie sur couche mince. Pour chacun des autres mutants glycosynthases testés, nous observons la formation des deux tétrasaccharides **40** et **41**, dans des proportions différentes selon l'enzyme considérée.

Les différentes glycosynthases dérivées de Cel7B d'*H. insolens* et préparées au cours de ces travaux ne présentent pas la même régiosélectivité au niveau de la nouvelle liaison formée. Les résultats obtenus sont présentés dans le **Tableau 20**.

Glycosynthase	Liaisons formées	Ordre de grandeur du rapport <u>40/41</u>
Cel7B E197A	(β 1-3)>(β 1-4)	20/80
Cel7B E197A H209A	(β 1-4)>(β 1-3)	80/20
Cel7B E197A H209G	(β 1-4)>(β 1-3)	80/20
Cel7B E197A H209A A211T	(β 1-3)	0/100

Tableau 20 : Proportions de chaque type de liaisons formées par chaque glycosynthase au cours de la réaction de couplage entre le donneur 16 et l'accepteur 39.

Ainsi, dans le cas de la glycosynthase Cel7B E197A, la substitution de la position 2 de l'extrémité non-réductrice de l'accepteur aurait permis de positionner l'unité glycosidique dans le sous-site +1 de l'enzyme de telle manière que l'hydroxyle en position 3 soit plus facilement activé par le résidu acide/base de l'enzyme. Ce positionnement du substrat accepteur aboutit alors préférentiellement à la formation du tétrasaccharide 41.

La mutation du résidu 209 en alanine ou en glycine change radicalement la régiosélectivité de la nouvelle liaison formée, de type β 1-4. Même si la poche créée par la mutation introduite en position 209 ne semble pas permettre la fixation d'un résidu *N*-acétylglucosaminyle, elle semble toutefois permettre un nouveau positionnement du résidu 2''-azidoglucosyle dans le sous-site +1 de l'enzyme. Par ce nouveau positionnement, l'hydroxyle en position 4 de l'extrémité non-réductrice du disaccharide 39 peut être activé par le résidu acide/base de la glycosynthase.

Le bénéfice de cette mutation au niveau du résidu 209 semble annulé par la mutation supplémentaire introduite au niveau du résidu 211. En effet, cette enzyme forme quasi-exclusivement une liaison de type β 1-3, comme la glycosynthase Cel7B E197A. La poche précédemment créée est très certainement remodelée par l'introduction d'une nouvelle chaîne latérale de type 2-hydroxyéthyle.

Ces derniers résultats indiquent que les mutations réalisées pour modifier la spécificité de substrat au niveau du sous-site +1, ont été réalisées sur des résidus impliqués dans cette spécificité, comme l'avaient prédit les études de modélisation moléculaire. En effet, nous avons réussi à modifier le comportement de la glycosynthase vis-à-vis du disaccharide 39.

Toutefois, ces mutations ne paraissent pas suffisantes pour modifier l'activité de la glycosynthase vis-à-vis d'un N^I, N^{II} -diacétylchitobiose.

IV.3 Conclusion

Nous avons donc au cours de ce chapitre tenté d'exprimer chez la bactérie *Escherichia coli* ou *Bacillus subtilis* l'enzyme Cel7B du champignon *Humicola insolens*. Les différentes souches et plasmides utilisés, ainsi que les différentes conditions testées n'ont pas permis de préparer sous une forme active et en quantité suffisante la protéine étudiée. Les différents mutants glycosynthases de Cel7B d'*H. insolens* ont donc été produits chez *Aspergillus oryzae*. Un mutant non-nucléophile dérivé de Cel7B de *Trichoderma reesei* a été produit chez *Saccharomyces cerevisiae*, et son activité glycosynthase a pu être mise en évidence en utilisant un cellobiose accepteur.

L'étude de l'activité glycosynthase des différents mutants de Cel7B E197A d'*H. insolens* vis-à-vis d'un cellobiose accepteur a été réalisée. Cette étude a permis de montrer l'influence des différentes mutations introduites au niveau du site actif de l'enzyme sur son activité.

La caractérisation de la spécificité de substrat des différents mutants préparés n'a pas permis de mettre en évidence un changement de la spécificité vis-à-vis d'un résidu *N*-acétylglucosaminyle dans le sous-site +1 de l'enzyme. La présence d'un groupement *N*-acétyle en position 2 de l'unité empêche toujours la fixation de l'accepteur dans le site actif, malgré les différentes mutations réalisées. La glycosynthase Cel7B E196A de *T. reesei* ne permet pas non plus le couplage entre le fluorure d' α -lactosyle et le N^I, N^{II} -diacétylchitobiose, ce qui ne concorde pas avec les tests d'hydrolyse du tétrasaccharide **34** par l'enzyme native.

Finalement, les résultats obtenus avec le disaccharide **39** permettent de penser que les mutations introduites dans l'enzyme peuvent modifier la spécificité de substrat pour différents accepteurs modifiés en position 2 de l'extrémité non-réductrice. Une nouvelle régiosélectivité de la glycosynthase Cel7B E197A d'*H. insolens* a pu être mise en évidence, avec la formation d'une liaison β 1-3 sur un accepteur 2-azidoglucosyle. Cette régiosélectivité diffère selon les

mutations introduites au niveau du site actif de l'enzyme. Ainsi, la mutation de l'His-209 en alanine ou en glycine favorise la formation d'une liaison de type β 1-4. Le mutant Cel7B E197A H209A A211T permet d'obtenir exclusivement la création d'une liaison β 1-3.

Chapitre V

PARTIE EXPERIMENTALE

V PARTIE EXPERIMENTALE

La nomenclature utilisée a été adaptée de celle publiée en anglais par : "International Union of Pure and Applied Chemistry and International Union of Biochemistry and Molecular Biology" : Nomenclature of carbohydrates, *J. Carbohydr. Chem.*, **1997**, *16*, 1191-1280; Nomenclature and Symbolism for Amino Acids and Peptides, *Biochem. J.*, **1984**, 219, 345-373.

V.1 Modélisation moléculaire

V.1.1 Alignements de séquence

Les alignements de séquences multiples ont été réalisés en utilisant le programme Clustal W version 1.82. Ils ont été édités grâce au logiciel BioEdit version 7.0.

V.1.2 Docking

Le docking a été réalisé en utilisant le pack AutoDock version 3.05 et ses modules AutoGrid version 3.06 et AutoTors.

Pour permettre au ligand de rechercher le site actif complet, un cube de dimensions 22.5×22.5×22.5 Å a été centré sur l'atome d'oxygène en 4 de l'unité saccharidique présente dans le sous-site +1. L'espacement entre chaque point de la matrice est de 0.375 Å. Les torsions possibles au niveau du ligand ont été définies grâce au module AutoTors. Chaque expérience de docking a consisté en une série de 100 cycles, utilisant l'algorithme génétique lamarckien. Tous les autres paramètres de docking sont placés à leur valeur de défaut.

Après une première étape de préparation des fichiers sources, différents fichiers sont générés : un fichier correspondant à la description de la macromolécule sous la forme de grilles, un autre pour la description du ligand, des fichiers de paramétrage d'AutoGrid (**Figure 49**) et de AutoDock (**Figure 50**).

Chapitre V : Partie expérimentale

receptor	2a39aD.pdbqs	#macromolecule
gridfld	2a39a_2Dn_maps.fld	#grid_data_file
npts	60 60 60	#numxyzpoints
spacing	0.375	#spacing/A,centering_flag
gridcenter	-7.000 -8.000 4.000	#center of grid
(Center of Mass of Subsites -1 and +1) *		
types	CNOHDQ	#atom_type_names *
map	2a39a_2Dn_C.map	#atomic_affinity_map
nbp_coeffs	1272653.000 1127.6840 12 6	#C-C_non_bond_parameters/Cn,Cm,n,m %
nbp_coeffs	610155.100 783.3452 12 6	#C-N_non_bond_parameters/Cn,Cm,n,m %
nbp_coeffs	588883.800 633.7542 12 6	#C-O_non_bond_parameters/Cn,Cm,n,m %
nbp_coeffs	1569268.000 1476.3640 12 6	#C-S_non_bond_parameters/Cn,Cm,n,m %
nbp_coeffs	88604.240 226.9102 12 6	#C-H_non_bond_parameters/Cn,Cm,n,m %
nbp_coeffs	88604.240 226.9102 12 6	#C-D_non_bond_parameters/Cn,Cm,n,m * %
nbp_coeffs	88604.240 226.9102 12 6	#C-D_non_bond_parameters/Cn,Cm,n,m * %
map	2a39a_2Dn_N.map	#atomic_affinity_map *
nbp_coeffs	610155.100 783.3452 12 6	#N-C_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	266862.200 546.7653 12 6	#N-N_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	249961.400 445.9175 12 6	#N-O_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	721128.600 1036.9320 12 6	#N-S_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	75570.000 23850.0000 12 10	#N-H_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	39093.660 155.9833 12 6	#N-D_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	39093.660 155.9833 12 6	#N-D_non_bond_parameters/Cn,Cm,n,m *
map	2a39a_2Dn_O.map	#atomic_affinity_map
nbp_coeffs	588883.800 633.7542 12 6	#O-C_non_bond_parameters/Cn,Cm,n,m
nbp_coeffs	249961.400 445.9175 12 6	#O-N_non_bond_parameters/Cn,Cm,n,m
nbp_coeffs	230584.400 368.6774 12 6	#O-O_non_bond_parameters/Cn,Cm,n,m
nbp_coeffs	675844.100 854.6872 12 6	#O-S_non_bond_parameters/Cn,Cm,n,m
nbp_coeffs	75570.000 23850.0000 12 10	#O-H_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	38919.640 124.0492 12 6	#O-D_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	38919.640 124.0492 12 6	#O-D_non_bond_parameters/Cn,Cm,n,m *
map	2a39a_2Dn_H.map	#atomic_affinity_map
nbp_coeffs	88604.240 226.9102 12 6	#H-C_non_bond_parameters/Cn,Cm,n,m
nbp_coeffs	75570.000 23850.0000 12 10	#H-N_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	75570.000 23850.0000 12 10	#H-O_non_bond_parameters/Cn,Cm,n,m
nbp_coeffs	126821.300 290.0756 12 6	#H-S_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	1908.578 46.73839 12 6	#H-H_non_bond_parameters/Cn,Cm,n,m
nbp_coeffs	1908.578 46.73839 12 6	#H-D_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	1908.578 46.73839 12 6	#H-D_non_bond_parameters/Cn,Cm,n,m *
map	2a39a_2Dn_D.map	#atomic_affinity_map *
nbp_coeffs	88604.240 226.9102 12 6	#D-C_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	39093.660 155.9833 12 6	#D-N_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	38919.640 124.0492 12 6	#D-O_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	126821.300 290.0756 12 6	#D-S_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	1908.578 46.73839 12 6	#D-H_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	1908.578 46.73839 12 6	#D-D_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	1908.578 46.73839 12 6	#D-D_non_bond_parameters/Cn,Cm,n,m *
map	2a39a_2Dn_Q.map	#atomic_affinity_map *
nbp_coeffs	0.0 0.0 12 6	#Q-C_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	0.0 0.0 12 6	#Q-N_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	0.0 0.0 12 6	#Q-O_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	0.0 0.0 12 6	#Q-S_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	0.0 0.0 12 6	#Q-H_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	0.0 0.0 12 6	#Q-D_non_bond_parameters/Cn,Cm,n,m *
nbp_coeffs	0.0 0.0 12 6	#Q-D_non_bond_parameters/Cn,Cm,n,m *
elecmap	2a39a_2Dn_e.map	#electrostatic_PE_map
dielectric	-1	#distance dependent
#fmap	2a39a_2Dn_f.map	#floating_grid *

Figure 49 : Fichier caractéristique regroupant les paramètres utilisés par AutoGrid.

Chapitre V : Partie expérimentale

```
Seed time pid # for random number generator
types COHD # atom type names
fld 2a39a_2Dn_maps.fld # grid data file
map 2a39a_2Dn_C.map # C-atomic affinity map file
map 2a39a_2Dn_O.map # O-atomic affinity map file
map 2a39a_2Dn_H.map # H-atomic affinity map file
map 2a39a_2Dn_D.map # N-atomic affinity map file
map 2a39a_2Dn_e.map # electrostatics map file

move rg3c_8cel.all # small molecule file
about -7.000 -8.000 4.000 # small molecule center

# Initial Translation, Quaternion and Torsions
tran0 -7.000 -8.000 4.000 # initial coordinates/A or "random"
quat0 1. 0. 0. 0. # initial quaternion or "random"
ndihe 18 # number of initial torsions
dihe0 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. # initial torsions

torsdof 0 0.3113 # num. non-Hydrogen torsional DOF & coeff.

# Initial Translation, Quaternion and Torsion Step Sizes and Reduction Factors
tstep 0.2 # translation step/A
qstep 5. # quaternion step/deg
dstep 15. # torsion step/deg
trnrf .9875 # trans reduction factor/per cycle
quarf .9875 # quat reduction factor/per cycle
dihrf .9875 # tors reduction factor/per cycle

# Hard Torsion Constraints
#hardtorcon 1 -180. 30. # constrain torsion, num., angle(deg), range(deg)

# Internal Non-Bonded Parameters
intnbp_coefs 1272653.000 1127.6840 12 6 # C-C internal energy non-bond parameters/Cn,Cm,n,m
intnbp_coefs 588883.800 633.7542 12 6 # C-O internal energy non-bond parameters/Cn,Cm,n,m
intnbp_coefs 88604.240 226.9102 12 6 # C-H internal energy non-bond parameters/Cn,Cm,n,m
intnbp_coefs 88604.240 226.9102 12 6 # C-D internal energy non-bond parameters/Cn,Cm,n,m *
intnbp_coefs 230584.400 368.6774 12 6 # O-O internal energy non-bond parameters/Cn,Cm,n,m
intnbp_coefs 75570.000 23850.0000 12 10 # O-H internal energy non-bond parameters/Cn,Cm,n,m *
intnbp_coefs 38919.640 124.0492 12 6 # O-D internal energy non-bond parameters/Cn,Cm,n,m
intnbp_coefs 1908.578 46.73839 12 6 # H-H internal energy non-bond parameters/Cn,Cm,n,m
intnbp_coefs 1908.578 46.73839 12 6 # H-D internal energy non-bond parameters/Cn,Cm,n,m *
intnbp_coefs 1908.578 46.73839 12 6 # D-D internal energy non-bond parameters/Cn,Cm,n,m *

#intelec # calculate internal electrostatic energy

#watch sman1.qamb.tors.watch.pdb # real-time monitoring file

outlev 0 # diagnostic output level

# Docked Conformation Clustering Parameters for "analysis" command
rmstol 1.0 # cluster tolerance (Angstroms)
rmsref Pacvmomin.t.pdbqs # reference structure file for RMS calc.
#rmsnosym # do no symmetry checking in RMS calc.
Write_all # write all conformations in a cluster

extnrg 1000. # external grid energy
e0max 0. 10000 # max. allowable initial energy, max. num. retries

# Genetic Algorithm (GA) and Lamarckian Genetic Algorithm Parameters (LGA)
ga_pop_size 50 # number of individuals in population
ga_num_evals 2000000 # maximum number of energy evaluations
ga_num_generations 500 # maximum number of generations
ga_elitism 1 # num. of top individuals that automatically survive
ga_mutation_rate 0.02 # rate of gene mutation
ga_crossover_rate 0.80 # rate of crossover
ga_window_size 10 # num. of generations for picking worst individual
ga_cauchy_alpha 0 # ~mean of Cauchy distribution for gene mutation
ga_cauchy_beta 1 # ~variance of Cauchy distribution for gene mutation
set_ga # set the above parameters for GA or LGA

# Local Search (Solis & Wets) Parameters (for LS alone and for LGA)
sw_max_its 300 # number of iterations of Solis & Wets local search
sw_max_succ 4 # number of consecutive successes before changing rho
sw_max_fail 4 # number of consecutive failures before changing rho
sw_rho 1.0 # size of local search space to sample
sw_lb_rho 0.01 # lower bound on rho
ls_search_freq 0.07 # probability of performing local search on an indiv.
Set_pswl # set the above pseudo-Solis & Wets parameters

ga_run 100 # do this many GA or LGA runs

# Perform Cluster Analysis
analysis # do cluster analysis on results
```

Figure 50 : Fichier caractéristique regroupant les paramètres utilisés par AutoDock.

Les grilles de potentiels sont ensuite calculées avec AutoGrid.

Finalement, les solutions de docking sont recherchées avec AutoDock. Pour chaque expérience, ces solutions ont été regroupées en clusters avec une tolérance de la déviation RMS de 1 Å. Seuls les clusters les plus représentatifs ont été analysés plus en détail.

Les additions des hydrogènes aux molécules utilisées ont été effectuées grâce au logiciel WHAT IF. Les charges atomiques partielles sont assignées à partir des paramètres PIM pour le champ de force TRIPOS^h ou par l'intermédiaire du logiciel MOPAC version 7 (pour le monosaccharide *N*-acétylglucosamine). Les effets de solvant n'ont pas été inclus aux études de docking.

Le logiciel BABEL version 1.6 nous a permis de convertir les fichiers entre les différents formats utilisés. Le programme TINKER version 3.9ⁱ nous a permis de minimiser les hydrogènes.

V.1.3 Visualisation des structures

Les figures du manuscrit ont été réalisées grâce aux logiciels DeepView/Swiss-Pdb Viewer version 3.7 ou Rasmol version 2.7.2. Certaines images ont été retravaillées avec le logiciel POV-Ray for Windows version 3.5.

V.2 Synthèses chimiques et enzymatiques

V.2.1 Matériels et méthodes générales

V.2.1.1 Produits de départ

Le *N*¹-acétylchitobiose et le *N*¹,*N*¹¹-diacétylchitobiose nous ont été fournis par Eric Samain. Ils ont été produits par ingénierie métabolique d'*Escherichia coli*.

^h Site Internet : <http://webenligne.cermav.cnrs.fr/databank/pim/>

ⁱ Site Internet : <http://dasher.wustl.edu/tinker/>

V.2.1.2 Solvants et tampons

Les solvants sont distillés [235] comme indiqués ci-dessous puis conservés sur tamis moléculaire, sur hydroxyde de potassium ou sur hydrure de calcium.

Le dichlorométhane est distillé sur anhydride phosphorique (P₂O₅). La pyridine est distillée sur hydroxyde de potassium. L'acétonitrile l'est sur hydrure de calcium.

Lors des étapes de glycosylation, les solvants utilisés sont fraîchement distillés.

Les différents tampons utilisés dans les réactions enzymatiques sont préparés comme décrit dans la littérature [236].

V.2.1.3 Enzymes utilisées

L'enzyme mutée Cel7B E197A d'*Humicola insolens* nous a été fournie par les Drs Martin Schülein et Jesper Vind (Novozymes, Danemark). Elle a été produite chez *Aspergillus oryzae* et purifiée [130].

L'enzyme mutée Cel7B E196A de *Trichoderma reesei* nous a été fournie par le Dr Anu Koivula (VTT, Finlande). Elle a été produite chez *Saccharomyces cerevisiae* et utilisée pour les tests enzymatiques après concentration du milieu de culture (cf. paragraphe V.3.4.2 pour la description).

V.2.1.4 Chromatographies

L'avancement des réactions a été suivi par chromatographie sur couche mince (CCM) sur plaque de gel de silice (Merk 60 F₂₅₄) en utilisant des systèmes d'éluant appropriés. La révélation est effectuée par irradiation sous lumière UV ($\lambda = 254$ nm) ainsi que par trempage dans des révélateurs spécifiques et chauffage à 300°C. Les révélateurs utilisés sont :

- un mélange acide (acide sulfurique/méthanol/eau, 5/45/45) pour les composés glycosidiques protégés;
- un mélange acide à base d'orcinol (1 g solubilisé dans 1 L d'un mélange acide sulfurique/éthanol/eau, 3/72.5/22.5) pour les composés glycosidiques libres;

- un mélange phosphomolybdique (50 g d'acide phosphomolybdique solubilisé dans 1 L d'éthanol) pour les dérivés peptidiques.

Les séparations par chromatographie rapide sont réalisées sous pression d'air comprimé sur colonne de gel de silice (Merk Geduran SI 60; 40-63 μm) avec l'éluant indiqué.

Les séparations par chromatographie sur colonne ouverte sont réalisées sous pression atmosphérique sur colonne de gel de silice (Merk 60; 0.063-0.200 mm) avec l'éluant indiqué.

V.2.1.5 Analyses physico-chimiques

Les pouvoirs rotatoires sont mesurés à 25°C sur un polarimètre Perkin Elmer 341.

Les spectres de masse basse résolution ont été effectués au CERMAV sur :

- un appareil quadripolaire NERMAG R 10-10C version 2000 en masse pour les méthodes d'ionisation de type bombardement par atomes accélérés (FAB) et ionisation chimique (DCI);
- un spectromètre Waters Micromass ZQ pour la méthode d'ionisation de type électrospray;
- un appareil Bruker Autoflex pour la technique MALDI.

V.2.1.6 Caractérisations par spectroscopie par Résonance Magnétique Nucléaire (RMN)

Les spectres RMN ont été réalisés sur des appareils Bruker AC300 et Avance 400.

Les déplacements chimiques (δ) sont exprimés en partie par million (ppm) par rapport au tétraméthylsilane utilisé comme référence interne et en utilisant comme solvants ceux indiqués; les constantes de couplage (J) sont mesurées en hertz (Hz). La notation utilisée pour la description des spectres est la suivante : s (singulet), d (doublet), dd (doublet de doublet), m (massif ou multiplet), t (triplet), q (quadruplet), $H_{\text{arom.}}$ (protons aromatiques) et $C_{\text{arom.}}$ (carbones aromatiques).

Pour les composés glycosidiques libres, les échantillons sont analysés après lyophilisation dans l'eau lourde à 99.9% afin d'échanger les protons des hydroxyles par des deutériums et les mesures sont effectuées dans l'eau lourde $\geq 99.97\%$. La référence utilisée est

alors le pic résiduel de l'eau pour les spectres protons (étalonné à $\delta = 4.77$ ppm pour une température de 298 K et à 4.72 ppm pour une température de 303 K).

Pour les composés glycosidiques protégés et/ou solubles dans le chloroforme, la référence utilisée est alors le signal du CHCl_3 pour les spectres protons (étalonné à $\delta = 7.24$ ppm quelque soit la température utilisée) ou la raie centrale du signal du CHCl_3 pour les spectres carbones (étalonné à $\delta = 77.0$ ppm).

Pour les dérivés peptidiques solubles dans le méthanol, la référence utilisée est le signal du CH_3OD pour les spectres protons (étalonné à $\delta = 3.31$ ppm) ou la raie centrale du CH_3OD pour les spectres carbones (étalonné à $\delta = 49$ ppm).

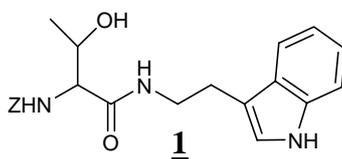
Les expériences bidimensionnelles homo- et hétéronucléaires sont calibrées par analogie avec les spectres à une dimension correspondant. Les programmes utilisés sont ceux de la bibliothèque de la société Bruker.

Pour l'attribution des signaux en RMN, les unités monosaccharidiques sont numérotées par ordre croissant (I, II, III, IV) de l'extrémité réductrice à l'extrémité non-réductrice.

V.2.2 Synthèses

V.2.2.1 Synthèse peptidique

***N*²-Benzyloxycarbonyl-*N*¹-indolyléthyl-L-thréonine [*Z*-Thr-Tryptamine] 1**



La *Z*-L-thréonine (1.61 g, 6.37 mmol, 1 éq.) et la tryptamine (1.02 g, 6.37 mmol, 1 éq.) sont solubilisées dans 40 mL de *N,N*-diméthylformamide et la solution est refroidie à 0°C. A celle-ci sont ajoutés de la triéthylamine (976 μL , 7 mmol, 1.1 éq.) puis le *N*-hydroxybenzotriazole sous forme hydratée (0.86 g, 6.37 mmol, 1 éq.) et finalement la dicyclohexylcarbodiimide (1.44 g, 7 mmol, 1 éq.). Le mélange réactionnel est agité pendant

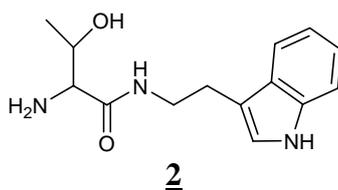
30 min à 0°C puis 24 h à température ambiante. Le précipité blanc de dicyclohexylurée formée est éliminé par filtration sur Büchner. Le filtrat est ensuite concentré à sec puis repris dans le chloroforme. La phase organique est lavée successivement avec une solution saturée d'hydrogénocarbonate de sodium et à l'eau, puis séchée sur sulfate de sodium anhydre et finalement concentrée. Après purification par chromatographie rapide sur gel de silice (cyclohexane/acétone, 3:1 → 2:1 → 3:2 → 1:1), le composé **1** est obtenu avec un rendement quantitatif (2.5 g)

SM (FAB) : $m/z = 396 [M+H]^+$.

RMN ¹H (CDCl₃, 400 MHz) : δ (ppm) = 8.57 (s, 1H, NH_{arom.}); 7.67-6.86 (m, H_{arom.}); 6.15 (d, 1H, OH^{Thr}); 5.38 (s, 2H, NH); 5.00 (s, 2H, CH₂^{Thr}); 4.24 (m, 1H, CH^{Thr}); 4.11 (m, 1H, CH^{Thr}); 3.45 (m, 2H, CH₂^{Tryp.}); 2.83 (m, 2H, CH₂^{Tryp.}); 1.07 (m, 3H, CH₃^{Thr}).

RMN ¹³C (CDCl₃, 400 MHz) : δ (ppm) = 170.9 (CO); 162.9-111.2 (C_{arom.}); 67.0, 66.9 (CH^{Thr}, CH₂^{Tryp.}); 59.3 (CH^{Thr}); 39.6 (CH₂^{Tryp.}); 24.8 (CH₂^{Tryp.}); 18.5 (CH₃^{Thr}).

N¹-indolyléthyl-L-thréonine [H₂N-Thr-Tryptamine] 2

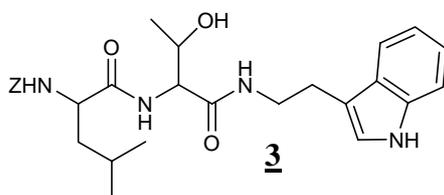


A une solution du composé **1** (2.5 g, 6.37 mmol) dans 25 mL de méthanol est ajoutée une quantité catalytique d'hydroxyde de palladium sur charbon. Le mélange réactionnel est alors agité vigoureusement sous atmosphère d'hydrogène à température ambiante pendant 4.5 h, puis est filtré sur célite et concentré. Le composé **2**, obtenu avec un rendement quantitatif (1.6 g), est utilisé sans autre purification.

SM (FAB) : $m/z = 262 [M+H]^+$.

***N*²-(*N*²-Benzyloxycarbonyl-L-leucyl)-*N*¹-indolyléthyl-L-thréonine**

[Z-Leu-Thr-Tryptamine] 3



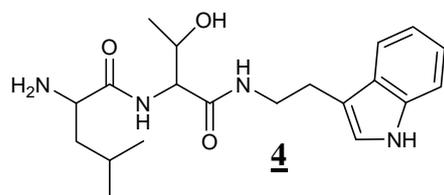
Le composé 3 est préparé dans les mêmes conditions que le composé 1, par couplage peptidique entre la Z-L-leucine (1.69 g, 6.37 mmol, 1 éq.) et le composé 2 (1.6 g, 6.37 mmol, 1 éq.). Après purification par chromatographie rapide sur gel de silice (cyclohexane/acétone, 3:2 → 1:1), le composé 3 est obtenu avec 96% de rendement (3.1 g, 6.1 mmol).

SM (FAB) : $m/z = 531 [M+Na]^+$.

RMN ¹H (CD₃OD, 400 MHz) : δ (ppm) = 7.49-6.92 (m, H_{arom.}); 4.19-4.09 (m, 3H, CH^{Leu,Thr,Thr}); 3.40 (m, 2H, CH₂^{Tryp.}); 3.25 (m, 1H, OH^{Thr}); 2.89 (m, 2H, CH₂^{Tryp.}); 1.64 (m, 1H, CH^{Leu}); 1.51 (m, 2H, CH₂^{Leu}); 1.06 (m, 3H, CH₃^{Thr}); 0.87 (m, 6H, CH₃^{Leu}).

RMN ¹³C (CD₃OD, 400 MHz) : δ (ppm) = 175.7, 172.1 (CO); 158.9-112.2 (C_{arom.}); 68.1, 67.8 (CH₂^{Leu}, CH^{Thr}); 59.9 (CH^{Thr}); 55.5 (CH^{Leu}); 41.6, 41.4 (CH₂^{Leu}, CH₂^{Tryp.}); 26.7, 26.1 (CH₂^{Tryp.}, CH^{Leu}); 23.4, 21.8, 20.2 (CH₃^{Thr,Leu}).

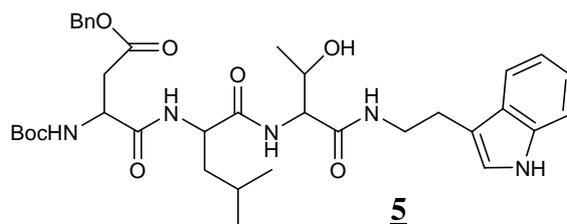
***N*²-(L-Leucyl)-*N*¹-indolyléthyl-L-thréonine [H₂N-Leu-Thr-Tryptamine] 4**



La préparation du composé 4 s'effectue selon les mêmes conditions de déprotection que pour le composé 2 à partir du composé 3 (3.1 g, 6.1 mmol). Le composé 4 est obtenu avec un rendement quantitatif (2.3 g).

SM (FAB) : $m/z = 375 [M+H]^+$.

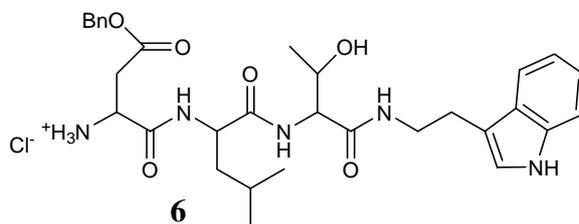
***N*²-(*N*²-*tert*-Butyloxycarbonyl- β -benzyl-L-aspartyl-L-leucyl)-*N*¹-indolyléthyl-L-thréonine
[Boc-Asp(OBn)-Leu-Thr-Tryptamine] 5**



Selon la même procédure que pour la préparation du composé 1, le couplage peptidique est effectué entre l'acide Boc- β -benzyl-L-aspartique (1.97 g, 6.1 mmol, 1 équ.) et le composé 4 (2.3 g, 6.1 mmol, 1 équ.). Après purification par chromatographie rapide sur gel de silice (cyclohexane/acétone, 3:2 \rightarrow 1:1), le composé 5 est obtenu avec 87% de rendement (3.6 g, 5.3 mmol).

SM (FAB) : $m/z = 702 [M+Na]^+$.

**Hydrochlorure de *N*²-(β -benzyl-L-aspartyl-L-leucyl)-*N*¹-indolyléthyl-L-thréonine
[Cl⁻H₃N-Asp(OBn)-Leu-Thr-Tryptamine] 6**



Le composé 5 (3.6 g, 5.3 mmol) est solubilisé à 0°C dans 30 mL d'une solution d'acide chlorhydrique (4 M) dans le dioxane anhydre. Le mélange réactionnel est agité à 0°C pendant 4 h puis concentré à sec. Le résidu est co-évacué plusieurs fois au toluène. Le composé 6 est

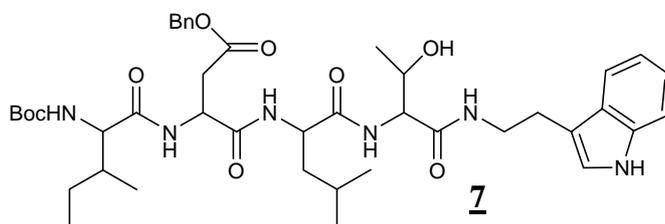
obtenu sous la forme d'un sel de chlorure d'ammonium avec un rendement quantitatif (3.1 g) et utilisé sans autre purification.

SM (FAB) : $m/z = 580 [M-Cl]^+$.

RMN 1H (CD₃OD, 400 MHz) : δ (ppm) = 7.50-6.90 (H_{arom.}); 5.12 (m, 2H, CH₂^{Asp}); 4.39 (m, 1H, CH^{Leu}); 4.24 (m, 2H, CH^{Thr,Asp}); 4.04 (m, 1H, CH^{Thr}); 3.41 (m, 2H, CH₂^{Tryp.}); 3.23 (m, 1H, OH^{Thr}); 2.89 (m, 4H, CH₂^{Tryp.,Asp}); 1.60 (m, 3H, CH^{Leu}, CH₂^{Leu}); 1.04 (m, 3H, CH₃^{Thr}); 0.86 (m, 6H, CH₃^{Leu}).

RMN ^{13}C (CD₃OD, 400 MHz) : δ (ppm) = 174.3-169.3 (CO); 145.5-112.2 (C_{arom.}); 68.4 (CH^{Thr}, CH₂^{Asp}); 60.0 (CH^{Thr}); 54.1 (CH^{Leu}); 50.7 (CH^{Asp}); 41.5 (CH₂^{Leu}, CH₂^{Tryp.}); 34.4 (CH₂^{Asp}); 26.6-25.8 (CH₂^{Tryp.}, CH^{Leu}); 23.4-20.1 (CH₃^{Thr,Leu}).

N*²-(*N*²-(*N*²-*tert*-Butyloxycarbonyl-L-isoleucyl)- β -benzyl-L-aspartyl-L-leucyl)-*N*¹-indolyléthyl-L-thréonine [Boc-Ile-Asp(OBn)-Leu-Thr-Tryptamine] **7*



Le couplage peptidique entre le Boc-L-isoleucine (1.22 g, 5.3 mmol, 1 éq.) et le composé **6** (3 g, 5.3 mmol, 1 éq.) est réalisé selon la même procédure que pour le composé **1**. Après purification par chromatographie rapide sur gel de silice (cyclohexane/acétone, 3:2 → 1:1), le composé **7** est obtenu avec 71% de rendement (3 g, 3.8 mmol).

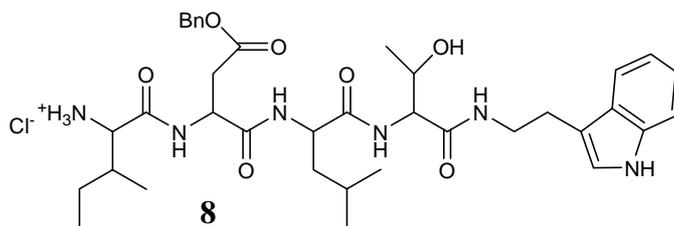
SM (FAB) : $m/z = 793 [M+H]^+$.

RMN 1H (CD₃OD, 400 MHz) : δ (ppm) = 7.52-6.90 (m, H_{arom.}); 5.04 (m, 1H, CH^{Ile}); 4.62 (m, 1H, CH^{Asp}); 4.36 (m, 1H, CH^{Leu}); 4.21 (m, 2H, CH^{Thr}); 3.59 (m, 1H, CH^{Ile}); 3.44 (m, 2H,

$\text{CH}_2^{\text{Tryp.}}$; 3.25 (OH^{Thr}); 2.89 (m, 4H, $\text{CH}_2^{\text{Asp, Tryp.}}$); 2.79 (m, 2H, CH_2^{Ile}); 1.67 (m, 3H, CH^{Leu} , CH_2^{Leu}); 1.40 (m, 9H, $\text{CH}_{3, \text{Boc}}^{\text{Ile}}$); 1.09 (m, 3H, CH_3^{Thr}); 0.85 (m, 12H, $\text{CH}_3^{\text{Leu, Ile}}$).

RMN ^{13}C (CD_3OD , 400 MHz) : δ (ppm) = 174.9-172.1 (CO); 138.1-112.2 ($\text{C}_{\text{arom.}}$); 81.0 ($\text{C}_{\text{Boc}}^{\text{Ile}}$); 68.1-67.9 ($\text{CH}^{\text{Ile, Thr}}$, CH_2^{Asp}); 61.0 (CH^{Thr}); 54.1 ($\text{CH}^{\text{Ile, Leu}}$); 51.7 (CH^{Asp}); 41.5 ($\text{CH}_2^{\text{Leu, Tryp.}}$); 38.2 (CH_2^{Asp}); 28.8 ($\text{CH}_{3, \text{Boc}}^{\text{Ile}}$); 26.2-25.7 (CH^{Leu} , $\text{CH}_2^{\text{Ile, Tryp.}}$); 23.5-20.1 ($\text{CH}_3^{\text{Thr, Leu}}$); 15.9 (CH_3^{Ile}); 11.8 (CH_3^{Ile}).

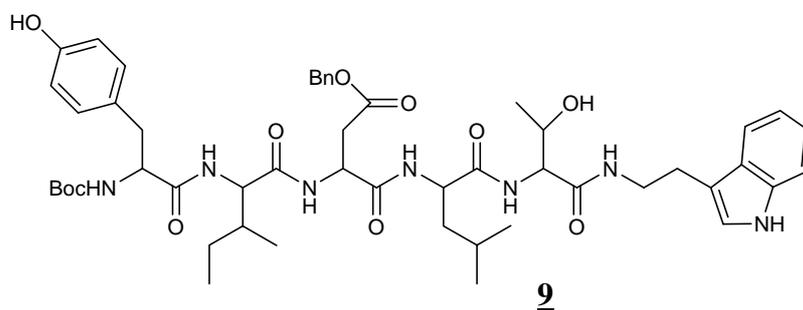
Hydrochlorure de N^2 -(N^2 -(L-Isoleucyl)- β -benzyl-L-aspartyl-L-leucyl)- N^1 -indolyléthyl-L-thréonine [$\text{Cl}^+\text{H}_3\text{N-Ile-Asp(OBn)-Leu-Thr-Tryptamine}$] **8**



La déprotection de la fonction amine du composé **7** (2.9 g, 3.7 mmol) est réalisée selon la même procédure que pour la préparation du composé **6**. Le composé **8** est ainsi obtenu sous la forme d'un sel de chlorure d'ammonium avec un rendement quantitatif (2.6 g) et utilisé sans autre purification.

SM (FAB) : $m/z = 693 [\text{M}-\text{Cl}]^+$.

N*²-(*N*²-(*N*²-*tert*-Butyloxycarbonyl-L-tyrosyl-L-isoleucyl)-β-benzyl-L-aspartyl-L-leucyl)-*N*¹-indolyléthyl-L-thréonine [Boc-Tyr-Ile-Asp(OBn)-Leu-Thr-Tryptamine] **9*



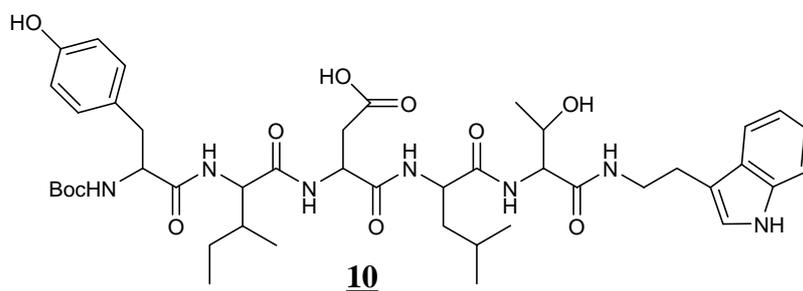
Le composé **9** est préparé dans les mêmes conditions que le composé **1**, par couplage peptidique entre le Boc-L-tyrosine (1.04 g, 3.7 mmol, 1 éq.) et le composé **8** (2.7 g, 3.7 mmol, 1 éq.). Après purification par chromatographie rapide sur gel de silice (chloroforme/éthanol, 20:1 → 15:1 → 11:1), le composé **9** est obtenu avec 76% de rendement (2.7 g, 2.8 mmol).

SM (FAB) : $m/z = 956 [M+H]^+$.

RMN ¹H (CD₃OD, 400 MHz) : δ (ppm) = 7.52-6.60 (m, H_{arom.}); 5.02 (m, 1H, CH^{Ile}); 4.68 (m, 1H, CH^{Asp}); 4.36-4.12 (m, 4H, CH^{Leu,Thr,Tyr}); 3.56 (m, 1H, CH^{Ile}); 3.47 (m, 2H, CH₂^{Tryp}); 3.25 (m, 1H, OH^{Thr}); 2.98-2.67 (m, 8H, CH₂^{Tyr,Asp,Ile,Tryp}); 1.68 (m, 3H, CH^{Leu}, CH₂^{Leu}); 1.30 (m, 9H, CH_{3,Boc}^{Tyr}); 1.09 (m, 3H, CH₃^{Thr}); 0.85 (m, 12H, CH₂^{Leu,Ile}).

RMN ¹³C (CD₃OD, 400 MHz) : δ (ppm) = 176.9-172.4 (CO); 157.2-112.2 (C_{arom.}); 80.9 (C_{Boc}^{Tyr}); 68.6-67.6 (CH^{Ile,Thr}, CH₂^{Asp}); 60.4-57.7 (CH^{Thr,Tyr}); 54.1-50.7 (CH^{Ile,Leu,Asp}); 41.5 (CH₂^{Leu,Tryp}); 28.7 (CH_{3,Boc}^{Tyr}); 26.4-25.7 (CH₂^{Leu,Ile,Tryp}); 23.6-20.1 (CH₃^{Leu,Thr}); 15.8 (CH₃^{Ile}); 11.6 (CH₃^{Ile}).

N*²-(*N*²-(*N*²-*tert*-Butyloxycarbonyl-L-tyrosyl-L-isoleucyl)-L-aspartyl-L-leucyl)-*N*¹-indolyléthyl-L-thréonine [Boc-Tyr-Ile-Asp-Leu-Thr-Tryptamine] **10*

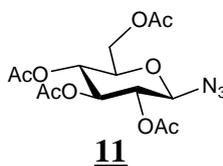


Le composé **9** (2.39 g, 2.5 mmol) est solubilisé dans 30 mL de méthanol. A cette solution est ajoutée une quantité catalytique d'hydroxyde de palladium sur charbon. Le mélange réactionnel est alors agité vigoureusement sous atmosphère d'hydrogène à température ambiante pendant 16 h, puis est filtré sur célite et concentré. Après purification par chromatographie rapide sur gel de silice (chloroforme/méthanol, 12:1 → 5:1 → 2:1), le composé **10** est obtenu avec un rendement quantitatif (2.1 g).

SM (ES) : $m/z = 864 [M-H]^-$.

V.2.2.2 Dérivés du D-glucose

Azide de 2,3,4,6-tétra-*O*-acétyl-β-D-glucoopyranosyle **11**



Dans une première étape, à une solution refroidie à 0°C de D-glucose per-*O*-acétylé commercial (2.2 g, 5.6 mmol) dans 25 mL de dichlorométhane anhydre sont ajoutés lentement 25 mL d'une solution à 33% de bromure d'hydrogène dans l'acide acétique. Le mélange réactionnel est alors agité à 0°C pendant 2 h puis dilué par du dichlorométhane. La phase organique est ensuite lavée successivement à l'eau glacée puis 3 fois par une solution saturée

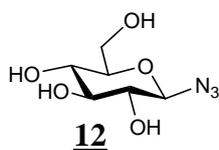
d'hydrogénocarbonate de sodium glacée. Elle est ensuite séchée sur sulfate de sodium anhydre et enfin concentrée. Le dérivé bromé est utilisé sans autre purification.

Dans une seconde étape, le résidu précédemment obtenu est repris dans 57 mL de *N,N*-diméthylformamide anhydre. A cette solution est alors ajouté de l'azoture de sodium (3.66 g, 56.4 mmol, 10 éq.) et la réaction est maintenue à température ambiante pendant 16 h sous agitation vigoureuse. Le mélange réactionnel est alors versé dans de l'eau puis la phase aqueuse est extraite par du dichlorométhane (3 fois). Les phases organiques sont rassemblées, séchées sur sulfate de sodium anhydre puis concentrées. Après purification par chromatographie rapide sur gel de silice (éther de pétrole/acétate d'éthyle, 3:1 → 3:2) et recristallisation dans l'éthanol absolu, le composé **11** [195] est obtenu avec un rendement de 84% (1.75 g, 4.7 mmol).

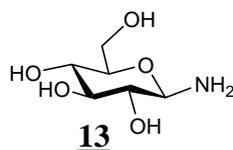
RMN ¹H (CDCl₃, 300 MHz) : δ (ppm) = 5.10 (t, $J_{2,3} = J_{3,4} = 9.5$ Hz, 1H, H-3); 4.96 (t, $J_{4,5} = 9.9$ Hz, 1H, H-4); 4.81 (t, $J_{1,2} = 8.9$ Hz, 1H, H-2); 4.56 (d, 1H, H-1); 4.14 (dd, $J_{5,6a} = 4.8$ Hz, $J_{6a,6b} = 12.4$ Hz, 1H, H-6a); 4.03 (dd, $J_{5,6b} = 4.8$ Hz, 1H, H-6b); 3.71 (m, 1H, H-5); 1.95-1.86 (m, 12H, COCH₃).

RMN ¹³C (CDCl₃, 300 MHz) : δ (ppm) = 170.2-168.8 (COCH₃); 87.5 (C-1); 73.7 (C-5); 72.3 (C-3); 70.4 (C-2); 67.7 (C-4); 61.4 (C-6); 20.3-20.1 (COCH₃).

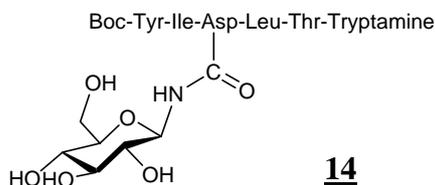
Azide de β-D-glucopyranosyle **12**



Le composé **11** (409 mg, 1.09 mmol, 1 éq.) en solution à 0°C dans 25 mL de méthanol est désacétylé par ajout d'une solution de méthylate de sodium 1 M (550 μL, 0.55 mmol, 0.5 éq.). Le mélange réactionnel est agité 30 min à 0°C puis 30 min à température ambiante. Il est ensuite neutralisé avec de la résine Amberlite IR-120 (H⁺), filtré sur coton et concentré. Le résidu est repris dans l'eau. Après lyophilisation, le composé **12** est obtenu avec un rendement quantitatif (225 mg).

β -D-Glucopyranosylamine 13

A une solution du composé 12 (225 mg, 1.09 mmol) dans 15 mL de méthanol est ajoutée une quantité catalytique d'hydroxyde de palladium sur charbon. Le mélange réactionnel est agité vigoureusement sous atmosphère d'hydrogène à température ambiante pendant 3 h, puis est filtré sur célite et concentré. Le composé 13, obtenu avec un rendement quantitatif (197 mg), est utilisé sans autre purification.

V.2.2.3 Synthèse du dérivé glucopeptidique **N^2 -(N^4 -(β -D-Glucopyranosyl)- N^2 -(*N*-*tert*-butyloxycarbonyl-L-tyrosyl-L-isoleucyl)-L-asparaginyll-L-leucyl)- N^1 -indolyléthyl-L-thréonine 14**

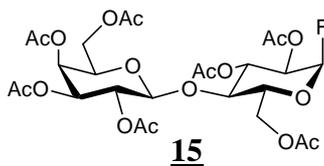
A une solution du composé 10 (316 mg, 0.365 mmol, 1 éq.) dans 13 mL de *N,N*-diméthylformamide sont ajoutés la *N*-hydroxybenzotriazole sous forme hydratée (49.3 mg, 0.365 mmol, 1 éq.) et le BOP (0.485 g, 1.09 mmol, 3 éq.). Le mélange réactionnel est agité pendant 30 min à température ambiante puis une solution du composé 13 (196 mg, 1.09 mmol, 3 éq.) dans 12 mL de *N,N*-diméthylformamide est ajoutée. Le mélange réactionnel est agité à température ambiante pendant 20 h puis concentré et co-éaporé au toluène. Le résidu est suspendu dans l'eau et lyophilisé. Le lyophilisat est alors lavé sur fritté avec de l'acétate d'éthyle. Après purification par chromatographie sur colonne ouverte de gel de silice (acétonitrile puis acétonitrile/eau, 98:2 \rightarrow 96:4), le composé 14 est obtenu avec 15% de rendement (56.3 mg, 55 μ mol).

SM (MALDI) : $m/z = 1049 [M+Na]^+$.

RMN 1H (CD₃OD, 400 MHz) : δ (ppm) = 7.80-6.69 (m, H_{arom.}); 4.91 ppm (d, $J_{1,2} = 9.0$ Hz, 1H, H-1); 4.48-2.62 (m, H-2, H-3, H-4, H-5, H-6a, H-6b, H-peptide); 1.86-0.87 (m, H-peptide).

V.2.2.4 Dérivés du lactose

Fluorure de (2,3,4,6-tétra-*O*-acétyl- β -D-galactopyranosyl)-(1 \rightarrow 4)-2,3,6-tri-*O*-acétyl- α -D-glucopyranosyle **15**



Du lactose per-*O*-acétylé majoritairement d'anométrie β (5.01 g, 7.38 mmol) est placé dans un récipient en plastique puis traité par 20 mL d'un mélange fluorure d'hydrogène/pyridine (7:3, v/v) pendant 2 heures à 0°C. Le milieu réactionnel est dilué avec du dichlorométhane puis versé dans une solution glacée d'ammoniaque (3M). La phase organique est récupérée et lavée avec une solution saturée d'hydrogénocarbonate de sodium glacée (3 fois), séchée sur sulfate de sodium anhydre et concentrée.

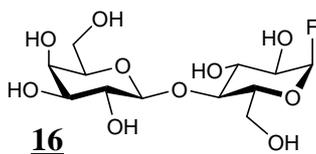
Le résidu est repris dans 15 mL de *N,N*-diméthylformamide puis chauffé à 50°C. De l'acétate d'hydrazine (51 mg, 0.55 mmol) est alors ajouté. Après 20 min à 50°C, le milieu réactionnel est dilué avec de l'acétate d'éthyle. La phase organique est alors lavée avec une solution saturée de chlorure de sodium (2 fois), séchée sur sulfate de sodium anhydre, finalement concentrée et co-évaporée plusieurs fois au toluène. Après purification par chromatographie rapide sur gel de silice (acétate d'éthyle/ éther de pétrole, 1:1), le fluorure **15** [204] est obtenu avec 80% de rendement (3.77 g, 5.9 mmol).

SM (DCI) : $m/z = 656 [M+NH_4]^+$.

RMN ^1H (CDCl_3 , 300 MHz) : δ (ppm) = 5.61 (dd, $J_{1,2} = 2.6$ Hz, $J_{\text{H,F}} = 53.0$ Hz, 1H, H-1^I); 5.43 (t, $J_{2,3} = J_{3,4} = 9.9$ Hz, 1H, H-3^I); 5.30 (d, $J_{3,4} = 3.3$ Hz, 1H, H-4^{II}); 5.05 (dd, $J_{1,2} = 7.7$ Hz, $J_{2,3} = 10.2$ Hz, 1H, H-2^{II}); 4.90 (dd, 1H, H-3^{II}); 4.83 (ddd, $J_{\text{H,F}} = 24.1$ Hz, 1H, H-2^I); 4.46 (d, 1H, H-1^{II}); 4.05 (m, 5H, H-4^I, H-6a^{II}, H-6b^{II}); 3.81 (m, 2H, H-5^{II}); 2.10-1.90 (m, 21H, COCH_3).

RMN ^{13}C (CDCl_3 , 75 MHz) : δ (ppm) = 170.2-168.8 (COCH_3); 103.6 (d, $J_{\text{C,F}} = 227.3$ Hz, C-1^I); 100.7 (C-1^{II}); 75.0, 70.9, 70.8, 70.7, 70.2, 69.0, 68.9, 66.6 (C-2^{I,II}, C-3^{I,II}, C-4^{I,II}, C-5^{I,II}); 61.2, 60.8 (C-6^{I,II}); 20.7-20.4 (COCH_3).

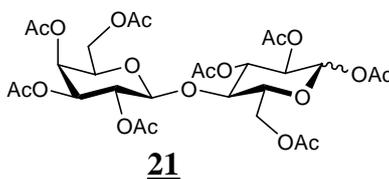
Fluorure de β -D-galactopyranosyl-(1 \rightarrow 4)- α -D-glucopyranosyle **16**



Le fluorure **15** (1.89 g, 2.96 mmol, 1 éq.) en solution à 0°C dans 150 mL de méthanol est désacétylé par ajout d'une solution de méthylate de sodium 1 M (1.5 mL, 1.48 mmol, 0.5 éq.). Le mélange réactionnel est agité 1 h à 0°C puis 1h30 à température ambiante. Il est ensuite neutralisé avec de la résine Amberlite IR-120 (H^+), filtré sur coton et concentré. Le résidu est repris dans l'eau. Après lyophilisation, le composé libre **16** est obtenu avec 98% de rendement (987 mg, 2.9 mmol).

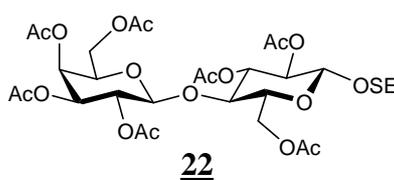
SM (DCI) : $m/z = 362$ $[\text{M}+\text{NH}_4]^+$.

(2,3,4,6-Tétra-*O*-acétyl- β -D-galactopyranosyl)-(1 \rightarrow 4)-1,2,3,6-tétra-*O*-acétyl- α,β -D-glucopyranose 21



Une suspension de lactose monohydraté commercial (20.7 g, 57.45 mmol) et d'acétate de potassium (7 g) dans 60 mL d'anhydride acétique est portée à reflux pendant 15 min. La solution obtenue est alors agitée pendant 24 h à température ambiante puis diluée par du méthanol à froid. Le mélange réactionnel est ensuite concentré et co-évaporé plusieurs fois au toluène. Le résidu est repris dans le dichlorométhane. La phase organique est lavée successivement à l'eau puis avec une solution saturée d'hydrogénocarbonate de sodium et enfin avec une solution saturée de chlorure de sodium. La phase organique est ensuite séchée sur sulfate de sodium anhydre puis concentrée. Après cristallisation dans l'éthanol à 96%, le composé 21 est obtenu avec 70% de rendement (27.1 g, 40.2 mmol).

(2,3,4,6-Tétra-*O*-acétyl- β -D-galactopyranosyl)-(1 \rightarrow 4)-2,3,6-tri-*O*-acétyl- β -D-glucopyranoside de 2-(triméthylsilyl)éthyle 22



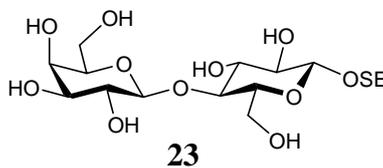
Dans une première étape, à une solution refroidie à 0°C de lactose per-*O*-acétylé 21 (7.4 g, 10.9 mmol, 1 éq.) dans 50 mL de dichlorométhane anhydre sont ajoutés lentement 50 mL d'une solution à 33% de bromure d'hydrogène dans l'acide acétique. Le mélange réactionnel est alors agité à 0°C pendant 30 min puis 1 h à température ambiante. Il est alors dilué par du dichlorométhane. La phase organique est ensuite lavée successivement à l'eau glacée puis 3 fois par une solution saturée d'hydrogénocarbonate de sodium glacée. Elle est ensuite séchée sur sulfate de sodium anhydre et enfin concentrée. Le dérivé bromé obtenu est utilisé sans autre purification.

Dans une seconde étape, à une suspension de tamis moléculaire 4 Å (10 g) préalablement activé (séchage à chaud sous vide), dans 50 mL de dichlorométhane anhydre contenant du carbonate d'argent (3 g, 10.9 mmol, 1 éq.) et du perchlorate d'argent (2.26 g, 10.9 mmol, 1 éq.) sont alors ajoutés le 2-(triméthylsilyl)éthanol (3.93 mL, 27.3 mmol, 2.5 éq.) puis le bromure précédemment préparé en solution dans 40 mL de dichlorométhane anhydre. Le mélange réactionnel est agité vigoureusement à température ambiante sous argon à l'abri de la lumière pendant 3.5 h. Il est ensuite filtré sur célite et lavé avec du dichlorométhane. Le filtrat est finalement concentré. Après purification par chromatographie rapide sur colonne (éther de pétrole/acétate d'éthyle, 2:1), le composé **22** [211] est obtenu avec 32% de rendement (2.6 g, 3.49 mmol).

RMN ¹H (CDCl₃, 300 MHz) : δ (ppm) = 5.29 (d, $J_{3,4} = 2.6$ Hz, 1H, H-4^{II}); 5.14 (t, $J_{2,3} = J_{3,4} = 9.1$ Hz, 1H, H-3^I); 5.05 (dd, $J_{1,2} = 7.7$ Hz, $J_{2,3} = 10.6$ Hz, 1H, H-2^{II}); 4.89 (dd, 1H, H-3^{II}); 4.82 (dd, $J_{1,2} = 9.5$ Hz, 1H, H-2^I); 4.42 (m, 2H, H-1^{I,II}); 4.05 (m, 4H, H-6a^{I,II}, H-6b^{I,II}); 3.85 (m, 2H, H-5^{II}, OCH₂); 3.74 (t, $J_{4,5} = 9.1$ Hz, 1H, H-4^I); 3.51 (m, 2H, H-5^I, OCH₂); 2.10-1.91 (m, 21H, COCH₃); 0.86 (m, 2H, CH₂SiMe₃); 0.00 (s, 9H, SiMe₃).

RMN ¹³C (CDCl₃, 300 MHz) : δ (ppm) = 170.2-169.0 (COCH₃); 101.0, 99.9 (C-1^{I,II}); 76.4, 73.0, 72.5, 71.8, 71.0, 70.6, 69.1, 67.4, 66.6 (C-2^{I,II}, C-3^{I,II}, C-4^{I,II}, C-5^{I,II}, OCH₂); 62.1, 60.8 (C-6^{I,II}); 20.7-20.4 (COCH₃); 17.8 (CH₂SiMe₃); 0.9 (SiMe₃).

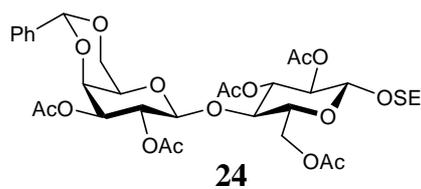
β-D-Galactopyranosyl-(1→4)-β-D-glucopyranoside de 2-(triméthylsilyl)éthyle **23**



Le composé **22** (1.5 g, 2.17 mmol, 1 éq.) en solution à 0°C dans 50 mL de méthanol est désacétylé par ajout d'une solution de méthylate de sodium 1 M (1.1 mL, 1.08 mmol, 0.5 éq.). Le mélange réactionnel est agité 30 min à 0°C puis 3h30 à température ambiante. Il est ensuite neutralisé avec de la résine Amberlite IR-120 (H⁺), filtré sur coton et concentré. Le

résidu est repris dans l'eau. Après lyophilisation, le composé libre **23** [211] est obtenu avec un rendement quantitatif (960 mg).

(2,3-Di-O-acétyl-4,6-O-benzylidène-β-D-galactopyranosyl)-(1→4)-2,3,6-tri-O-acétyl-β-D-glucopyranoside de 2-(triméthylsilyl)éthyle **24**



A une suspension du composé **23** (240.3 mg, 0.543 mmol, 1 éq.) dans 8.4 mL d'acétonitrile anhydre sont ajoutés l' α,α -diméthoxytoluène (163 μ L, 1.086 mmol, 2 éq.) et une quantité catalytique de l'acide *p*-toluène sulfonique (4.8 mg). Le mélange réactionnel est agité pendant 1.5 h à température ambiante puis de la triéthylamine est ajoutée pour rendre le milieu basique. Le mélange réactionnel est ensuite concentré. Le résidu est repris dans 4.8 mL de pyridine, puis sont ajoutés 4.8 mL d'anhydride acétique ainsi qu'une quantité catalytique de DMAP (10 mg). Le mélange réactionnel est agité à l'abri de la lumière pendant 16 h à température ambiante. L'excès d'anhydride acétique est détruit par addition de méthanol à froid. Le mélange réactionnel est alors concentré et co-évaporé plusieurs fois au toluène. Après purification par chromatographie rapide sur gel de silice (éther de pétrole/acétate d'éthyle, 2:1), le composé **24** [210] est obtenu avec 81% de rendement (326 mg, 0.434 mmol).

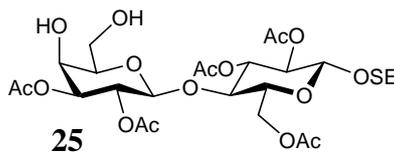
SM (DCI) : $m/z = 758 [M+NH_4]^+$.

RMN 1H (CDCl₃, 300 MHz) : δ (ppm) = 7.36-7.23 (m, 5H, H_{arom.} de Ph); 5.37 (s, 1H, PhCH); 5.10 (m, 2H, H-2^{II}, H-3^I); 4.80 (m, 2H, H-2^I, H-3^{II}); 4.39 (m, 2H, H-1^{I,II}); 4.20 (d, $J_{3,4} = 3.3$ Hz, 1H, H-4^{II}); 4.13-3.78 (m, 6H, OCH₂, H-6a^{I,II}, H-6b^{I,II}); 3.68 (t, $J_{3,4} = 9.5$ Hz, 1H, H-4^I); 3.53-3.36 (m, 2H, H-5^{I,II}); 1.97-1.89 (m, 15H, COCH₃); 0.80 (m, 2H, CH₂SiMe₃); 0.00 (s, 9H, SiMe₃).

RMN ^{13}C (CDCl₃, 300 MHz) : δ (ppm) = 170.5-168.5 (COCH₃); 137.4, 128.7, 127.8, 126.1 (C_{arom.} de Ph); 100.7, 100.6, 99.7 (C-1^{I,II}, PhCH); 75.9, 72.8, 72.4, 71.6, 71.3, 68.8, 68.0, 66.9,

66.0 (C-2^{I,II}, C-3^{I,II}, C-4^{I,II}, C-5^{I,II}, OCH₂); 61.9, 59.9 (C-6^{I,II}); 20.5-20.2 (COCH₃); 17.5 (CH₂SiMe₃); 0.9 (SiMe₃).

(2,3-Di-O-acétyl-β-D-galactopyranosyl)-(1→4)-2,3,6-tri-O-acétyl-β-D-glucopyranoside de 2-(triméthylsilyl)éthyle 25



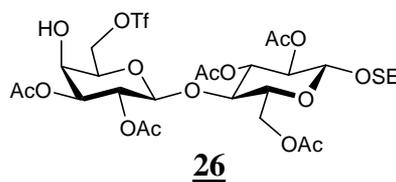
Le composé 24 (438.5 mg, 0.592 mmol) est solubilisé dans 11 mL d'une solution d'acide acétique dans l'eau à 80%. Le mélange réactionnel est agité à 50°C pendant 9 h puis concentré et co-évaporé plusieurs fois au toluène. Après purification par chromatographie rapide sur gel de silice (éther de pétrole/acétate d'éthyle, 1:1 → 1:2 → 1:3), le composé 25 [210] est obtenu avec 60% de rendement (233 mg, 0.358 mmol).

SM (DCI) : m/z = 670 [M+NH₄]⁺.

RMN ¹H (CDCl₃, 300 MHz) : δ (ppm) = 5.19 (m, 2H, H-2^{II}, H-3^I); 4.85 (m, 2H, H-2^I, H-3^{II}); 4.46 (m, 3H, H-1^{I,II}, H-6a^I); 4.06 (m, 2H, H-4^{II}, H-6b^I); 3.94-3.76 (m, 4H, H-4^I, H-5^{II}, H-6a^{II}, OCH₂); 3.62-3.48 (m, 3H, H-5^I, H-6b^{II}, OCH₂); 2.07-1.99 (m, 15H, COCH₃); 0.86 (m, 2H, CH₂SiMe₃); 0.00 (s, 9H, SiMe₃).

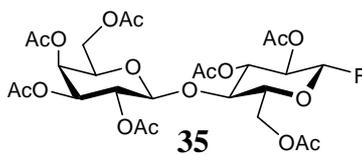
RMN ¹³C (CDCl₃, 300 MHz) : δ (ppm) = 170.7-169.5 (COCH₃); 101.1, 99.9 (C-1^{I,II}); 76.5, 74.5, 73.5, 73.4, 72.5, 71.8, 69.7, 67.8, 67.4 (C-2^{I,II}, C-3^{I,II}, C-4^{I,II}, C-5^{I,II}, OCH₂); 62.4, 62.0 (C-6^{I,II}); 20.9-20.6 (COCH₃); 17.9 (CH₂SiMe₃); 0.90 (SiMe₃).

(2,3-Di-*O*-acétyl-6-*O*-trifluorométhanesulfonyl-β-D-galactopyranosyl)-(1→4)-2,3,6-tri-*O*-acétyl-β-D-glucopyranoside de 2-(triméthylsilyl)éthyle 26



Le diol 25 (64 mg, 98 μmol, 1 éq.) est solubilisé dans 1.7 mL de dichlorométhane. A cette solution est ajoutée de la pyridine (32 μL, 0.382 mmol, 3.9 éq.). Le mélange réactionnel, gardé sous argon, est refroidi à -30°C puis l'anhydride triflique (18 μL, 0.108 mmol, 1.1 éq.) est ajouté goutte à goutte. Le mélange réactionnel est agité pendant 1 h à -30°C puis dilué avec du dichlorométhane. La phase organique est lavée successivement avec une solution d'acide chlorhydrique 1 M, une solution saturée d'hydrogénocarbonate de sodium et à l'eau. Elle est ensuite séchée sur sulfate de sodium anhydre puis concentrée. Le triflate 26 [212] (77 mg) est utilisé directement sans autre purification dans l'étape suivante.

Fluorure de (2,3,4,6-tétra-*O*-acétyl-β-D-galactopyranosyl)-(1→4)-2,3,6-tri-*O*-acétyl-β-D-glucopyranosyle 35



Dans une première étape, à une solution refroidie à 0°C de lactose per-*O*-acétylé 21 (4.24 g, 6.2 mmol, 1 éq.) dans 30 mL de dichlorométhane anhydre sont ajoutés lentement 30 mL d'une solution à 33% de bromure d'hydrogène dans l'acide acétique. Le mélange réactionnel est alors agité à 0°C pendant 2 h, puis dilué par du dichlorométhane. La phase organique est ensuite lavée successivement à l'eau glacée puis 3 fois par une solution saturée d'hydrogénocarbonate de sodium glacée. Elle est ensuite séchée sur sulfate de sodium anhydre et enfin concentrée. Le dérivé bromé obtenu est utilisé sans autre purification.

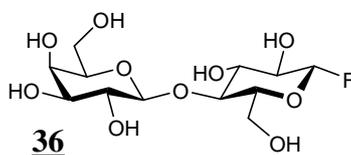
Dans une seconde étape, à une solution du bromure précédent dans 50 mL d'acétonitrile anhydre est ajouté le fluorure d'argent (1.58 g, 12.5 mmol, 2 éq.). Le mélange réactionnel est alors agité vigoureusement à l'abri de la lumière pendant 3 h à température

ambiante. Il est ensuite filtré deux fois sur célite puis concentré. Le résidu est repris dans du dichlorométhane puis lavé successivement par une solution saturée d'hydrogencarbonate de sodium glacée (2 fois) et à l'eau glacée. La phase organique est alors séchée sur sulfate de sodium anhydre puis concentrée. Après purification par chromatographie rapide sur gel de silice (éther de pétrole/acétate d'éthyle, 3:2), le composé **35** [237] est obtenu avec 83% de rendement (3.3 g, 5.2 mmol).

RMN ^1H (CDCl_3 , 300 MHz) : δ (ppm) = 5.36 (dd, $J_{1,2} = 5.1$ Hz, $J_{\text{H,F}} = 53.0$ Hz, 1H, H-1^I); 4.52 (d, $J_{1,2} = 7.7$ Hz, 1H, H-1^{II}); 2.12-1.94 (m, 21H, COCH_3).

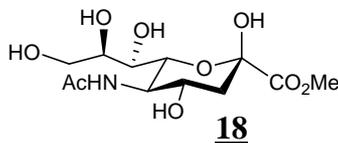
RMN ^{13}C (CDCl_3 , 300 MHz) : δ (ppm) = 170.2-169.1 (COCH_3); 106.0 (d, $J_{\text{C,F}} = 216.9$ Hz, C-1^I); 101.3 (C-1^{II}); 75.2, 72.5, 72.2, 72.1, 71.4, 70.9, 69.1, 66.7 (C-2^{I,II}, C-3^{I,II}, C-4^{I,II}, C-5^{I,II}); 61.9, 60.8 (C-6^{I,II}); 20.7-20.4 (COCH_3).

Fluorure de β -D-galactopyranosyl-(1 \rightarrow 4)- β -D-glucopyranosyle **36**

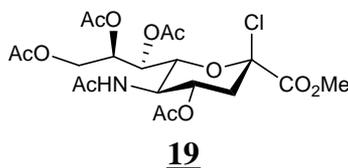


Le composé **35** (1.79 g, 2.81 mmol, 1 éq.) en solution à 0°C dans 150 mL de méthanol est désacétylé par ajout d'une solution de méthylate de sodium 1 M (1.4 mL, 1.4 mmol, 0.5 éq.). Le mélange réactionnel est agité 4.5 h à 0°C. Il est ensuite neutralisé avec de la résine Amberlite IR-120 (H^+), filtré sur coton et concentré. Le résidu est repris dans l'eau et congelé immédiatement. Après lyophilisation, le composé libre **36** est obtenu avec un rendement quantitatif (970 mg).

V.2.2.5 Dérivés de l'acide neuraminique

5-Acétamido-3,5-didésoxy-D-glycero-β-D-galacto-non-2-ulopyranosylonate de méthyle 18

A une suspension de l'acide *N*-acétylneuraminique (1.15 g, 3.72 mmol) dans 20 mL de méthanol est ajoutée de la résine acide Amberlite IR-120 (0.77 g). Le mélange réactionnel est agité vigoureusement pendant une nuit à température ambiante. La solution limpide obtenue est ensuite filtrée sur coton. Le filtrat est concentré jusqu'à un volume d'environ 2 mL puis de l'éther diéthylique est ajouté pour cristalliser le produit **18** [238], qui est obtenu avec un rendement de 80% (0.96 g, 2.97 mmol).

5-Acétamido-4,7,8,9-tétra-*O*-acétyl-2-chloro-2,3,5-tridésoxy-D-glycero-β-D-galacto-non-2-ulopyranosylonate de méthyle 19

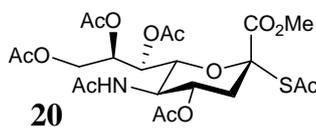
Le composé **18** (0.86 g, 2.66 mmol) est ajouté à une solution de chlorure d'acétyle (19 mL) et d'acide acétique (9.5 mL) préalablement refroidie à 0°C. Le ballon est ensuite fermé hermétiquement par un septum. Le mélange réactionnel est agité pendant 48 h à température ambiante, puis concentré et co-évalué plusieurs fois au toluène. Après purification par chromatographie rapide sur gel de silice (acétate d'éthyle), le composé **19** [239] est obtenu avec 75% de rendement (1.01 g, 2 mmol).

RMN ¹H (CDCl₃, 300 MHz) : δ (ppm) = 6.63 (d, $J_{5,NH} = 10.2$ Hz, 1H, NH); 5.33 (dd, $J_{6,7} = 2.2$ Hz, $J_{7,8} = 6.6$ Hz, 1H, H-7); 5.24 (m, $J_{3e,4} = 4.8$ Hz, $J_{3a,4} = J_{4,5} = 10.6$ Hz, 1H, H-4); 5.02 (m, $J_{8,9} = 2.9$ Hz, $J_{8,9'} = 6.2$ Hz, 1H, H-8); 4.31 (m, 1H, H-9); 4.27 (m, 1H, H-6); 4.08 (t, $J_{5,6} =$

10.6 Hz, 1H, H-5); 3.93 (dd, $J_{9,9'} = 12.8$ Hz, 1H, H-9'); 3.72 (s, 3H, OMe); 2.64 (dd, $J_{3a,3e} = 13.9$ Hz, 1H, H-3e); 2.09 (dd, 1H, H-3a); 1.97-1.78 (m, 15H, COCH₃).

RMN ¹³C (CDCl₃, 300 MHz) : δ (ppm) = 170.8-169.6 (COCH₃); 165.3 (C-1); 96.4 (C-2); 73.5 (C-6); 69.9 (C-8); 68.5 (C-4); 66.8 (C-7); 61.9 (C-9); 53.4 (OMe); 48.2 (C-5); 40.3 (C-3); 22.6 (NCOCH₃); 20.5-20.4 (OCOCH₃).

5-Acétamido-4,7,8,9-tétra-O-acétyl-2-S-acétyl-3,5-didésoxy-2-thio-D-glycéro- α -D-galactonon-2-ulopyranosylonate de méthyle 20



A une solution du composé **19** (1.01 g, 1.99 mmol, 1 éq.) dans un mélange de 10 mL d'acétate d'éthyle (5 mL/mmol de composé **19**) et de 10 mL d'une solution de carbonate de sodium 1 M sont ajoutés l'hydrogénosulfate de tétrabutylammonium (0.67 g, 1.99 mmol, 1 éq.) et le thioacétate de potassium (0.27 g, 2.39 mmol, 1.2 éq.). Le mélange réactionnel est agité vigoureusement pendant 45 min à température ambiante, puis dilué à l'acétate d'éthyle. La phase organique est séparée et lavée successivement avec une solution saturée d'hydrogénocarbonate de sodium, à l'eau et avec une solution saturée de chlorure de sodium, séchée sur sulfate de sodium anhydre et concentrée. Après purification par chromatographie rapide sur gel de silice (dichlorométhane/acétone, 8:2), le composé **20** [240] est obtenu avec 76% de rendement (830 mg, 1.52 mmol).

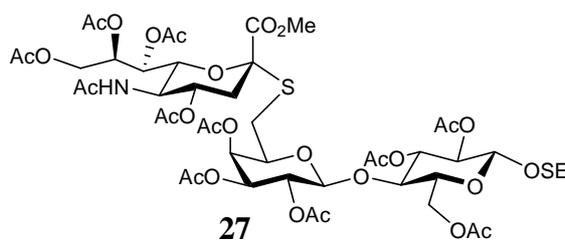
SM (FAB) : $m/z = 572$ [M+Na]⁺.

RMN ¹H (CDCl₃, 300 MHz) : δ (ppm) = 5.58 (d, $J_{5,NH} = 10.2$ Hz, 1H, NH); 5.44 (m, 1H, H-7); 5.27 (m, 1H, H-8); 4.83 (m, 1H, H-4); 4.58 (m, 1H, H-6); 4.33 (m, 1H, H-9); 4.13 (q, 1H, H-5); 4.02 (m, 1H, H-9'); 3.73 (s, 3H, OMe); 2.56 (m, 1H, H-3e); 2.20 (m, 3H, SCOCH₃); 2.07-1.98 (m, 3H, COCH₃); 1.86 (m, 1H, H-3a); 1.85 (s, 3H, NCOCH₃).

RMN ^{13}C (CDCl_3 , 300 MHz) : δ (ppm) = 170.7-170.0 (COCH₃); 161.5 (C-1); 84.5 (C-2); 75.1 (C-6); 70.4 (C-8); 68.9 (C-4); 67.8 (C-7); 62.4 (C-9); 53.4 (OMe); 49.0 (C-5); 37.5 (C-3); 30.0 (SCOCH₃); 23.1 (NCOCH₃); 21.0-20.7 (OCOCH₃).

V.2.2.6 Synthèse et fonctionnalisation du trisaccharide sialylé

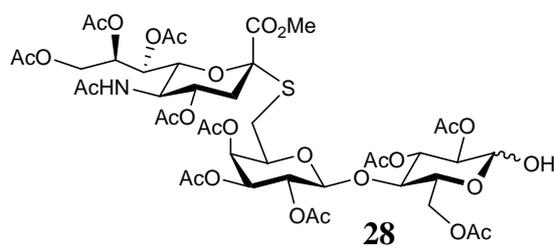
S*-(Méthoxyoxycarbonyl-5-acétamido-4,7,8,9-tétra-*O*-acétyl-3,5-didésoxy-*D*-glycéro- α -*D*-galacto-non-2-ulopyranosylonyl)-(2 \rightarrow 6)-(2,3-di-*O*-acétyl-6-thio- β -*D*-galactopyranosyl)-(1 \rightarrow 4)-2,3,6-tri-*O*-acétyl- β -*D*-glucopyranoside de 2-(triméthylsilyl)éthyle **27*



Le triflate **26** (77 mg, 98 μmol , 1 éq.) et le dérivé sialylé **20** (55 mg, 100 μmol , 1 éq.) sont solubilisés dans 800 μL de *N,N*-diméthylformamide et la solution sous argon est refroidie à -30°C . De la diéthylamine (103 μL , 0.98 mmol, 10 éq.) est ajoutée à la solution. Le milieu réactionnel est agité pendant 2 h à -30°C puis concentré. Le résidu est repris dans 100 μL de pyridine, puis sont ajoutés 100 μL d'anhydride acétique ainsi qu'une quantité catalytique de DMAP (10 mg). Le mélange réactionnel est agité à l'abri de la lumière pendant 16 h à température ambiante. L'excès d'anhydride acétique est détruit par addition de méthanol à froid. Le mélange réactionnel est alors concentré et co-évaporé plusieurs fois au toluène. Après purification par chromatographie rapide sur gel de silice (dichlorométhane puis dichlorométhane/méthanol, 99:1 \rightarrow 94:6), le composé **27** [**212**] est obtenu avec 50% de rendement (60 mg, 50.6 μmol).

SM (MALDI) : $m/z = 1206$ $[\text{M}+\text{Na}]^+$.

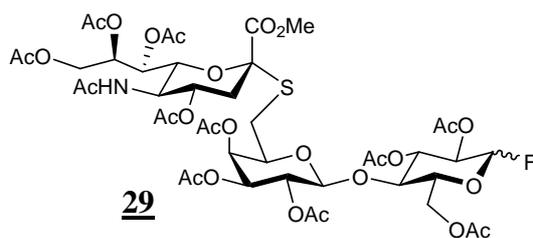
***S*-(Méthyloxy-carbonyl-5-acétamido-4,7,8,9-tétra-*O*-acétyl-3,5-didésoxy-*D*-glycéro- α -*D*-galacto-non-2-ulopyranosylonyl)-(2 \rightarrow 6)-(2,3-di-*O*-acétyl-6-thio- β -*D*-galactopyranosyl)-(1 \rightarrow 4)-2,3,6-tri-*O*-acétyl- α,β -*D*-glucopyranose 28**



A une solution du composé 27 (49 mg, 41.4 μ mol) dans 500 μ L de dichlorométhane anhydre est ajouté 1 mL d'acide trifluoroacétique. Le mélange réactionnel est agité à température ambiante pendant 3 h puis concentré et co-éaporé plusieurs fois au toluène. Le composé 28 [241] (45 mg) est directement utilisé dans l'étape suivante sans autre purification.

SM (MALDI) : $m/z = 1106 [M+Na]^+$.

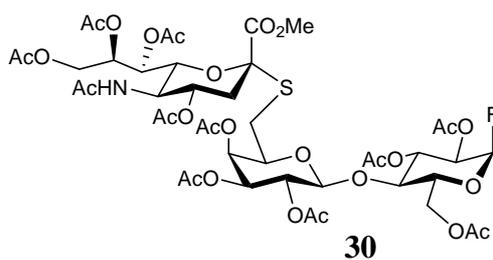
Fluorure de *S*-(méthyloxy-carbonyl-5-acétamido-4,7,8,9-tétra-*O*-acétyl-3,5-didésoxy-*D*-glycéro- α -*D*-galacto-non-2-ulopyranosylonyl)-(2 \rightarrow 6)-(2,3-di-*O*-acétyl-6-thio- β -*D*-galactopyranosyl)-(1 \rightarrow 4)-2,3,6-tri-*O*-acétyl- α,β -*D*-glucopyranosyle 29



A une solution refroidie à -30°C de l'hémiacétal 28 (45 mg, 41.4 μ mol, 1 éq.) dans 2 mL de dichlorométhane est ajouté du DAST (50 μ L, 0.366 mmol, 10 éq.). Le mélange réactionnel est agité pendant 10 h à -30°C . Le mélange réactionnel est concentré après addition de quelques gouttes de méthanol pour détruire l'excès de DAST. Après purification par chromatographie rapide sur gel de silice (dichlorométhane puis dichlorométhane/méthanol, 97:3), le composé 29 est obtenu avec 90% de rendement sur les deux dernières étapes (40 mg, 36.8 μ mol)

SM (MALDI) : $m/z = 1108 [M+Na]^+$.

Fluorure de *S*-(méthyloxycarbonyl-5-acétamido-4,7,8,9-tétra-*O*-acétyl-3,5-didésoxy-*D*-glycéro- α -*D*-galacto-non-2-ulopyranosylonyl)-(2 \rightarrow 6)-(2,3-di-*O*-acétyl-6-thio- β -*D*-galactopyranosyl)-(1 \rightarrow 4)-2,3,6-tri-*O*-acétyl- α -*D*-glucopyranosyle 30



Un mélange fluorure d'hydrogène/pyridine (1.2 mL; 7:3, v/v) est ajouté au fluorure **29** (48 mg, 44 μ mol), placé dans un récipient en plastique à -50°C . Le mélange réactionnel est agité à -50°C pendant 15 min puis lentement porté à -10°C pendant 2 h et finalement agité à -10°C pendant 1 h. Le milieu réactionnel est dilué avec du dichlorométhane puis versé dans une solution glacée d'ammoniaque (3M). La phase organique est récupérée et lavée avec une solution saturée d'hydrogénocarbonate de sodium glacée (3 fois), séchée sur sulfate de sodium anhydre et concentrée. Après purification par chromatographie rapide sur gel de silice (dichlorométhane/méthanol, 98:2), le composé **30** est obtenu avec 52% de rendement (24.8 mg, 23 μ mol).

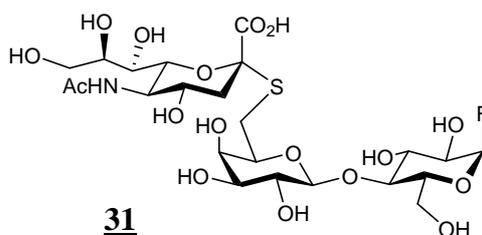
$[\alpha]_D : +22$ (c 1.0, chloroforme).

RMN ^1H (CDCl_3 , 400 MHz) : δ (ppm) = 5.65 (dd, $J_{1,2} = 2.7$ Hz, $J_{\text{H,F}} = 53.0$ Hz, 1H, H-1^I); 5.47 (m, 1H, H-4^{II}); 5.45 (t, $J_{2,3} = J_{3,4} = 9.6$ Hz, 1H, H-3^I); 5.28 (dd, $J_{6,7} = 1.6$ Hz, $J_{7,8} = 9.8$ Hz, 1H, H-7^{III}); 5.19 (d, $J_{5,\text{NH}} = 9.0$ Hz, 1H, NH); 5.17 (m, 1H, H-8^{III}); 5.03 (dd, $J_{1,2} = 7.6$ Hz, $J_{2,3} = 10.4$ Hz, 1H, H-2^{II}); 4.96 (dd, 1H, H-3^{II}); 4.90 (m, 2H, H-2^I, H-4^{III}); 4.63 (d, 1H, H-1^{II}); 4.48 (m, 1H, H-6a^I); 4.24 (dd, $J_{8,9} = 2.9$ Hz, $J_{9,9'} = 12.5$ Hz, 1H, H-9^{III}); 4.12 (m, 3H, H-5^I, H-6b^I, H-9^{III}); 3.93-3.81 (m, 4H, H-4^I, H-5^{II}, H-5^{III}, H-6^{III}); 3.81 (s, 3H, OMe); 2.79 (dd, $J_{5,6a} = 7.4$ Hz, $J_{6a,6b} = 14.3$ Hz, 1H, H-6a^{II}); 2.68 (dd, $J_{3e,4} = 4.7$ Hz, $J_{3a,3e} = 12.8$ Hz, 1H, H-3e^{III});

2.53 (dd, $J_{5,6b} = 7.0$ Hz, 1H, H-6b^{II}); 2.14-1.99 (m, 27H, COCH₃); 1.97 (m, 1H, H-3a^{III}); 1.90-1.86 (m, 6H, COCH₃).

RMN ¹³C (CDCl₃, 400 MHz) : δ (ppm) = 170.8-169.1 (COCH₃); 167.9 (C-1^{III}); 103.5 (d, $J_{C,F} = 227.9$ Hz, C-1^I); 100.1 (C-1^{II}); 83.9 (C-2^{III}); 74.8 (C-4^I); 73.8 (C-6^{III}); 72.6 (C-5^{II}); 71.2 (C-3^{II}); 71.0 (d, $J_{C,F} = 4.2$ Hz, C-5^I); 70.5 (d, $J_{C,F} = 24.4$ Hz, C-2^I); 69.1, 69.0 (C-2^{II}, C-4^{III}); 68.8 (C-3^I); 68.0 (C-4^{II}); 67.6 (C-8^{III}); 66.7 (C-7^{III}); 62.1 (C-9^{III}); 61.6 (C-6^I); 53.2 (OMe); 49.5 (C-5^{III}); 38.0 (C-3^{III}); 29.7 (C-6^{II}); 23.2 (NCOCH₃); 21.1-20.5 (OCOCH₃).

Fluorure de S-(carboxy-5-acétamido-3,5-didésoxy-D-glycéro- α -D-galacto-non-2-uloypyranosylonyl)-(2 \rightarrow 6)-(6-thio- β -D-galactopyranosyl)-(1 \rightarrow 4)- α -D-glucopyranosyle 31



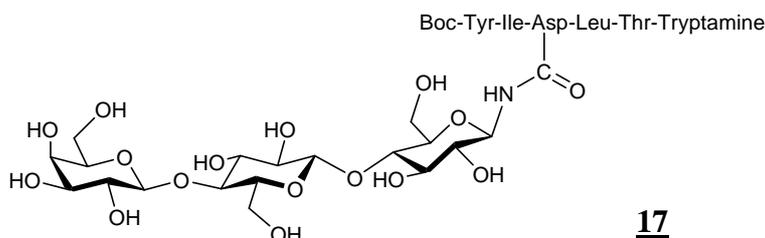
Dans une première étape, le composé **30** (5 mg, 4.6 μ mol, 1 éq.) en solution à 0°C dans 1 mL de méthanol est désacétylé par ajout d'une solution de méthylate de sodium 0.05 M (46 μ L, 2.3 μ mol, 0.5 éq.). Le mélange réactionnel est agité 1.5 h à 0°C puis 3h à température ambiante, puis concentré.

Dans une seconde étape, le résidu est repris dans 2 mL d'une solution d'hydroxyde de sodium 0.1 M. Le mélange réactionnel est agité pendant 45 min à température ambiante puis neutralisé avec de la résine Amberlite IR-120 (H⁺), filtré sur coton et concentré. Le résidu est repris dans l'eau. Après lyophilisation, le composé libre **31** est obtenu avec un rendement quantitatif (3 mg).

SM (ES) : $m/z = 650$ [M-H]⁻.

V.2.2.7 *Couplages enzymatiques à l'échelle analytique*

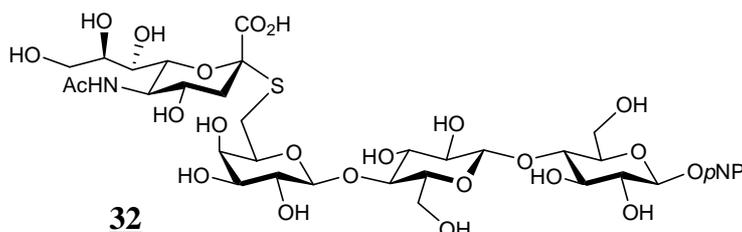
***N*²-(*N*⁴-(β-D-Galactopyranosyl-(1→4)-β-D-glucofuranosyl-(1→4)-β-D-glucofuranosyl)-*N*²-(*N*-*tert*-butyloxycarbonyl-L-tyrosyl-L-isoleucyl)-L-asparaginyl-L-leucyl)-*N*¹-indolyléthyl-L-thréonine 17**



45 µL d'une solution du donneur **16** (12.8 mg/mL) dans le tampon phosphate de sodium 50 mM pH 7 sont mélangés avec 45 µL d'une solution de l'accepteur **14** (13 mg/mL) dans un mélange 1:1 de tampon phosphate de sodium 50 mM pH 7 et d'acétonitrile. A ce mélange est ajoutée une solution de l'enzyme Cel7B E197A d'*Humicola insolens* (10 µL, 0.12 mg) dans le tampon Tris-H₃PO₄ 200 mM pH 7.5. Le mélange réactionnel est maintenu à 40°C. Le suivi de la réaction est réalisé par chromatographie sur couche mince (acétonitrile/eau, 8:2). Après 24 h d'incubation, le mélange réactionnel est lyophilisé, dessalé par traitement avec une résine mixte TMD-8 puis analysé par spectrométrie de masse.

SM (MALDI) : $m/z = 1373 [M+Na]^+$.

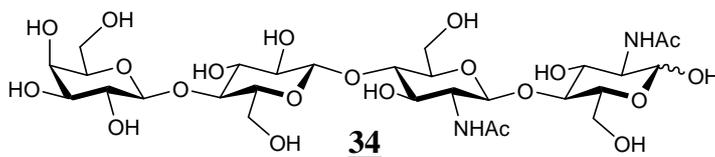
***S*-(Carboxy-5-acétamido-3,5-didésoxy-D-glycéro-α-D-galacto-non-2-ulopyranosylonyl)-(2→6)-(6-thio-β-D-galactopyranosyl)-(1→4)-β-D-glucofuranosyl-(1→4)-β-D-glucofuranoside de *p*-nitrophényle 32**



45 μL d'une solution du donneur **31** (14.3 mg/mL) dans le tampon phosphate de sodium 50 mM pH 7 sont mélangés avec 45 μL d'une solution de β -D-glucopyranoside de *p*-nitrophényle (11 mg/mL) dans le tampon phosphate de sodium 50 mM pH 7. A ce mélange est ajoutée une solution de l'enzyme Cel7B E197A d'*Humicola insolens* (10 μL , 0.12 mg) dans le tampon Tris- H_3PO_4 200 mM pH 7.5. Le mélange réactionnel est maintenu à 40°C. Le suivi de la réaction est réalisé par chromatographie sur couche mince (acétonitrile/eau, 8:2). Après 24 h d'incubation, le mélange réactionnel est lyophilisé, dessalé par traitement avec une résine mixte TMD-8 puis analysé par spectrométrie de masse.

SM (ES) : $m/z = 931$ $[\text{M}-\text{H}]^-$.

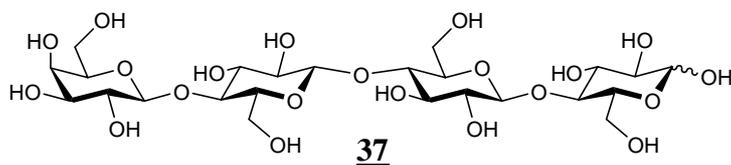
β -D-Galactopyranosyl-(1 \rightarrow 4)- β -D-glucopyranosyl-(1 \rightarrow 4)-(2-acétamido-2-désoxy- β -D-glucopyranosyl)-(1 \rightarrow 4)-2-acétamido-2-désoxy- α,β -D-glucopyranose **34**



Le fluorure **36** (20 mg, 58 μmol , 3 éq.) est solubilisé dans un mélange de tampon acétate de sodium 50 mM pH 5 (200 μL) et d'acétonitrile (600 μL). A cette solution est ajoutée une solution de l'enzyme Cel7B de *Trichoderma reesei* (14.7 μL , 11.7 μg). Après 1 min d'agitation à 40°C, le N^I, N^{II} -diacétylchitobiose (8.2 mg, 19.3 μmol , 1 éq.) en solution dans le tampon acétate de sodium 50 mM pH 5 (200 μL) est ajouté au mélange réactionnel. Il est alors agité à 40°C pendant 40 min. Le pH de la solution est ensuite amené à 9 par l'addition d'ammoniaque concentré et le mélange réactionnel est porté à ébullition pendant 10 min. Après concentration, lyophilisation et dessalage par traitement avec une résine mixte TMD-8, le brut réactionnel est analysé par spectrométrie de masse.

SM (ES) : $m/z = 771$ $[\text{M}+\text{Na}]^+$.

β -D-Galactopyranosyl-(1 \rightarrow 4)- β -D-glucofuranosyl-(1 \rightarrow 4)- β -D-glucofuranosyl-(1 \rightarrow 4)- α,β -D-glucofuranose 37



Couplage avec Cel7B E197A d'*Humicola insolens* : 5 mg du donneur **16** et 5 mg de cellobiose sont solubilisés dans 96 μ L de tampon phosphate de sodium 50 mM pH 7. A ce mélange est ajoutée une solution de l'enzyme Cel7B E197A d'*H. insolens* (4.1 μ L, 49.2 μ g) dans le tampon Tris-H₃PO₄ 200 mM pH 7.5. Le mélange réactionnel est maintenu à 40°C. Le suivi de la réaction est réalisé par chromatographie sur couche mince (acétonitrile/eau, 7:3). Après 96 h d'incubation, le mélange réactionnel est lyophilisé, dessalé par traitement avec une résine mixte TMD-8 puis analysé par spectrométrie de masse.

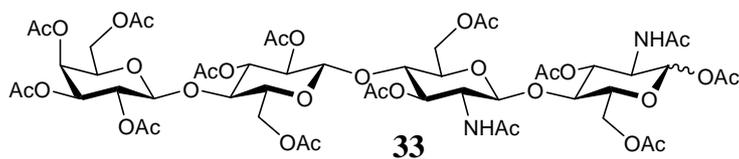
SM (FAB) : $m/z = 689 [M+Na]^+$.

Couplage avec Cel7B E196A de *Trichoderma reesei* : 5 mg du donneur **16** et 5 mg de cellobiose sont solubilisés dans 100 μ L d'une solution de l'enzyme Cel7B E196A de *T. reesei* (concentration en protéines totales de 0.52 mg/mL) dans le tampon acétate de sodium 25 mM pH 4.6. Le mélange réactionnel est maintenu à 40°C. Le suivi de la réaction est réalisé par chromatographie sur couche mince (acétonitrile/eau, 7:3). Après 96 h d'incubation, le mélange réactionnel est lyophilisé, dessalé par traitement avec une résine mixte TMD-8 puis analysé par spectrométrie de masse.

SM (FAB) : $m/z = 689 [M+Na]^+$.

V.2.2.8 Préparation des substrats utilisés pour la caractérisation des enzymes

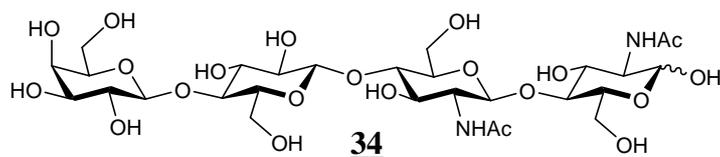
(2,3,4,6-Tétra-O-acétyl-β-D-galactopyranosyl)-(1→4)-(2,3,6-tri-O-acétyl-β-D-glucopyranosyl)-(1→4)-(2-acétamido-3,6-di-O-acétyl-2-désoxy-β-D-glucopyranosyl)-(1→4)-2-acétamido-1,3,6-tri-O-acétyl-2-désoxy-α,β-D-glucopyranose 33



A une solution de fluorure de lactosyle **16** (151.1 mg, 0.44 mmol, 1.2 éq.) dans 6 mL de tampon carbonate-bicarbonate de sodium 50 mM pH 9.2 sont ajoutés le *N*¹-acétylchitobiose (140.7 mg, 0.37 mmol, 1 éq.) et une solution de l'enzyme Cel7B E197A d'*Humicola insolens* (240 μL, 3 mg) dans le tampon Tris-H₃PO₄ 200 mM pH 7.5. Le mélange réactionnel est maintenu à 40°C pendant 24 h puis lyophilisé. Le résidu est repris dans 16 mL de pyridine et la solution est refroidie à 0°C, puis sont ajoutés 8 mL d'anhydride acétique ainsi qu'une quantité catalytique de DMAP (1 mg). Le mélange réactionnel est agité à l'abri de la lumière à 0°C pendant 10 min puis 48 h à température ambiante. L'excès d'anhydride acétique est détruit par addition de méthanol à froid. Le mélange réactionnel est alors concentré et co-évaporé plusieurs fois au toluène. Après purification par chromatographie rapide sur gel de silice (toluène/acétone, 5:1 → 3:1 → 1:1), le composé **33** [130] est obtenu avec 61% de rendement (283.7 mg, 0.22 mmol).

SM (MALDI) : $m/z = 1275 [M+Na]^+$.

β-D-Galactopyranosyl-(1→4)-β-D-glucopyranosyl-(1→4)-(2-acétamido-2-désoxy-β-D-glucopyranosyl)-(1→4)-2-acétamido-2-désoxy-α,β-D-glucopyranose 34



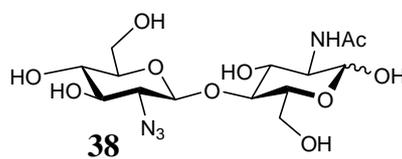
Le composé **33** (283 mg, 0.226 mmol, 1 éq.) en solution à 0°C dans 10 mL de méthanol est désacétylé par ajout d'une solution de méthylate de sodium 1 M (113 µL, 0.113 mmol, 0.5 éq.). Le mélange réactionnel est agité 30 min à 0°C puis 5h30 à température ambiante. Il est ensuite neutralisé avec de la résine Amberlite IR-120 (H⁺), filtré sur coton et concentré. Le résidu est repris dans l'eau puis lyophilisé. Après purification par chromatographie rapide sur gel de silice (acétonitrile/eau, 8:2), le composé libre **34** est obtenu avec 53% de rendement (90.5 mg, 0.120 mmol).

SM (FAB) : m/z = 749 [M+H]⁺.

RMN ¹H (D₂O, 400 MHz) : δ (ppm) = 5.13 (d, $J_{1,2} = 2.6$ Hz, H-1^Iα); 4.63 (d, $J_{1,2} = 8.1$ Hz, H-1^Iβ); 4.54 (2d, $J_{1,2} = 8.1$ Hz, 1H, H-1^{II}α, H-1^{II}β); 4.49 (d, $J_{1,2} = 7.9$ Hz, 1H, H-1^{III}); 4.38 (d, $J_{1,2} = 7.8$ Hz, 1H, H-1^{IV}); 3.94-3.57 (m, 22H, H-2^{I,II}, H-3^{I,II,III,IV}, H-4^{I,II,III,IV}, H-5^{I,II,III,IV}, H-6a^{I,II,III,IV}, H-6b^{I,II,III,IV}); 3.48 (dd, $J_{2,3} = 9.9$ Hz, 1H, H-2^{IV}); 3.30 (m, 1H, H-2^{III}); 2.00 (s, 3H, COCH₃); 1.98 (s, 3H, COCH₃).

RMN ¹³C (D₂O, 400 MHz) : δ (ppm) = 176.0-175.7 (COCH₃); 104.2, 103.6 (C-1^{III}, C-1^{IV}); 96.1 (C-1^Iβ); 91.7 (C-1^Iα); 81.0, 80.5, 79.6, 79.3, 76.6, 76.1, 76.0, 75.8, 75.4, 74.1, 73.8, 73.2, 72.2, 71.2, 70.5, 69.8 (C-2^{III,IV}, C-3^{I,II,III,IV}, C-4^{I,II,III,IV}, C-5^{I,II,III,IV}); 62.3-61.1 (C-6^{I,II,III,IV}); 57.3-54.9 (C-2^{I,II}); 23.4-23.1 (COCH₃).

(2-Azido-2-désoxy-β-D-glucopyranosyl)-(1→4)-2-acétamido-2-désoxy-α,β-D-glucopyranose 38



Préparation de l'azoture triflique : une solution d'azoture de sodium (0.95 g, 14.69 mmol, 6 éq.) dans un mélange de 2.4 mL d'eau et de 2.4 mL de dichlorométhane est refroidie à 0°C. L'anhydride triflique (1.21 mL, 7.35 mmol, 3 éq.) est ajouté lentement à la solution. Le mélange réactionnel est alors agité pendant 2 h à 0°C et sous argon, puis une solution saturée

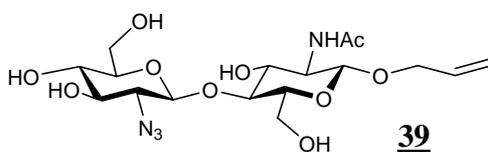
d'hydrogencarbonate de sodium est ajoutée lentement. La phase organique est séparée et la phase aqueuse est extraite avec 2 fois 1.8 mL de dichlorométhane. Le volume total de dichlorométhane utilisé dans la réaction et les traitements doit être approximativement 2.5 fois le volume d'eau utilisé dans la réaction. Les phases organiques sont combinées et lavées avec une solution saturée d'hydrogencarbonate de sodium. La solution de TfN₃ ainsi obtenue, d'une concentration d'environ 0.6 M calculée sur la base d'une conversion de 50%, est utilisée sans autre purification.

Le *N*^I-acétylchitobiose (0.94 g, 2.45 mmol, 1 éq.) et le chlorure de zinc (3.3 mg, 24 μmol, 0.01 éq.) sont solubilisés dans 6 mL d'eau (même volume que le volume de solution de TfN₃ utilisé). A cette solution sont ajoutés de la triéthylamine (1.03 mL, 7.35 mmol, 3 éq.) puis au goutte à goutte 20.4 mL de méthanol, le rapport final eau/méthanol/dichlorométhane étant de 3:10:3. La solution de TfN₃ (6.1 mL) est alors ajoutée et le mélange réactionnel est agité pendant 3.5 h à température ambiante. De l'hydrogencarbonate de sodium (617 mg, 7.35 mmol, 3 éq.) est ajouté au mélange réactionnel qui est ensuite concentré. Après purification par chromatographie rapide sur gel de silice (acétonitrile/eau, 95:5 → 90:10), le composé **38** est obtenu avec 71% de rendement (711 mg, 1.74 mmol).

SM (ES) : $m/z = 431 [M+Na]^+$.

RMN ¹H (D₂O, 400 MHz) : δ (ppm) = 5.15 (s, H-1^Iα); 4.66 (d, $J_{1,2} = 7.3$ Hz, H-1^Iβ); 4.53 (d, $J_{1,2} = 7.9$ Hz, 1H, H-1^{II}); 3.97-3.30 (m, 12 H, H-2^{I,II}, H-3^{I,II}, H-4^{I,II}, H-5^{I,II}, H-6a^{I,II}, H-6b^{I,II}); 1.99 (s, 3H, COCH₃).

RMN ¹³C (D₂O, 400 MHz) : δ (ppm) = 176.0-175.7 (COCH₃); 102.3 (C-1^{II}); 96.1 (C-1^Iβ); 91.8 (C-1^Iα); 79.9 (C-4^Iα); 79.4 (C-4^Iβ); 77.3 (C-5^{II}); 76.1 (C-5^Iβ); 75.5 (C-3^{II}); 73.5 (C-3^Iβ); 71.6 (C-5^Iα); 70.6-70.3 (C-3^Iα; C-4^{II}); 67.3 (C-2^{II}); 61.7-61.2 (C-6^Iα/β, C-6^{II}); 57.7 (C-2^Iβ); 55.1 (C-2^Iα); 23.4-23.1 (COCH₃).

(2-Azido-2-désoxy-β-D-glucopyranosyl)-(1→4)-2-acétamido-2-désoxy-β-D-glucopyranoside d'allyle 39

A une solution du composé 38 (250.4 mg, 0.61 mmol, 1 éq.) dans 1.25 mL de *N,N*-diméthylformamide sont ajoutés du chlorure de lithium (77.8 mg, 1.83 mmol, 3 éq.) puis de l'hydrure de sodium (19.1 mg, 0.796 mmol, 1.3 éq.). Le mélange est agité 15 min à température ambiante et sous argon, puis du bromure d'allyle (312 μ L, 3.7 mmol, 6 éq.) est ajouté (DMF/AllBr, 4:1, v/v). Le mélange réactionnel est alors agité vigoureusement pendant 24 h à température ambiante et sous argon, puis est refroidi à 0°C. Du méthanol est alors ajouté, et le mélange agité à nouveau pendant 15 min, avant d'être concentré. Après purification par chromatographie rapide sur gel de silice (acétonitrile/eau, 95:5), le composé 39 est obtenu avec 50% de rendement (137 mg, 0.30 mmol).

SM (ES) : $m/z = 471 [M+Na]^+$.

RMN 1H (D_2O , 400 MHz) : δ (ppm) = 5.86 (m, 1H, $CH=CH_2$); 5.23 (m, 1H, $CH=CH_2$); 4.54-4.51 (2 d, $J_{1,2} = 8.1$ Hz, $J_{1,2} = 8.0$ Hz, 2H, $H-1^I$, $H-1^{II}$); 4.28 (dd, $J = 5.2$ Hz, $J = 13.2$ Hz, 1H, OCH_2); 4.10 (dd, $J = 6.3$ Hz, $J = 13.2$ Hz, 1H, OCH_2); 3.97-3.30 (m, 12 H, $H-2^{I,II}$, $H-3^{I,II}$, $H-4^{I,II}$, $H-5^{I,II}$, $H-6a^{I,II}$, $H-6b^{I,II}$); 1.98 (s, 3H, $COCH_3$).

RMN ^{13}C (D_2O , 400 MHz) : δ (ppm) = 175.9 ($COCH_3$); 134.6 ($CH=CH_2$); 119.5 ($CH=CH_2$); 102.3 ($C-1^{II}$); 101.2 ($C-1^I$); 79.5 ($C-4^I$); 77.3 ($C-5^{II}$); 76.1 ($C-5^I$); 75.5 ($C-3^{II}$); 73.5 ($C-3^I$); 71.8 (OCH_2); 70.6 ($C-4^{II}$); 67.3 ($C-2^{II}$); 61.7-61.3 ($C-6^I$, $C-6^{II}$); 56.5 ($C-2^I$); 23.4 ($COCH_3$).

V.2.2.9 Couplage enzymatique entre le fluorure d' α -lactosyle 16 et le disaccharide 39

Couplage avec Cel7B E197A d'*Humicola insolens* : à une solution de fluorure de lactosyle 16 (46 mg, 0.134 mmol, 1.2 éq.) dans 9.55 mL de tampon phosphate de sodium 50

mM pH 7 sont ajoutés le *N*^l-acétyl-2^{II}-azidochitobioside d'allyle **39** (50 mg, 0.111 mmol, 1 équ.) et une solution de l'enzyme Cel7B E197A d'*Humicola insolens* (450 µL, 5 mg) dans le tampon Tris-H₃PO₄ 200 mM pH 7.5. Le mélange réactionnel est maintenu à 40°C pendant 22 h, puis concentré. Après purification par chromatographie rapide sur gel de silice (acétonitrile puis acétonitrile/eau, 9:1 → 8:2), un mélange des composés **40** et **41** contaminé par du lactose est obtenu et acétylé. Le mélange est repris dans 5 mL de pyridine. A cette solution sont alors ajoutés 2.5 mL d'anhydride acétique ainsi qu'une quantité catalytique de DMAP (1 mg). Le mélange réactionnel est agité à l'abri de la lumière et à température ambiante pendant 48 h. L'excès d'anhydride acétique est détruit par addition de méthanol à froid. Le mélange réactionnel est alors concentré et co-éaporé plusieurs fois au toluène. Un mélange des composés **42** et **43** est obtenu après purification par chromatographie rapide sur gel de silice (dichlorométhane/acétone, 10:1 → 5:1 → 2:1).

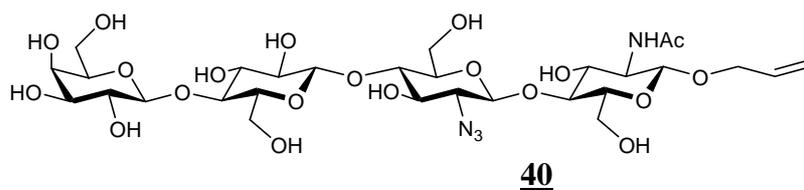
Le mélange **42** et **43** (68.3 mg, 55.3 µmol, 1 équ.) en solution à 0°C dans 3 mL de méthanol est désacétylé par ajout d'une solution de méthylate de sodium 1 M (30 µL, 30 µmol, 0.5 équ.). Le mélange réactionnel est agité 2 h à 0°C puis 10 h à température ambiante. Il est ensuite neutralisé avec de la résine Amberlite IR-120 (H⁺), filtré sur coton et concentré. Le résidu est repris dans l'eau puis lyophilisé. Les composés **40** (6.9 mg) et **41** (17 mg) sont isolés après purification par chromatographie sur colonne ouverte de gel de silice (acétonitrile puis acétonitrile/eau, 97.5:2.5 → 95:5 → 92.5:7.5 → 90:10 → 87.5:12.5).

Couplage avec Cel7B E197A H209A d'*Humicola insolens* : à une solution de fluorure de lactosyle **16** (46 mg, 0.134 mmol, 1.2 équ.) dans 3 mL de tampon phosphate de sodium 50 mM pH 7 sont ajoutés le *N*^l-acétyl-2^{II}-azidochitobioside d'allyle **39** (50 mg, 0.111 mmol, 1 équ.) et une solution de l'enzyme Cel7B E197A H209A d'*Humicola insolens* (7 mL, 5 mg) dans le tampon Tris-Acétate 50 mM pH 8. Le mélange réactionnel est maintenu à 40°C pendant 13 h, puis concentré. Après purification par chromatographie rapide sur gel de silice (acétonitrile puis acétonitrile/eau, 9:1 → 8:2), un mélange des composés **40** et **41** contaminé par du lactose est obtenu et acétylé. Le mélange est repris dans 5 mL de pyridine. A cette solution sont alors ajoutés 2.5 mL d'anhydride acétique ainsi qu'une quantité catalytique de DMAP (1 mg). Le mélange réactionnel est agité à l'abri de la lumière et à température ambiante pendant 48 h. L'excès d'anhydride acétique est détruit par addition de méthanol à froid. Le mélange réactionnel est alors concentré et co-éaporé plusieurs fois au toluène. Un

mélange des composés **42** et **43** est obtenu après purification par chromatographie rapide sur gel de silice (dichlorométhane/acétone, 10:1 → 5:1 → 2:1).

Le mélange **42** et **43** (71.9 mg, 58.2 μmol, 1 éq.) en solution à 0°C dans 3 mL de méthanol est désacétylé par ajout d'une solution de méthylate de sodium 1 M (30 μL, 30 μmol, 0.5 éq.). Le mélange réactionnel est agité 2 h à 0°C puis 10 h à température ambiante. Il est ensuite neutralisé avec de la résine Amberlite IR-120 (H⁺), filtré sur coton et concentré. Le résidu est repris dans l'eau puis lyophilisé. Les composés **40** (23.1 mg) et **41** (6.7 mg) sont isolés après purification par chromatographie sur colonne ouverte de gel de silice (acétonitrile puis acétonitrile/eau, 97.5:2.5 → 95:5 → 92.5:7.5 → 90:10 → 87.5:12.5).

β-D-Galactopyranosyl-(1→4)-β-D-glucopyranosyl-(1→4)-(2-azido-2-désoxy-β-D-glucopyranosyl)-(1→4)-2-acétamido-2-désoxy-β-D-glucopyranoside d'allyle **40**

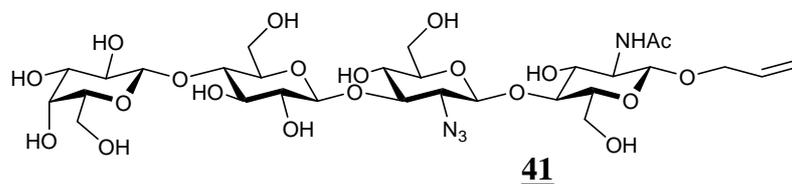


SM (ES) : $m/z = 795 [M+Na]^+$.

RMN ¹H (D₂O, 300 MHz) : δ (ppm) = 5.81 (m, 1H, CH=CH₂); 5.18 (m, 2H, CH=CH₂); 4.51, 4.48 (2d, $J_{1,2} = 8.2$ Hz, $J_{1,2} = 7.9$ Hz, 2H, H-1^{I,II}); 4.43 (d, $J_{1,2} = 7.8$ Hz, 1H, H-1^{III}); 4.34 (d, $J_{1,2} = 7.9$ Hz, 1H, H-1^{IV}); 4.24 (m, 1H, OCH₂); 3.99 (m, 1H, OCH₂); 3.92-3.46 (m, 22H, H-2^{I,II}, H-3^{I,II,III,IV}, H-4^{I,II,III,IV}, H-5^{I,II,III,IV}, H-6a^{I,II,III,IV}, H-6b^{I,II,III,IV}); 3.28 (m, 2H, H-2^{III,IV}); 1.93 (s, 3H, COCH₃).

RMN ¹³C (D₂O, 300 MHz) : δ (ppm) = 175.5 (COCH₃); 134.5 (CH=CH₂); 119.5 (CH=CH₂); 104.2, 103.6, 102.1, 101.2 (C-1^{I,II,III,IV}); 79.4, 79.3, 79.2 (C-4^{I,II,III}); 76.6, 76.2, 76.1, 76.0, 75.4, 74.1, 74.0, 73.8, 73.5, 72.2, 71.8 (C-2^{III,IV}, C-3^{I,II,III,IV}, C-5^{I,II,III,IV}, OCH₂); 69.8 (C-4^{IV}); 66.9 (C-2^{II}); 62.3-61.0 (C-6^{I,II,III,IV}); 56.5 (C-2^I); 23.4 (COCH₃).

β -D-Galactopyranosyl-(1 \rightarrow 4)- β -D-glucopyranosyl-(1 \rightarrow 3)-(2-azido-2-désoxy- β -D-glucopyranosyl)-(1 \rightarrow 4)-2-acétamido-2-désoxy- β -D-glucopyranoside d'allyle 41

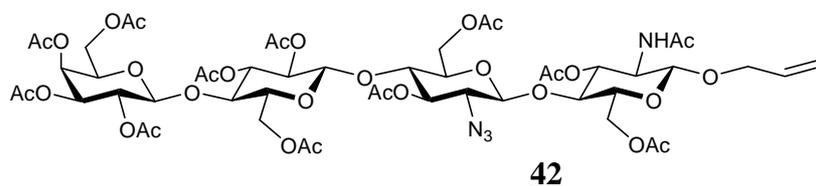


SM (ES) : $m/z = 795 [M+Na]^+$.

RMN 1H (D₂O, 400 MHz) : δ (ppm) = 5.79 (m, 1H, CH=CH₂); 5.17 (m, 2H, CH=CH₂); 4.64 (d, $J_{1,2} = 7.9$ Hz, 1H, H-1^{III}); 4.50 (2d, $J_{1,2} = 8.1$ Hz, $J_{1,2} = 8.0$ Hz, 2H, H-1^{I,II}); 4.35 (d, $J_{1,2} = 7.8$ Hz, 1H, H-1^{IV}); 4.24 (m, 1H, OCH₂); 3.99 (m, 1H, OCH₂); 3.92-3.28 (m, 24H, H-2^{I,II,III,IV}, H-3^{I,II,III,IV}, H-4^{I,II,III,IV}, H-5^{I,II,III,IV}, H-6a^{I,II,III,IV}, H-6b^{I,II,III,IV}); 1.93 (s, 3H, COCH₃).

RMN ^{13}C (D₂O, 400 MHz) : δ (ppm) = 175.9 (COCH₃); 134.5 (CH=CH₂); 119.5 (CH=CH₂); 104.2, 103.7, 102.5, 101.2 (C-1^{I,II,III,IV}); 83.1 (C-3^{II}); 79.5 (C-4^{I,III}); 77.0, 76.6, 76.1, 76.0, 75.4, 74.1, 74.0, 73.5, 72.2, 71.7 (C-2^{III,IV}, C-3^{I,III,IV}, C-5^{I,II,III,IV}, OCH₂); 69.8, 69.0 (C-4^{II,IV}); 67.1 (C-2^{II}); 62.3-61.2 (C-6^{I,II,III,IV}); 56.5 (C-2^I); 23.4 (COCH₃).

(2,3,4,6-Tétra-*O*-acétyl- β -D-galactopyranosyl)-(1 \rightarrow 4)-(2,3,6-tri-*O*-acétyl- β -D-glucopyranosyl)-(1 \rightarrow 4)-(3,6-di-*O*-acétyl-2-azido-2-désoxy- β -D-glucopyranosyl)-(1 \rightarrow 4)-3,6-di-*O*-acétyl-2-acétamido-2-désoxy- β -D-glucopyranoside d'allyle 42



Le composé 40 (19.9 mg, 25.7 μ mol) est repris dans 2 mL de pyridine. A cette solution sont alors ajoutés 2 mL d'anhydride acétique ainsi qu'une quantité catalytique de DMAP (1 mg). Le mélange réactionnel est agité à l'abri de la lumière et à température ambiante pendant 24 h. L'excès d'anhydride acétique est détruit par addition de méthanol à

froid. Le mélange réactionnel est alors concentré et co-évaporé plusieurs fois au toluène. Après purification par chromatographie rapide sur gel de silice (toluène/acétone, 4:1 → 3:1 → 2:1 → 1:1), le composé **42** est obtenu avec 75% de rendement (24 mg, 19.4 μmol).

SM (FAB) : $m/z = 1257 [M+Na]^+$.

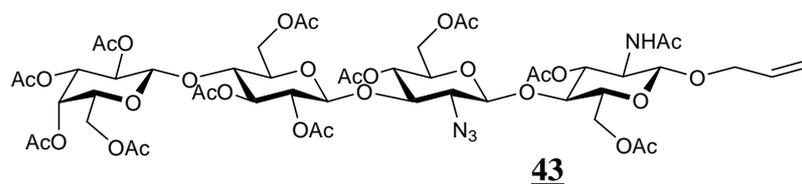
RMN ¹H (CDCl₃, 400 MHz) : δ (ppm) et *J* (Hz)

	H-1	H-2	H-3	H-4	H-5	H-6a	H-6b
Unité I	4.52 d (8.1)	3.93 m	5.14 t (9.1)	3.79 t (9.5)	3.62 m	4.57 dd (2.0; 12.1)	4.18 dd (4.4; 12.2)
Unité II	4.28 d (8.0)	3.33 dd (8.0; 10.2)	4.95 t (9.9)	3.61 t (9.6)	3.54 m	4.36 m	4.08 m
Unité III	4.41 d (8.0)	4.77 dd (7.8; 9.0)	5.09 t (9.0)	3.75 t (9.1)	3.55 m	4.36 m	4.08 m
Unité IV	4.41 d (8.0)	5.05 dd (7.8; 10.4)	4.90 dd (3.4; 10.4)	5.30 d (2.7)	3.82 t (6.5)	4.05 m	4.05 m

5.81 (m, 1H, CH=CH₂); 5.57 (d, *J* = 9.2 Hz, 1H, NH); 5.22 (dq, *J* = 17.2 Hz, *J* = 1.6 Hz, 1H, CH=CH₂); 5.15 (m, 1H, CH=CH₂); 4.26 (m, 1H, OCH₂); 4.01 (m, 1H, OCH₂); 2.32-1.91 (m, 36H, COCH₃).

RMN ¹³C (CDCl₃, 100 MHz) : δ (ppm) = 170.9-169.1 (COCH₃); 133.5 (CH=CH₂); 117.7 (CH=CH₂); 101.0 (C-1^{IV}); 100.8 (C-1^{II}); 100.4 (C-1^{III}); 99.6 (C-1^I); 76.1-75.8 (C-4^{I,II,III}); 72.9-72.7 (C-3^{III} et C-5^{I,II,III}); 71.9-71.8 (C-2^{III} et C-3^{I,II}); 70.9 (C-3^{IV}); 70.7 (C-5^{IV}); 69.7 (OCH₂); 69.0 (C-2^{IV}); 66.5 (C-4^{IV}); 64.3 (C-2^{II}); 62.2-61.9 (C-6^{I,II,III,IV}); 54.1 (C-2^I); 23.2 (NHCOCH₃); 20.8-20.4 (COCH₃).

(2,3,4,6-Tétra-*O*-acétyl- β -D-galactopyranosyl)-(1 \rightarrow 4)-(2,3,6-tri-*O*-acétyl- β -D-glucopyranosyl)-(1 \rightarrow 3)-(4,6-di-*O*-acétyl-2-azido-2-désoxy- β -D-glucopyranosyl)-(1 \rightarrow 4)-3,6-di-*O*-acétyl-2-acétamido-2-désoxy- β -D-glucopyranoside d'allyle **43**



Le composé **41** (17.1 mg, 22.1 μ mol) est repris dans 2 mL de pyridine. A cette solution sont alors ajoutés 2 mL d'anhydride acétique ainsi qu'une quantité catalytique de DMAP (1 mg). Le mélange réactionnel est agité à l'abri de la lumière et à température ambiante pendant 24 h. L'excès d'anhydride acétique est détruit par addition de méthanol à froid. Le mélange réactionnel est alors concentré et co-éaporé plusieurs fois au toluène. Après purification par chromatographie rapide sur gel de silice (toluène/acétone, 4:1 \rightarrow 3:1 \rightarrow 2:1 \rightarrow 1:1), le composé **43** est obtenu avec 50% de rendement (16.9 mg, 13.7 μ mol).

SM (FAB) : $m/z = 1257 [M+Na]^+$.

RMN 1H (CDCl₃, 400 MHz) : δ (ppm) et J (Hz)

	H-1	H-2	H-3	H-4	H-5	H-6a	H-6b
Unité I	4.54 d (8.1)	3.96 m	5.12 t (9.0)	3.76 m	3.57 m	4.36 m	4.10 m
Unité II	4.23 d (8.1)	3.27 m	3.45 t (9.3)	4.83 t (9.2)	3.49 m	4.24 m	3.98 m
Unité III	4.72 d 7.9	4.81 t (8.1)	5.14 t (9.2)	3.77 m	3.66 m	4.54 m	4.19 m
Unité IV	4.41 d (7.9)	5.05 dd (7.8; 10.4)	4.90 dd (3.4; 10.4)	5.30 d (5.3)	3.82 m	4.06 m	4.06 m

5.81 (m, 1H, CH=CH₂); 5.57 (d, $J = 9.1$ Hz, 1H, NH); 5.24 (dq, $J = 17.3$ Hz, $J = 1.4$ Hz, 1H, CH=CH₂); 5.15 (m, 1H, CH=CH₂); 4.28 (m, 1H, OCH₂); 4.02 (m, 1H, OCH₂); 2.30-1.92 (m, 36H, COCH₃).

RMN ^{13}C (CDCl_3 , 100 MHz) : δ (ppm) = 171.0-169.0 (COCH_3); 133.5 ($\text{CH}=\text{CH}_2$); 117.7 ($\text{CH}=\text{CH}_2$); 101.4 (C-1^{II}); 100.9 (C-1^{IV}); 100.4 (C-1^{III}); 99.7 (C-1^{I}); 78.6 (C-3^{II}); 76.0-75.9 ($\text{C-4}^{\text{I,III}}$); 72.9 ($\text{C-3}^{\text{I,III}}$); 72.5 (C-5^{I}); 72.0-71.9 ($\text{C-5}^{\text{II,III}}$); 71.6 (C-2^{III}); 70.9 (C-3^{IV}); 70.7 (C-5^{IV}); 69.8 (OCH_2); 69.0 (C-2^{IV}); 67.5 (C-4^{II}); 66.6 (C-4^{IV}); 66.5 (C-2^{II} et C-4^{IV}); 62.2-60.8 ($\text{C-6}^{\text{I,II,III,IV}}$); 53.9 (C-2^{I}); 23.2 (NHCOCH_3); 20.8-20.4 (COCH_3).

V.3 Biologie moléculaire et biochimie

V.3.1 Matériels

V.3.1.1 Produits chimiques et réactifs

L'isopropylthio- β -D-galactoside (IPTG) et le 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-Gal) proviennent de Euromedex ou Acros. L'IPTG est stocké en solution dans l'eau à -20°C à une concentration de 50 mg/mL. L'X-Gal est stocké en solution dans le DMF à -20°C à une concentration de 25 mg/mL.

Le réactif de Bradford (Bio-Rad) utilisé pour la détermination de la concentration en protéine est composé d'acide phosphorique, de méthanol et de bleu de Coomassie Brilliant Blue G-250.

Le réactif au DNS utilisé pour le dosage des sucres réducteurs a la composition suivante pour 1 L de solution : 10 g d'acide 3,5-dinitrosalicylique, 2 g de phénol, 0.5 g de sulfite de sodium anhydre et 200 g de tartrate double de potassium et de sodium tétrahydraté; le volume de la solution est ajusté à 1 L avec une solution de soude à 2%.

Les différents glycosides de *p*-nitrophényle utilisés proviennent de Sigma.

V.3.1.2 Endoglucanases natives

L'enzyme Cel7B d'*Humicola insolens* nous a été fournie par le Dr Martin Schüleïn (Novozymes, Danemark). Elle a été produite chez *Aspergillus oryzae* et purifiée [137].

L'enzyme Cel7B de *Trichoderma reesei* nous a été fournie par le Dr Anu Koivula (VTT, Finlande).

V.3.1.3 Souches bactériennes, plasmides et milieux de culture

Les cellules d'*E. coli* sont cultivées sur un milieu LB solide (+1.5% d'agar) ou liquide sélectif à 37°C. Un volume de 100 µL d'une solution de l'antibiotique approprié (solution mère 100x) est ajouté pour 100 mL de LB.

Milieu LB : 20 g de peptone, 20 g de NaCl, 10 g d'extrait de levure, 2 L d'eau.

Les souches d'*E. coli* utilisées pour la préparation de plasmides ou pour la surexpression des protéines sont présentées dans le **Tableau 21**.

Souches	Résistance	Provenance
TOP10	-	Invitrogen
JM109	-	Stratagene
DH5α	-	Invitrogen
XL1-Blue	tétracycline	Stratagene
MC1061	streptomycine	Invitrogen
BL21 (DE3)	-	Novagen
BL21 (DE3) pLysS	chloramphénicol	Novagen
Origami (DE3)	kanamycine, tétracycline	Novagen

Tableau 21 : Liste des différentes souches d'*Escherichia coli* utilisées.

Les vecteurs plasmidiques utilisés chez *E. coli* pour les différents clonages et pour l'expression des protéines sont donnés dans le **Tableau 22**.

Vecteurs plasmidiques	Résistance	Provenance
pCR4-TOPO	ampicilline, kanamycine	Invitrogen
pBluescript II KS+	ampicilline	Stratagene
pET-28a(+)	kanamycine	Novagen
pCR4blunt-TOPO	ampicilline, kanamycine	Invitrogen
pET-20b(+)	ampicilline	Novagen
pET-32a(+)	ampicilline	Novagen

Tableau 22 : Liste des différents vecteurs plasmidiques utilisés.

Le vecteur d'expression utilisé chez *B. subtilis* est le vecteur navette pDP66K. Le vecteur d'expression utilisé chez *S. cerevisiae* est le vecteur navette pMA91.

V.3.1.4 Enzymes

Les enzymes de restriction, les enzymes de modification telles que les ADN polymérases ou la T4 ADN ligase proviennent de Fermentas, Gibco, Stratagene, Qiagen, Promega, Invitrogen ou Q-Biogen.

V.3.2 Travaux de biologie moléculaire

V.3.2.1 Méthodes générales

➤ *Transformation de cellules compétentes d'E. coli*

La préparation des cellules compétentes et la transformation d'*Escherichia coli* ont été réalisées selon les protocoles décrits dans la littérature [242].

➤ *Extraction des plasmides*

Les extractions plasmidiques sont effectuées grâce au kit miniprep QIAprep Spin Miniprep (Qiagen).

➤ *Analyse et traitement de l'ADN*

L'ensemble des réactions de digestion, de ligature a été effectué selon les indications des différents fournisseurs.

La séparation, l'identification et la quantification des fragments d'ADN sont réalisées par électrophorèse sur gel d'agarose (1%, p/v) dans du tampon Tris-Acétate-EDTA (TAE) 1x. Le bromure d'éthidium (BET) est inclus directement dans le gel d'agarose et permet la visualisation de l'ADN aux rayons ultraviolets. Le marqueur de taille déposé à chaque électrophorèse est l'ADN du phage λ digéré par l'enzyme *PstI*.

La purification des fragments d'ADN à partir des gels d'agarose est faite à l'aide du kit QIAEX II Gel Extraction (Qiagen). Les fragments obtenus par PCR sont purifiés grâce au kit MinElute PCR Purification (Qiagen).

➤ *Amplification de l'ADN et mutagénèse dirigée*

Les réactions de PCR ont été effectuées avec la Proof-Start ADN polymérase ou la *PfuTurbo* ADN polymérase avec un thermocycleur de type Biometra T Personal. Les amorces ont été synthétisées par la société MWG Biotech.

Une réaction typique de PCR est réalisée dans un volume de 25 μ L. La matrice d'ADN (de l'ordre de 10 ng) est incubée pendant 30 cycles de PCR avec les amorces (de l'ordre de 200 ng de chaque amorce), les désoxyribonucléosides triphosphates (dATP, dCTP, dGTP et dTTP à une concentration finale de 400 μ M pour chaque dNTP), de la Proof-Start ADN polymérase (1.25 U) et du tampon de réaction de l'enzyme (solution mère 10x). Chaque cycle comprend une étape de dénaturation de l'ADN (1 min, 95°C), une étape d'hybridation des amorces sur la matrice (1 min à une température variable selon les amorces utilisées) et une étape d'élongation (à 72°C pendant une durée variable selon la taille du fragment à amplifier).

La mutagénèse dirigée pour la construction des mutants est réalisée par PCR dans des conditions similaires basées sur le protocole du kit QuikChange Site-Directed Mutagenesis de Stratagene. La polymérase utilisée est la *PfuTurbo* ADN polymérase (2.5 U) dans un volume de réaction de 50 μ L. Le nombre de cycles (12 ou 15) dépend du type de mutation désirée. L'étape d'élongation a lieu à 68°C pendant une durée variable (2 min/kb). Les brins d'ADN parents sont finalement digérés par l'enzyme *DpnI* avant la transformation de cellules compétentes d'*E. coli*.

➤ *Analyse des séquences d'ADN*

Le séquençage des fragments d'ADN a été réalisé par la société GENOME express.

V.3.2.2 Constructions des plasmides d'expression

Les plasmides pCR4-TOPO contenant les gènes codant pour Cel7B (noté pHW 704) et pour Cel7B E197A (noté pCE 10) sont fournis par le Dr Jesper Vind (Novozymes, Danemark).

Dans le cas de Cel7B, un site de restriction *NotI* est inséré par PCR en 3' du gène codant pour Cel7B. Les amorces utilisées sont données dans le **Tableau 23**. Le fragment PCR amplifié *XhoI/NotI* (390 pb) est sous-cloné dans le plasmide pBlueScript KS(+) pour vérifier les mutations par séquençage.

Mutation	Séquences des amorces
insertion du site <i>NotI</i> en 3'	5'-GGCAAGCTCGAGAAGATCC-3' (sens) 5'-CTTGGGCG CGGCCG CCTGAACCTCC-3' (antisens)

Tableau 23 : Séquences des amorces pour l'insertion du site de restriction *NotI* en 3' de Cel7B d'*H. insolens*.

En gras, les bases mutées; les bases soulignées sont caractéristiques du site de restriction *XhoI* (cas de l'amorce sens) et du site de restriction *NotI* (cas de l'amorce antisens).

Les fragments *BamHI/XhoI* (906 pb) et *XhoI/NotI* après digestion (370 pb) du gène codant pour Cel7B sont sous-clonés dans pBlueScript KS(+). Finalement, le fragment *NcoI/NotI* (1269 pb) du gène codant pour Cel7B est cloné dans le vecteur d'expression pET-20b(+), conduisant au plasmide noté **pSB-1**, ainsi que dans le vecteur pET-32a(+), conduisant au plasmide noté **pSB-2**.

Dans le cas de Cel7B E197A, un site de restriction *NotI* est inséré en 3' du gène codant pour Cel7B E197A, de la même manière que pour Cel7B. Les fragments *BamHI/XhoI* (906 pb) et *XhoI/NotI* (370 pb) du gène codant pour Cel7B E197A sont ensuite sous-clonés dans pBS KS(+). A ce niveau, la même stratégie que pour Cel7B ne peut pas être suivie puisque le site de restriction *NcoI* n'est pas unique dans le gène. Dans une étape intermédiaire, le fragment *BamHI/NotI* (1276 pb) du gène codant pour Cel7B E197A est cloné dans le vecteur pET-28a(+), puis le fragment *PstI/SacI* (690 pb) contenant le site *NcoI* à muter, est sous-cloné dans pBS KS(+), qui sera la matrice pour l'étape suivante de mutagenèse dirigée par PCR. Les amorces utilisées sont données dans le **Tableau 24**.

Mutation	Séquences des amorces
mutation du site <i>NcoI</i>	5'-GGCTCGTGCTGCAACGCGATGGATATCTGGGAGG-3' (sens) 5'-CCTCCCAGATATCCATCGCGTTGCAGCACGAGCC-3' (antisens)

Tableau 24 : Séquences des amorces pour la mutation du site de restriction *NcoI* au niveau de l'alanine 197 de *Cel7B E197A d'H. insolens*.

En gras, la base mutée.

Le fragment *PstI/SacI* (690 pb) est sous-cloné dans pET-28a(+) contenant les fragments complémentaires du gène de *Cel7B E197A*. Finalement, le fragment *NcoI/NotI* (1269 pb) du gène codant pour *Cel7B E197A* est cloné dans le vecteur d'expression pET-20b(+), conduisant au plasmide noté **pSB-3**.

La délétion du peptide signal est effectuée par PCR en utilisant comme matrice le plasmide pHW 704 contenant le gène codant pour *Cel7B*. Les amorces utilisées sont données dans le **Tableau 25**. L'insertion d'un nouveau site de restriction *NcoI* au niveau de l'ATG est réalisée en même temps que la délétion du peptide signal.

Mutation	Séquences des amorces
délétion du peptide signal	5'-GGCTGGTCCATGGC █ CAGAAG-3' (sens) 5'-GGATCTTCTCGAGCTTGCC-3' (antisens)

Tableau 25 : Séquences des amorces utilisées pour la délétion du peptide signal de *Cel7B d'H. insolens*.

En italique, le codon correspondant au premier acide aminé de la protéine mature; en gras, les bases mutées; en blanc sur fond noir, la base insérée; les bases soulignées sont caractéristiques du site de restriction *NcoI* (cas de l'amorce sens) et du site de restriction *XhoI* (cas de l'amorce antisens).

Le fragment PCR *NcoI/XhoI* (843 pb) du gène codant pour *Cel7B* est sous-cloné dans le plasmide pCRB, un plasmide dérivé de pCR4blunt-TOPO (Eric Samain, travaux non publiés), pour vérifier les mutations par séquençage. Les fragments *NcoI/XhoI* (843 pb) et *XhoI/NotI* (370 pb) du gène codant pour *Cel7B* mature sont sous-clonés dans pET-20b(+), conduisant au plasmide noté **pSB-4**.

Pour l'expression de Cel7B mature dans *Bacillus subtilis*, les sites de restriction *NcoI* et *NotI* ont dans un premier temps été introduits respectivement en amont et en aval du gène codant pour la CGTase de *B. circulans* dans le vecteur d'expression pDP66K [230]. L'insertion de ces deux sites a été réalisée par PCR en utilisant les amorces données dans le **Tableau 26**.

Mutation	Séquences des amorces
insertion du site <i>NcoI</i>	5'-GATACCGAGGTATCCGCCATGGCCTGGGCAGGGCTG-3' (sens) 5'-CAGCCCTGCCCAGGCCATGGCGGATACCTCGGTATC-3' (antisens)
insertion du site <i>NotI</i>	5'-CGATGAATGTAGCCGCCTATCGCGGCCGCCTTATGGCTGCCAATT-3' (sens) 5'-GAATTGGCAGCCATAAGGCGGCCGCGATAGGCGGCTACATTCATC-3' (antisens)

Tableau 26 : Séquences des amorces utilisées pour l'insertion des sites de restriction *NcoI* et *NotI* dans pDP66K.

Le fragment *NcoI/NotI* (1213 pb) codant pour Cel7B mature obtenu par digestion de **pSB-4** a ensuite été cloné dans le vecteur pDP66K ainsi modifié, conduisant au plasmide noté **pSB-5**.

V.3.2.3 Mutagenèse dirigée

V.3.2.3.1 Cellulases d'*Humicola insolens*

Pour réaliser la mutation du résidu 209, le fragment *EcoRV/SacI* (380 pb) obtenu par digestion du plasmide pHW 704 est sous-cloné dans pBS KS(+), qui sera alors utilisé comme matrice pour la mutagenèse dirigée par PCR. Les amorces utilisées dans le cas des mutations H209A et H209G sont présentées dans le **Tableau 27**. L'insertion d'un site de restriction *NruI* permet le criblage des clones obtenus.

Mutation	Séquences des amorces
H209A	5'-GCCAACT <u>TCGCGA</u> GCCTCG GG CGTGGCTCCCCAC-3' (sens)
	5'-GTGGGGAGCCAC GG CCGAGGCT <u>TCGCGA</u> GTTGGC-3' (antisens)
H209G	5'-GCCAACT <u>TCGCGA</u> GCCTCG GG CGTGGCTCCCCAC-3' (sens)
	5'-GTGGGGAGCCAC GG CCGAGGCT <u>TCGCGA</u> GTTGGC-3' (antisens)
A211T	5'-GGGAGG CGAATT CGCGAGCCTCGGCGTGACTCCCCACACC-3' (sens)
	5'-GGTGTGGGGAG TCACGG CCGAGGCTCGCG GAATT CGCCTCCC-3' (antisens)

Tableau 27 : Séquences des amorces utilisées pour la mutagenèse dirigée sur Cel7B d'*H. insolens*.

En italique, le codon correspondant au résidu muté; en gras, les bases mutées; les bases soulignées sont caractéristiques du site de restriction *NruI* (cas des mutations H209A et H209G) et du site de restriction *EcoRI* (cas de la mutation A211T) insérés lors de la mutagenèse.

Les fragments mutés *EcoRV/SacI* (380 pb) du gène codant pour Cel7B sont ensuite sous-clonés dans **pSB-3**.

Concernant la mutation du résidu 211, le plasmide pBS KS(+) contenant le fragment *EcoRV/SacI* (380 pb) de Cel7B avec la mutation H209A est utilisé comme matrice pour la mutagenèse dirigée par PCR. Les amorces utilisées dans le cas de la mutation A211T sont présentées dans le **Tableau 27**. L'insertion d'un site de restriction *EcoRI* permet le criblage des clones obtenus.

Le fragment muté *EcoRV/SacI* (380 pb) du gène codant pour Cel7B est ensuite sous-cloné dans **pSB-3**.

V.3.2.3.2 Cellulase de *Trichoderma reesei*

Le plasmide noté pMP311 contenant l'ADNc de Cel7B de *Trichoderma reesei* nous a été donné par le Dr Anu Koivula (VTT, Finlande).

La mutation E196A est réalisée en utilisant le plasmide pMP311 comme matrice pour la mutagenèse dirigée par PCR. Les amorces utilisées sont données dans le **Tableau 28**. L'insertion d'un site de restriction *NcoI* permet le criblage des clones obtenus.

Mutation	Séquences des amorces
E196A	5'-GGGCTTCTGCTGCAACG CC ATGGATATCCTGGAGG-3' (sens) 5'-CCTCCAGGATAT CC AT GG CGTTGCAGCAGAAGCCC-3' (antisens)

Tableau 28 : Séquences des amorces utilisées pour la mutation du nucléophile de Cel7B de *T. reesei*.

En italique, le codon correspondant au résidu muté; en gras, les bases mutées; les bases soulignées sont caractéristiques du site de restriction *NcoI* inséré lors de la mutagenèse.

V.3.3 Expression de Cel7B d'*H. insolens* chez la bactérie

V.3.3.1 Méthodes générales

V.3.3.1.1 Extraction protéique

➤ *Extraits cellulaires*

Ils sont préparés selon le protocole suivant décrit ci-dessous.

40 mL de culture bactérienne sont prélevés et centrifugés à 6000 rpm pendant 10 min à 4°C. Les cellules sont resuspendues dans 5 mL de tampon I puis congelées pour la nuit. Le lendemain, la suspension est lentement décongelée dans la glace puis est soumise aux ultrasons à froid par cycles de 2 min (Bioblock Scientific Vibra Cell 72434). Finalement, les débris cellulaires sont éliminés par centrifugation à 13000 rpm pendant 30 min à 4°C. Le surnageant constitue l'extrait cellulaire qui sera analysé.

Tampon I : tampon Tris-HCl (20 mM, pH 8) contenant 100 mM de NaCl.

➤ *Fractions périplasmiques*

Elles sont préparées par choc osmotique selon le protocole suivant.

40 mL de culture bactérienne sont prélevés et centrifugés à 6000 rpm pendant 10 min à 4°C. Le culot est resuspendu dans 10 mL de tampon II (0.25 fois le volume de culture) puis congelé pour la nuit. Le lendemain, la suspension de cellules est lentement décongelée dans la

glace puis centrifugée à 6000 rpm pendant 5 min à 4°C. Le culot bactérien est resuspendu dans 2 mL de tampon III préalablement refroidi à 4°C. 4 µL d'une solution préalablement refroidie d'EDTA (0.5 M) sont ajoutés à la solution précédente (concentration finale de 1mM); le tout est alors incubé dans la glace pendant 30 min. Le milieu est alors centrifugé à 8000 rpm pendant 10 min à 4°C et les cellules sont resuspendues dans 2 mL d'une solution de MgSO₄ (5 mM) à 4°C. La suspension est incubée dans la glace pendant 30 min. Après centrifugation à 8000 rpm pendant 10 min à 4°C, le surnageant constituant la fraction périplasmique est prélevé.

Tampon II : tampon Tris-HCl (20 mM, pH 8).

Tampon III : tampon Tris-HCl (30 mM, pH 8) contenant 20% de saccharose

V.3.3.1.2 Dosage des protéines

La concentration protéique a été déterminée par la méthode de Bradford [243], en utilisant la sérum albumine bovine (SAB) comme référence.

V.3.3.1.3 Electrophorèse en conditions dénaturantes sur gel de polyacrylamide (SDS-PAGE)

L'analyse qualitative des protéines a été réalisée par électrophorèse sur gel de polyacrylamide à 10% après dénaturation de l'échantillon [244].

V.3.3.1.4 Dosage des sucres réducteurs par la méthode au DNS

Les sucres réducteurs peuvent être dosés grâce à leurs propriétés réductrices, en milieu alcalin et à chaud, vis-à-vis de l'acide 3,5-dinitrosalicylique. En effet, ce dernier, de couleur jaune, peut être réduit en acide 3-amino-5-nitrosalicylique, de couleur orange-rouge, que l'on peut doser par colorimétrie à 550 nm.

Le protocole suivi pour le dosage des sucres réducteurs est le suivant.

100 ou 200 μL d'extrait cellulaire ou de fraction périplasmique sont dilués par le tampon acétate (25 mM, pH 5) contenant 1% de CMC jusqu'à un volume total de 400 μL . Le mélange réactionnel est incubé pendant une durée variable à 37°C. 600 μL de réactif au DNS (cf. paragraphe V.3.1.1 pour la composition) sont alors ajoutés. Le milieu est porté à ébullition pendant 10 min puis refroidi rapidement par un bain de glace. L'absorbance est finalement lue à 550 nm.

Des contrôles sans CMC ou sans milieu enzymatique sont effectués dans les mêmes conditions.

V.3.3.2 *Etude de l'expression chez E. coli*

L'étude d'expression de Cel7B d'*H. insolens* chez *Escherichia coli* est généralement effectuée de la manière suivante, pour chacune des souches testées.

250 mL de LB sélectif sont inoculés avec 2 mL de préculture et agités à 200 rpm à 37°C jusqu'à atteindre une absorbance à 550 nm de 0.5-0.6. Un premier prélèvement de 40 mL est effectué avant l'induction par l'IPTG à différentes concentrations. L'induction est réalisée à différentes températures (25 ou 37°C). Des prélèvements de 40 mL sont effectués à différents temps après induction. Un contrôle non induit est réalisé dans les mêmes conditions.

Des extraits cellulaires ou des fractions périplasmiques préparés à partir des différents prélèvements sont analysés par SDS-PAGE et par mesure de l'activité cellulasique.

V.3.3.3 *Etude de l'expression chez B. subtilis*

La préparation des cellules compétentes et la transformation de *Bacillus subtilis* ont été réalisées d'après le protocole décrit par Bron [245].

La souche de *Bacillus subtilis* DB104A transformée avec le plasmide **pSB-5** est cultivée en Erlenmeyer contenant 150 mL de milieu de culture – composé de tryptone (2%), d'extrait de levure (0.5%), de chlorure de sodium (1%), de casamino-acides (1%) et de kanamycine (5 $\mu\text{g}/\text{mL}$) – à pH7 pendant 16 h à 37°C sous agitation (200 rpm).

La production de Cel7B est induite par l'addition d'érythromycine (10 µg/mL). Après 8 h d'induction à 37°C, le milieu de culture est centrifugée à 4°C pendant 30 min à 10000 rpm. Les protéines sécrétées dans le milieu de culture sont alors analysées par SDS-PAGE.

V.3.4 Production des enzymes dans des organismes hétérologues

V.3.4.1 Enzymes d'*H. insolens* chez *Aspergillus oryzae*

En collaboration avec Novozymes (Danemark), les différents mutants de Cel7B E197A d'*H. insolens* préparés par mutagenèse dirigée ont été produits chez le champignon *Aspergillus oryzae*, sous le contrôle du promoteur de la Taka-amylase A [231].

Après fermentation, les protéines extracellulaires sont purifiées par chromatographie échangeuse d'ions [130] et sont utilisées ensuite pour nos tests enzymatiques.

V.3.4.2 Enzymes de *T. reesei* chez *Saccharomyces cerevisiae*

Le plasmide du type pMA91 porteur du gène de Cel7B E196A de *Trichoderma reesei* a été utilisé par l'institut VTT (Finlande) pour exprimer la protéine chez la levure *Saccharomyces cerevisiae*.

Le mutant Cel7B E196A de *T. reesei* a été produit chez la levure sous le contrôle du promoteur du gène *pgk1* de la phosphoglycérate kinase de la levure et sécrété dans le milieu de culture, comme décrit précédemment [232].

Le milieu de culture est concentré, en utilisant les systèmes Nanosep et Microsep (PALL Life Sciences), avec une membrane dont le seuil de coupure est de 10 kDa. La concentration des protéines est réalisée selon les indications données par le fournisseur, jusqu'à l'obtention d'une concentration protéique de 0.52 mg/mL.

Le milieu de culture ainsi concentré est utilisé sans autre purification pour nos tests enzymatiques.

V.3.5 Détermination des activités enzymatiques

V.3.5.1 Détection des activités

La détection des activités des différentes enzymes étudiées est réalisée par chromatographie sur couche mince. L'analyse par spectrométrie de masse permet d'attester ensuite de la formation des produits attendus et visibles sur CCM.

V.3.5.2 Quantification par HPLC

V.3.5.2.1 Matériel

Les séparations par Chromatographie Liquide Haute Performance (HPLC) ont été réalisées avec un système équipé d'une colonne en phase amine NH₂ fournie par Interchim (Interchrom) de type μ Bondapak 10 μ NH₂ (300 \times 3.9).

Les échantillons sont élués avec différents mélanges d'acétonitrile et d'eau dans des conditions isocratiques, délivrés par une pompe Beckmann, de type System Gold 127P Solvent Module, avec un débit de 0.5 mL/min.

La détection est effectuée avec un détecteur à diffusion de lumière DDL 31 de chez Eurosep Instruments, utilisé dans les conditions suivantes :

- sensibilité/gain du photomultiplicateur (HT PM) : 500;
- température de la chambre d'évaporation (T2) : 45°C;
- température du nébuliseur (T1) : 35°C.

Le volume d'échantillon injecté est de 20 μ L. Les analyses sont réalisées à 25°C.

L'acquisition et le traitement des données sont réalisés grâce au logiciel Azur version 2.0 de Datalys.

Les conditions d'éluion des échantillons pour les différentes analyses HPLC sont les suivantes :

- suivi de la réaction d'hydrolyse du tétrasaccharide **34** : MeCN/H₂O, 80:20;
- suivi de la réaction d'hydrolyse du cellotétraose : MeCN/H₂O, 75:25;
- suivi de la réaction de couplage entre le fluorure d' α -lactosyle et le cellobiose : MeCN/H₂O, 80:20;

- suivi de la réaction de couplage entre le fluorure d' α -lactosyle et le N^1 -acétylchitobiose : MeCN/H₂O, 75:25;
- suivi de la réaction de couplage entre le fluorure d' α -lactosyle et le composé **38** : MeCN/H₂O, 80:20;
- suivi de la réaction de couplage entre le fluorure d' α -lactosyle et le composé **39** : MeCN/H₂O, 85:15;

Toutes les solutions ont été filtrées sur filtre de 0.45 μ m.

Les analyses LC-MS ont été effectuées avec la même colonne. Le système HPLC utilisé est composé d'un système de deux pompes (Beckmann, de type System Gold LC-125S Solvent Module) et d'un injecteur avec une boucle de 20 μ L. Le débit de la colonne est de 0.5 mL/min dont une partie est menée à une source d'ionisation de type électrospray. Les spectres de masse (ES⁺) sont effectués sur un appareil Waters Micromass ZQ.

V.3.5.2.2 Etudes HPLC

La quantification de l'activité d'hydrolyse des deux endoglucanases Cel7B d'*Humicola insolens* et de *Trichoderma reesei* vis-à-vis du tétrasaccharide **34** et du cellotétraose a été réalisée par une analyse HPLC des prélèvements du milieu réactionnel effectués à différents temps de réaction.

Les réactions ont lieu dans le tampon phosphate de sodium 50 mM pH 7 pour Cel7B d'*H. insolens* ou dans le tampon acétate de sodium 25 mM pH 4.6 pour Cel7B de *T. reesei*.

Dans le cas de l'hydrolyse du tétrasaccharide **34**, le milieu réactionnel est composé de 160 μ L d'une solution du tétrasaccharide **34** (12.6 mg/mL) à laquelle sont ajoutés 40 μ L d'une solution de l'enzyme Cel 7B d'*H. insolens* (1 mg/mL) ou de *T. reesei* (0.16 mg/mL).

Dans le cas de l'hydrolyse du cellotétraose, le milieu réactionnel est composé de 240 μ L d'une solution de cellotétraose (13.8 mg/mL) à laquelle sont ajoutés 60 μ L d'une solution de l'enzyme Cel 7B d'*H. insolens* (5 μ g/mL) ou de *T. reesei* (8 μ g/mL).

Selon une méthode comparable, la quantification de l'activité glycosynthase des différents mutants de l'enzyme Cel7B E197A d'*H. insolens* a été effectuée par l'analyse HPLC

des prélèvements du milieu de réaction de couplage entre le fluorure d' α -lactosyle **16** et le cellobiose.

Pour chaque enzyme étudiée, la composition du milieu réactionnel est décrite dans le **Tableau 29**. Chaque réaction a lieu dans le tampon phosphate de sodium 50 mM pH 7. Dans chaque cas, les concentrations du donneur et de l'accepteur sont égales à 4.5 mg/mL.

Donneur : lactose α-F 16	Accepteur : cellobiose	Volume de tampon ajouté	Enzyme
1.8 mg	1.8 mg	360 μ L	40 μ L d'une solution de Cel7B E197A (0.56 mg/mL)
2.25 mg	2.25 mg	250 μ L	250 μ L d'une solution de Cel7B E197A H209A (0.71 mg/mL)
2.25 mg	2.25 mg	250 μ L	250 μ L d'une solution de Cel7B E197A H209G (0.61 mg/mL)
2.25 mg	2.25 mg	-	500 μ L d'une solution de Cel7B E197A H209A A211T (0.36 mg/mL)

Tableau 29 : Composition des milieux réactionnels de couplage entre le donneur **16 et le cellobiose.**

Des prélèvements de 25 μ L du milieu réactionnel sont effectués régulièrement et sont rapidement congelés à -196°C dans l'azote liquide pour stopper la réaction enzymatique. Ils sont ensuite décongelés juste avant leur injection en HPLC.

L'activité d'hydrolyse du tétrasaccharide **34** est déterminée en fonction de l'apparition du N^I, N^{II} -diacétylchitobiose au cours du temps. L'activité glycosynthase est quant à elle exprimée en fonction de la disparition du cellobiose.

Dans les deux cas, les gammes étalons appropriées ont été réalisées afin de corréler les aires des différents pics à la quantité de substrat correspondante.

CONCLUSION ET PERSPECTIVES

CONCLUSION ET PERSPECTIVES

Ces travaux effectués au sein de l'équipe "Glycochimie et Biotechnologie des Oligosaccharides" du CERMAV ont concerné la modification de la spécificité d'une enzyme, la glycosynthase Cel7B E197A d'*Humicola insolens*, couramment utilisée dans divers projets pour la synthèse enzymatique d'oligosaccharides d'intérêt. Il s'agissait d'un véritable challenge compte-tenu du peu d'exemples disponibles dans la littérature. Ce projet original et ambitieux visait à utiliser cette glycosynthase pour la préparation de glycoconjugués, notamment pour le remodelage de la glycosylation des protéines recombinantes. L'ingénierie de la glycosynthase nécessaire pour cette application constituait une thématique jusqu'à maintenant inexplorée.

L'étude du comportement de la glycosynthase Cel7B E197A d'*H. insolens* vis-à-vis de substrats non-naturels a permis dans un premier temps de montrer la capacité de la glycosynthase à transférer un oligosaccharide complexe sur un glycopeptide modèle, validant ainsi la stratégie envisagée pour le remodelage de la glycosylation des protéines.

Les études de modélisation moléculaire ont été effectuées pour déterminer de manière rationnelle les mutations ponctuelles nécessaires à la modification de la spécificité de substrat de la glycosynthase. Les résultats obtenus ont laissé penser que des mutations au niveau des résidus 209 et 211 pouvaient permettre la fixation d'un groupement *N*-acétyle dans le sous-site +1 du site actif de l'enzyme et ainsi rendre le greffage d'un oligosaccharide sur une unité *N*-acétylglucosaminyle possible. Des tests réalisés avec l'enzyme native Cel7B de *Trichoderma reesei* ont confirmé la pertinence de ces mutations.

L'ingénierie de la glycosynthase Cel7B E197A d'*H. insolens* et la préparation de la glycosynthase Cel7B E196A de *T. reesei* ont été réalisées par mutagenèse dirigée et les enzymes ont été produites respectivement chez *Aspergillus oryzae* et *Saccharomyces cerevisiae*, des essais d'expression de Cel7B d'*H. insolens* chez des bactéries n'ayant pas permis d'obtenir avec succès une protéine active.

La caractérisation de l'activité des différents mutants vis-à-vis d'un accepteur tel que le cellobiose a tout d'abord été réalisée. L'activité glycosynthase du mutant Cel7B E196A de *T. reesei* a pu ainsi être mise en évidence, ajoutant une nouvelle *endo*-glycosynthase au répertoire des glycosynthases déjà existantes. Les mutations H209A, H209G et A211T réalisées sur la glycosynthase Cel7B E197A d'*H. insolens* ont eu une incidence plus ou moins

importante sur l'activité glycosynthase de celles-ci. Les mutations introduites au niveau du résidu His-209 ont notamment provoqué une chute de l'activité spécifique des mutants glycosynthases.

L'étude de la spécificité de substrat des glycosynthases vis-à-vis d'un accepteur tel que le N^I, N^{II} -diacétylchitobiose n'a pas mis en évidence d'activité glycosynthase avec cet accepteur. Les différentes mutations réalisées ne sont donc pas suffisantes. Les tests effectués avec Cel7B E196A de *T. reesei* ont abouti aux mêmes observations. Cette absence d'activité vis-à-vis de ce type de substrat est en contradiction avec l'activité d'hydrolyse de l'enzyme native vis-à-vis du tétrasaccharide **34**.

La caractérisation de l'activité des différentes glycosynthases de Cel7B E197A d'*H. insolens* vis-à-vis du N^I -acétyl-2^{II}-azidochitobioside d'allyle **39** a permis de mettre en évidence un comportement atypique chez les endo-glycosynthases. En effet, l'introduction d'une substitution de type azido en position 2 de l'extrémité non-réductrice de l'accepteur a permis d'obtenir une nouvelle régiosélectivité pour la glycosynthase Cel7B E197A, formant majoritairement des liaisons du type β 1-3 en présence de cet accepteur. Par ailleurs, cette régiosélectivité est modifiée par des mutations au niveau de l'His-209, favorisant la formation de liaisons β 1-4 ainsi qu'au niveau de l'Ala-211, permettant de former exclusivement une liaison β 1-3. Ces différents comportements laissent penser que les résidus 209 et 211 pourraient être impliqués dans la spécificité de substrat au niveau d'une substitution de la position 2 de l'unité présente dans le sous-site +1 de l'enzyme.

Compte-tenu des résultats obtenus par l'ingénierie rationnelle de l'enzyme, la modification de la spécificité de substrat de la glycosynthase Cel7B E197A d'*H. insolens* pourrait maintenant être envisagée par évolution dirigée. Lin et coll. [246] ont d'ailleurs récemment appliqué leur approche de "Chemical Complementation", une méthode de sélection basée sur l'utilisation d'une levure triple hybride, à l'évolution dirigée de la glycosynthase Cel7B E197A.

D'autre part, le comportement inattendu de la glycosynthase Cel7B E197A pourrait être analysé plus en détail par des études de cristallographie. Il s'agirait de comprendre l'influence des mutations introduites sur le repliement des différentes enzymes pour interpréter les différentes activités observées. Ces études permettraient de comprendre également le positionnement du N^I -acétyl-2^{II}-azidochitobioside d'allyle **39** dans les sous-sites accepteurs de l'enzyme et donc la régiosélectivité de la réaction de couplage.

Enfin, cette nouvelle régiosélectivité de la glycosynthase, formant des liaisons β 1-3 entre des disaccharides portant des liaisons β 1-4, pourrait constituer une alternative à l'utilisation de la glycosynthase dérivée de la β 1,3-1,4-glucanase *Bacillus licheniformis* dans la synthèse de dérivés de β 1,3-1,4-glucanes.

REFERENCES BIBLIOGRAPHIQUES

REFERENCES BIBLIOGRAPHIQUES

- [1] Kobata, A. *Eur. J. Biochem.* **1992**, *209*, 483-501.
- [2] Spiro, R. G. *Glycobiology* **2002**, *12*, 43R-56R.
- [3] Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4-8.
- [4] Johansen, P. G.; Marshall, R. D.; Neuberger, A. *Biochem. J.* **1961**, *78*, 518-527.
- [5] Sears, P.; Wong, C. H. *Cell. Mol. Life Sci.* **1998**, *54*, 223-252.
- [6] Rademacher, T. W.; Parekh, R. B.; Dwek, R. A. *Annu. Rev. Biochem.* **1988**, *57*, 785-838.
- [7] Cunningham, L. W.; Clouse, R. W.; Ford, J. D. *Biochim. Biophys. Acta* **1963**, *78*, 379-381.
- [8] Hubbard, S. C.; Ivatt, R. J. *Annu. Rev. Biochem.* **1981**, *50*, 555-583.
- [9] Abeijon, C.; Hirschberg, C. B. *Trends Biochem. Sci.* **1992**, *17*, 32-36.
- [10] Lehrman, M. A. *Glycobiology* **1991**, *1*, 553-562.
- [11] Pan, Y. T.; Elbein, A. D. *Prog. Drug Res.* **1990**, *34*, 163-207.
- [12] Silberstein, S.; Gilmore, R. *FASEB J.* **1996**, *10*, 849-858.
- [13] Gavel, Y.; Von Heijne, G. *Protein Eng.* **1990**, *3*, 433-442.
- [14] Imperiali, B.; Shannon, K. L. *Biochemistry* **1991**, *30*, 4374-4380.
- [15] Imperiali, B.; Shannon, K. L.; Unno, M.; Rickert, K. W. *J. Am. Chem. Soc.* **1992**, *114*, 7944-7945.
- [16] Bause, E. *Biochem. Soc. Trans.* **1984**, *12*, 514-517.
- [17] Bause, E.; Breuer, W.; Peters, S. *Biochem. J.* **1995**, *312*, 979-985.
- [18] Helenius, A.; Trombetta, E. S.; Nebert, D. N.; Simons, F. *Trends Cell Biol.* **1997**, *7*, 193-200.
- [19] Goochee, C. F.; Monica, T. *Bio/Technology* **1990**, *8*, 421-427.
- [20] Van den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 151-208.

- [21] Varki, A. *Glycobiology* **1993**, 3, 97-130.
- [22] Trombetta, E. S.; Helenius, A. *Curr. Opin. Struct. Biol.* **1998**, 8, 587-592.
- [23] Trombetta, E. S.; Helenius, A. *J. Cell Biol.* **2000**, 148, 1123-1129.
- [24] Gervais, A.; Hammel, Y.-A.; Pelloux, S.; Lepage, P.; Baer, G.; Carte, N.; Sorokine, O.; Strub, J.-M.; Koerner, R.; Leize, E.; Van Dorsselaer, A. *Glycobiology* **2003**, 13, 179-189.
- [25] Dahms, N. M.; Lobel, P.; Kornfeld, S. *J. Biol. Chem.* **1989**, 264, 12115-12118.
- [26] Bruneau, N.; Nganga, A.; Fisher, E. A.; Lombardo, D. *J. Biol. Chem.* **1997**, 272, 27353-27361.
- [27] Van den Steen, P.; Rudd, P. M.; Proost, P.; Martens, E.; Paemen, L.; Kuster, B.; van Damme, J.; Dwek, R. A.; Opdenakker, G. *Biochim. Biophys. Acta* **1998**, 1425, 587-598.
- [28] Bratosin, D.; Mazurier, J.; Debray, H.; Lecocq, M.; Boilly, B.; Alonso, C.; Moisei, M.; Motas, C.; Montreuil, J. *Glycoconjugate J.* **1995**, 12, 258-267.
- [29] Jaeken, J.; Matthijs, G. *Annu. Rev. Genomics Hum. Genet.* **2001**, 2, 129-151.
- [30] Freeze, H. H. *Glycobiology* **2001**, 11, 129R-143R.
- [31] Martin-Rendon, E.; Blake, D. J. *Trends Pharmacol. Sci.* **2003**, 24, 178-183.
- [32] Schachter, H. *Cell. Mol. Life Sci.* **2001**, 58, 1085-1104.
- [33] Dennis, J. W.; Granovsky, M. In *Carbohydrates in Chemistry and Biology*; Ernst, B., Hart, G. W., Sinay, P., Eds.; 2000; Vol. 4, pp 923-943.
- [34] Jordon, D.; Jagirdar, J.; Kaneko, M. *Am. J. Pathol.* **1989**, 135, 931-937.
- [35] Cooper, H. S.; Malecha, M. J.; Bass, C.; Fagel, P. L.; Steplewski, Z. *Am. J. Pathol.* **1991**, 138, 103-110.
- [36] Sakamoto, S.; Watanabe, T.; Tokumaru, T.; Takagi, H.; Nakazato, H.; Lloyd, K. O. *Cancer Res.* **1989**, 49, 745-752.
- [37] Meanwell, N. A.; Krystal, M. *Drug Discov. Today* **1996**, 1, 388-397.
- [38] Boren, T.; Falk, P.; Roth, K. A.; Larson, G.; Normark, S. *Science* **1993**, 262, 1892-1895.
- [39] Hooper, L. V.; Gordon, J. I. *Glycobiology* **2001**, 11, 1R-10R.
- [40] Linn, S.; Arber, W. *Proc. Natl. Acad. Sci. U. S. A.* **1968**, 59, 1300-1306.

- [41] Maxam, A. M.; Gilbert, W. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 560-564.
- [42] Sanger, F.; Coulson, A. R. *J. Mol. Biol.* **1975**, *94*, 441-448.
- [43] Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 5463-5467.
- [44] Swartz, J. R. *Curr. Opin. Biotechnol.* **2001**, *12*, 195-201.
- [45] Andersen, D. C.; Krummen, L. *Curr. Opin. Biotechnol.* **2002**, *13*, 117-123.
- [46] Baneyx, F. *Curr. Opin. Biotechnol.* **1999**, *10*, 411-421.
- [47] Cereghino, G. P. L.; Cereghino, J. L.; Ilgen, C.; Cregg, J. M. *Curr. Opin. Biotechnol.* **2002**, *13*, 329-332.
- [48] Maras, M.; Van Die, I.; Contreras, R.; Van den Hondel, C. A. M. J. *J. Glycoconjugate J.* **1999**, *16*, 99-107.
- [49] Tanner, W.; Lehle, L. *Biochim. Biophys. Acta* **1987**, *906*, 81-99.
- [50] Herscovics, A.; Orlean, P. *FASEB J.* **1993**, *7*, 540-550.
- [51] Hamilton, S. R.; Bobrowicz, P.; Bobrowicz, B.; Davidson, R. C.; Li, H.; Mitchell, T.; Nett, J. H.; Rausch, S.; Stadheim, T. A.; Wischnewski, H.; Wildt, S.; Gerngross, T. U. *Science* **2003**, *301*, 1244-1246.
- [52] Altmann, F.; Staudacher, E.; Wilson, I. B. H.; Marz, L. *Glycoconjugate J.* **1999**, *16*, 109-123.
- [53] Marchal, I.; Jarvis, D. L.; Cacan, R.; Verbert, A. *Biol. Chem.* **2001**, *382*, 151-159.
- [54] Giddings, G.; Allison, G.; Brooks, D.; Carter, A. *Nat. Biotechnol.* **2000**, *18*, 1151-1155.
- [55] Larrick, J. W.; Thomas, D. W. *Curr. Opin. Biotechnol.* **2001**, *12*, 411-418.
- [56] Ma, J. K. C.; Drake, P. M. W.; Christou, P. *Nat. Rev. Genet.* **2003**, *4*, 794-805.
- [57] Fischer, R.; Stoger, E.; Schillberg, S.; Christou, P.; Twyman, R. M. *Curr. Opin. Plant Biol.* **2004**, *7*, 152-158.
- [58] Lerouge, P.; Bardor, M.; Pagny, S.; Gomord, V.; Faye, L. *Curr. Pharm. Biotechnol.* **2000**, *1*, 347-354.
- [59] Cumming, D. A. *Glycobiology* **1991**, *1*, 115-130.
- [60] Jenkins, N.; Curling, E. M. A. *Enzyme Microb. Technol.* **1994**, *16*, 354-364.
- [61] Jenkins, N.; Paekh, R.; James, D. C. *Nat. Biotechnol.* **1996**, *14*, 975-981.

- [62] Ashwell, G.; Harford, J. *Annu. Rev. Biochem.* **1982**, *51*, 531-554.
- [63] Sunder-Plassmann, G.; Horl, W. H. *Expert Opin. Biol. Th.* **2001**, *1*, 733-739.
- [64] Davis, B. G. *Chem. Rev.* **2002**, *102*, 579-601.
- [65] Davis, B. G. *Science* **2004**, *303*, 480-482.
- [66] Peri, F.; Nicotra, F. *Chem. Commun.* **2004**, 623-627.
- [67] Davis, N. J.; Flitsch, S. L. *Tetrahedron Lett.* **1991**, *32*, 6793-6796.
- [68] Wong, S. Y. C.; Guile, G. R.; Dwek, R. A.; Arsequell, G. *Biochem. J.* **1994**, *300*, 843-850.
- [69] Davis, B. G.; Maughan, M. A. T.; Green, M. P.; Ullman, A.; Jones, J. B. *Tetrahedron: Asymmetry* **2000**, *11*, 245-262.
- [70] Gamblin, D. P.; Garnier, P.; Ward, S. J.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G. *Org. Biomol. Chem.* **2003**, *1*, 3642-3644.
- [71] Gamblin, D. P.; Garnier, P.; van Kasteren, S.; Oldham, N. J.; Fairbanks, A. J.; David, B. G. *Angew. Chem., Int. Ed.* **2004**, *43*, 828-833.
- [72] Rodriguez, E. C.; Marcaurelle, L. A.; Bertozzi, C. R. *J. Org. Chem.* **1998**, *63*, 7134-7135.
- [73] Ni, J.; Singh, S.; Wang, L.-X. *Bioconjugate Chem.* **2003**, *14*, 232-238.
- [74] Liu, H.; Wang, L.; Brock, A.; Wong, C.-H.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 1702-1703.
- [75] Luchansky, S. J.; Argade, S.; Hayes, B. K.; Bertozzi, C. R. *Biochemistry* **2004**, *43*, 12358-12366.
- [76] Takegawa, K.; Fan, J.-Q. In *Methods in Enzymology*; Lee, Y. C., Lee, R. T., Eds.; 2003; Vol. 362, pp 64-74.
- [77] Witte, K.; Sears, P.; Wong, C.-H. *J. Am. Chem. Soc.* **1997**, *119*, 2114-2118.
- [78] Bailey, J. E. *Science* **1991**, *252*, 1668-1675.
- [79] Jacobsen, J. R.; Khosla, C. *Curr. Opin. Chem. Biol.* **1998**, *2*, 133-137.
- [80] Stephanopoulos, G. *Curr. Opin. Biotechnol.* **1994**, *5*, 196-200.
- [81] Cameron, D. C.; Tong, I. T. *Appl. Biochem. Biotechnol.* **1993**, *38*, 105-140.

- [82] Warner, T. G. In *Carbohydrates in Chemistry and Biology*; Ernst, B., Hart, G. W., Sinay, P., Eds.; 2000; Vol. 4, pp 1043-1064.
- [83] Roy, S. K.; Chiba, Y.; Jigami, Y. *Biotechnol. Bioprocess Eng.* **2000**, *5*, 219-226.
- [84] Bretthauer, R. K. *Trends Biotechnol.* **2003**, *21*, 459-462.
- [85] Bobrowicz, P.; Davidson, R. C.; Li, H.; Potgieter, T. I.; Nett, J. H.; Hamilton, S. R.; Stadheim, T. A.; Miele, R. G.; Bobrowicz, B.; Mitchell, T.; Rausch, S.; Renfer, E.; Wildt, S. *Glycobiology* **2004**, *14*, 757-766.
- [86] Vervecken, W.; Kaigorodov, V.; Callewaert, N.; Geysens, S.; De Vusser, K.; Contreras, R. *Appl. Environ. Microbiol.* **2004**, *70*, 2639-2646.
- [87] Weikert, S.; Papac, D.; Briggs, J.; Cowfer, D.; Tom, S.; Gawlitzek, M.; Lofgren, J.; Mehta, S.; Chisholm, V.; Modi, N.; Eppler, S.; Carroll, K.; Chamow, S.; Peers, D.; Berman, P.; Krummen, L. *Nat. Biotechnol.* **1999**, *17*, 1116-1121.
- [88] Grabenhorst, E.; Schlenke, P.; Pohl, S.; Nimtz, M.; Conradt, H. S. *Glycoconjugate J.* **1999**, *16*, 81-97.
- [89] Bragonzi, A.; Distefano, G.; Buckberry, L. D.; Acerbis, G.; Foglieni, C.; Lamotte, D.; Campi, G.; Marc, A.; Soria, M. R.; Jenkins, N.; Monaco, L. *Biochim. Biophys. Acta* **2000**, *1474*, 273-282.
- [90] Jarvis, D. L.; Kwar, Z. S.; Hollister, J. R. *Curr. Opin. Biotechnol.* **1998**, *9*, 528-533.
- [91] Aumiller, J. J.; Hollister, J. R.; Jarvis, D. L. *Glycobiology* **2003**, *13*, 497-507.
- [92] Gomord, V.; Faye, L. *Curr. Opin. Plant Biol.* **2004**, *7*, 171-181.
- [93] Henrissat, B. *Biochem. J.* **1991**, *280*, 309-316.
- [94] Gaboriaud, C.; Bissery, V.; Benchetrit, T.; Mornon, J. P. *FEBS Lett.* **1987**, *224*, 149-155.
- [95] Lemesle-Varloot, L.; Henrissat, B.; Gaboriaud, C.; Bissery, V.; Morgat, A.; Mornon, J. P. *Biochimie* **1990**, *72*, 555-574.
- [96] Chothia, C.; Lesk, A. M. *EMBO J.* **1986**, *5*, 823-826.
- [97] Davies, G.; Henrissat, B. *Structure* **1995**, *3*, 853-859.
- [98] Coutinho, P. M.; Henrissat, B. In *Recent Advances in Carbohydrate Bioengineering*; Gilbert, H. J., Davies, G. J., Henrissat, B., Svensson, B., Eds.; 1999; Vol. 246, pp 3-12.
- [99] Henrissat, B.; Callebaut, I.; Fabrega, S.; Lehn, P.; Mornon, J.-P.; Davies, G. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 7090-7094.

- [100] Henrissat, B.; Bairoch, A. *Biochem. J.* **1993**, *293*, 781-788.
- [101] Henrissat, B.; Bairoch, A. *Biochem. J.* **1996**, *316*, 695-696.
- [102] Rye, C. S.; Withers, S. G. *Curr. Opin. Chem. Biol.* **2000**, *4*, 573-580.
- [103] Rajan, S. S.; Yang, X.; Collart, F.; Yip, V. L. Y.; Withers, S. G.; Varrot, A.; Thompson, J.; Davies, G. J.; Anderson, W. F. *Structure* **2004**, *12*, 1619-1629.
- [104] Yip, V. L. Y.; Varrot, A.; Davies, G. J.; Rajan, S. S.; Yang, X.; Thompson, J.; Anderson, W. F.; Withers, S. G. *J. Am. Chem. Soc.* **2004**, *126*, 8354-8355.
- [105] Yip, V. L. Y.; Withers, S. G. *Org. Biomol. Chem.* **2004**, *2*, 2707-2713.
- [106] Koshland, D. E., Jr. *Biol. Rev., Camb. Philos. Soc.* **1953**, *28*, 416-436.
- [107] Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171-1202.
- [108] McCarter, J. D.; Withers, S. G. *Curr. Opin. Struct. Biol.* **1994**, *4*, 885-892.
- [109] Wang, Q.; Graham, R. W.; Trimbur, D.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1994**, *116*, 11594-11595.
- [110] Henrissat, B.; Davies, G. *Curr. Opin. Struct. Biol.* **1997**, *7*, 637-644.
- [111] Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schuelein, M.; Davies, G. J. *Biochemistry* **1996**, *35*, 15280-15287.
- [112] Withers, S. G.; Mackenzie, L.; Wang, Q. *Methods and compositions for synthesis of oligosaccharides using mutant glycosidase enzymes* (The University of British Columbia, Can.; Withers, Stephen G.; Mackenzie, Lloyd; Wang, Quingping). PCT Int. Appl. 9721822, 1997, 45 pp.
- [113] Mackenzie, L. F.; Wang, Q.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1998**, *120*, 5583-5584.
- [114] Prade, H.; Mackenzie, L. F.; Withers, S. G. *Carbohydr. Res.* **1998**, *305*, 371-381.
- [115] Perugino, G.; Trincone, A.; Rossi, M.; Moracci, M. *Trends Biotechnol.* **2004**, *22*, 31-37.
- [116] Mayer, C.; Jakeman, D. L.; Mah, M.; Karjala, G.; Gal, L.; Warren, R. A. J.; Withers, S. G. *Chem. Biol.* **2001**, *8*, 437-443.
- [117] Mayer, C.; Zechel, D. L.; Reid, S. P.; Warren, R. A. J.; Withers, S. G. *FEBS Lett.* **2000**, *466*, 40-44.
- [118] Okuyama, M.; Mori, H.; Watanabe, K.; Kimura, A.; Chiba, S. *Biosci. Biotechnol. Biochem.* **2002**, *66*, 928-933.

- [119] Hrmova, M.; Imai, T.; Rutten, S. J.; Fairweather, J. K.; Pelosi, L.; Bulone, V.; Driguez, H.; Fincher, G. B. *J. Biol. Chem.* **2002**, *277*, 30102-30111.
- [120] Tolborg, J. F.; Petersen, L.; Jensen, K. J.; Mayer, C.; Jakeman, D. L.; Warren, R. A. J.; Withers, S. G. *J. Org. Chem.* **2002**, *67*, 4143-4149.
- [121] Trincone, A.; Perugino, G.; Rossi, M.; Moracci, M. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 365-368.
- [122] Moracci, M.; Trincone, A.; Cobucci-Ponzano, B.; Perugino, G.; Ciaramella, M.; Rossi, M. *Extremophiles* **2001**, *5*, 145-152.
- [123] Perugino, G.; Trincone, A.; Giordano, A.; Van der Oost, J.; Kaper, T.; Rossi, M.; Moracci, M. *Biochemistry* **2003**, *42*, 8484-8493.
- [124] Trincone, A.; Giordano, A.; Perugino, G.; Rossi, M.; Moracci, M. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4039-4042.
- [125] Nashiru, O.; Zechel, D. L.; Stoll, D.; Mohammadzadeh, T.; Warren, R. A. J.; Withers, S. G. *Angew. Chem., Int. Ed.* **2001**, *40*, 417-420.
- [126] Jakeman, D. L.; Withers, S. G. *Can. J. Chem.* **2002**, *80*, 866-870.
- [127] Malet, C.; Planas, A. *FEBS Lett.* **1998**, *440*, 208-212.
- [128] Fairweather, J. K.; Faijes, M.; Driguez, H.; Planas, A. *ChemBioChem* **2002**, *3*, 866-873.
- [129] Faijes, M.; Perez, X.; Perez, O.; Planas, A. *Biochemistry* **2003**, *42*, 13304-13318.
- [130] Fort, S.; Boyer, V.; Greffe, L.; Davies, G. J.; Moroz, O.; Christiansen, L.; Schuelein, M.; Cottaz, S.; Driguez, H. *J. Am. Chem. Soc.* **2000**, *122*, 5429-5437.
- [131] Fairweather, J. K.; Hrmova, M.; Rutten, S. J.; Fincher, G. B.; Driguez, H. *Chem. Eur. J.* **2003**, *9*, 2603-2610.
- [132] Jahn, M.; Stoll, D.; Warren, R. A. J.; Szabo, L.; Singh, P.; Gilbert, H. J.; Ducros, V. M. A.; Davies, G. J.; Withers, S. G. *Chem. Commun.* **2003**, 1327-1329.
- [133] Jahn, M.; Marles, J.; Warren, R. A. J.; Withers, S. G. *Angew. Chem., Int. Ed.* **2003**, *42*, 352-354.
- [134] Jahn, M.; Withers, S. G. *Biocatal. Biotransform.* **2003**, *21*, 159-166.
- [135] Jahn, M.; Chen, H.; Muellegger, J.; Marles, J.; Warren, R. A. J.; Withers, S. G. *Chem. Commun.* **2004**, 274-275.
- [136] Stick, R. V.; Stubbs, K. A.; Watts, A. G. *Aust. J. Chem.* **2004**, *57*, 779-786.
- [137] Schuelein, M. *J. Biotechnol.* **1997**, *57*, 71-81.

- [138] Kleywegt, G. J.; Zou, J.-Y.; Divne, C.; Davies, G. J.; Sinning, I.; Stahlberg, J.; Reinikainen, T.; Srisodsuk, M.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1997**, *272*, 383-397.
- [139] Sulzenbacher, G.; Schuelein, M.; Davies, G. J. *Biochemistry* **1997**, *36*, 5902-5911.
- [140] Davies, G.; Tolley, S.; Wilson, K.; Schulein, M.; Woldike, H. F.; Dodson, G. *J. Mol. Biol.* **1992**, *228*, 970-972.
- [141] Mackenzie, L. F.; Sulzenbacher, G.; Divne, C.; Jones, T. A.; Woldike, H. F.; Schulein, M.; Withers, S. G.; Davies, G. J. *Biochem. J.* **1998**, *335*, 409-416.
- [142] Schou, C.; Rasmussen, G.; Kaltoft, M. B.; Henrissat, B.; Schulein, M. *Eur. J. Biochem.* **1993**, *217*, 947-953.
- [143] Ducros, V. M. A.; Tarling, C. A.; Zechel, D. L.; Brzozowski, A. M.; Frandsen, T. P.; von Ossowski, I.; Schulein, M.; Withers, S. G.; Davies, G. J. *Chem. Biol.* **2003**, *10*, 619-628.
- [144] Fort, S.; Christiansen, L.; Schulein, M.; Cottaz, S.; Driguez, H. *Isr. J. Chem.* **2000**, *40*, 217-221.
- [145] Fajjes, M.; Fairweather, J. K.; Driguez, H.; Planas, A. *Chem. Eur. J.* **2001**, *7*, 4651-4655.
- [146] Boyer, V.; Fort, S.; Frandsen, T. P.; Schulein, M.; Cottaz, S.; Driguez, H. *Chem. Eur. J.* **2002**, *8*, 1389-1394.
- [147] Penning, T. M.; Jez, J. M. *Chem. Rev.* **2001**, *101*, 3027-3046.
- [148] Powell, K. A.; Ramer, S. W.; Del Cardayre, S. B.; Stemmer, W. P. C.; Tobin, M. B.; Longchamp, P. F.; Huisman, G. W. *Angew. Chem., Int. Ed.* **2001**, *40*, 3948-3959.
- [149] Cherry, J. R.; Fidantsef, A. L. *Curr. Opin. Biotechnol.* **2003**, *14*, 438-443.
- [150] Dalby, P. A. *Curr. Opin. Struct. Biol.* **2003**, *13*, 500-505.
- [151] Harris, J. L.; Craik, C. S. *Curr. Opin. Chem. Biol.* **1998**, *2*, 127-132.
- [152] Jones, J. B.; DeSantis, G. *Acc. Chem. Res.* **1999**, *32*, 99-107.
- [153] Cedrone, F.; Menez, A.; Quemeneur, E. *Curr. Opin. Struct. Biol.* **2000**, *10*, 405-410.
- [154] Hult, K.; Berglund, P. *Curr. Opin. Biotechnol.* **2003**, *14*, 395-400.
- [155] Davis, B. G. *Curr. Opin. Biotechnol.* **2003**, *14*, 379-386.
- [156] Quijoch, F. A.; Richards, F. M. *Proc. Natl. Acad. Sci. U. S. A.* **1964**, *52*, 833-839.

- [157] DeSantis, G.; Jones, J. B. *Curr. Opin. Biotechnol.* **1999**, *10*, 324-330.
- [158] Marshall, S. A.; Lazar, G. A.; Chirino, A. J.; Desjarlais, J. R. *Drug Discov. Today* **2003**, *8*, 212-221.
- [159] Saviranta, P.; Haavisto, T.; Rappu, P.; Karp, M.; Loevgren, T. *Bioconjugate Chem.* **1998**, *9*, 725-735.
- [160] Parker, L. L.; Atherton-Fessler, S.; Lee, M. S.; Ogg, S.; Falk, J. L.; Swenson, K. I.; Piwnica-Worms, H. *EMBO J.* **1991**, *10*, 1255-1263.
- [161] Zhang, Z. Y.; Clemens, J. C.; Schubert, H. L.; Stuckey, J. A.; Fischer, M. W. F.; Hume, D. M.; Saper, M. A.; Dixon, J. E. *J. Biol. Chem.* **1992**, *267*, 23759-23766.
- [162] Hutchison, C. A., III; Phillips, S.; Edgell, M. H.; Gillam, S.; Jahnke, P.; Smith, M. J. *Biol. Chem.* **1978**, *253*, 6551-6560.
- [163] Winter, G.; Fersht, A. R.; Wilkinson, A. J.; Zoller, M.; Smith, M. *Nature* **1982**, *299*, 756-758.
- [164] Saiki, R. K.; Scharf, S.; Faloona, F.; Mullis, K. B.; Horn, G. T.; Erlich, H. A.; Arnheim, N. *Science* **1985**, *230*, 1350-1354.
- [165] Ling, M. M.; Robinson, B. H. *Anal. Biochem.* **1997**, *254*, 157-178.
- [166] Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.
- [167] Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182-188.
- [168] Bain, J. D.; Diala, E. S.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. *J. Am. Chem. Soc.* **1989**, *111*, 8013-8014.
- [169] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235-242.
- [170] Cunningham, B. C.; Wells, J. A. *Science* **1989**, *244*, 1081-1085.
- [171] Stemmer, W. P. C. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 10747-10751.
- [172] Liao, H.; McKenzie, T.; Hageman, R. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 576-580.
- [173] Chen, K.; Arnold, F. H. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 5618-5622.
- [174] Stemmer, W. P. C. *Nature* **1994**, *370*, 389-391.
- [175] Strausberg, S. L.; Alexander, P. A.; Gallagher, D. T.; Gilliland, G. L.; Barnett, B. L.; Bryan, P. N. *Bio/Technology* **1995**, *13*, 669-673.

- [176] Zhang, J. H.; Dawes, G.; Stemmer, W. P. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4504-4509.
- [177] Reetz, M. T.; Zonta, A.; Schimossek, K.; Liebeton, K.; Jaeger, K.-E. *Angew. Chem., Int. Ed.* **1998**, *36*, 2830-2832.
- [178] Zhao, H.; Arnold, F. H. *Protein Eng.* **1999**, *12*, 47-53.
- [179] van der Veen, B. A.; Potocki-Veronese, G.; Albenne, C.; Joucla, G.; Monsan, P.; Remaud-Simeon, M. *FEBS Lett.* **2004**, *560*, 91-97.
- [180] Taguchi, S.; Ozaki, A.; Momose, H. *Appl. Environ. Microbiol.* **1998**, *64*, 492-495.
- [181] Christians, F. C.; Loeb, L. A. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 6124-6128.
- [182] Lin, H.; Cornish, V. W. *Angew. Chem., Int. Ed.* **2002**, *41*, 4402-4425.
- [183] Deacon, S. E.; Mahmoud, K.; Spooner, R. K.; Firbank, S. J.; Knowles, P. F.; Phillips, S. E. V.; McPherson, M. J. *ChemBioChem* **2004**, *5*, 972-979.
- [184] Sun, L.; Bulter, T.; Alcalde, M.; Petrounia, I. P.; Arnold, F. H. *ChemBioChem* **2002**, *3*, 781-783.
- [185] Thorson, J. S.; Barton, W. A.; Hoffmeister, D.; Albermann, C.; Nikolov, D. B. *ChemBioChem* **2004**, *5*, 16-25.
- [186] Andrews, S. R.; Charnock, S. J.; Lakey, J. H.; Davies, G. J.; Claeysens, M.; Nerinckx, W.; Underwood, M.; Sinnott, M. L.; Warren, R. A. J.; Gilbert, H. J. *J. Biol. Chem.* **2000**, *275*, 23027-23033.
- [187] Kaper, T.; Lebbink, J. H. G.; Pouwels, J.; Kopp, J.; Schulz, G. E.; Van der Oost, J.; de Vos, W. M. *Biochemistry* **2000**, *39*, 4963-4970.
- [188] Corbett, K.; Fordham-Skelton, A. P.; Gatehouse, J. A.; Davis, B. G. *FEBS Lett.* **2001**, *509*, 355-360.
- [189] Kim, Y.-W.; Lee, S. S.; Warren, R. A. J.; Withers, S. G. *J. Biol. Chem.* **2004**, *279*, 42787-42793.
- [190] Schou, C.; Rasmussen, G.; Schulein, M.; Henrissat, B.; Driguez, H. *J. Carbohydr. Chem.* **1993**, *12*, 743-752.
- [191] Suzuki, Y.; Sato, K.; Kiso, M.; Hasegawa, A. *Glycoconjugate J.* **1990**, *7*, 349-354.
- [192] Cohen-Anisfeld, S. T.; Lansbury, P. T., Jr. *J. Am. Chem. Soc.* **1993**, *115*, 10531-10537.
- [193] Wang, L.-X.; Tang, M.; Suzuki, T.; Kitajima, K.; Inoue, Y.; Inoue, S.; Fan, J.-Q.; Lee, Y. C. *J. Am. Chem. Soc.* **1997**, *119*, 11137-11146.

- [194] Windridge, G.; Jorgensen, E. C. *J. Am. Chem. Soc.* **1971**, *93*, 6318-6319.
- [195] Masuda, M.; Shimizu, T. *J. Carbohydr. Chem.* **1998**, *17*, 405-416.
- [196] Zemplén, G.; Pacsu, E. *Ber.* **1929**, *62B*, 1613-1614.
- [197] Shiozaki, M.; Mochizuki, T.; Hanzawa, H.; Haruyama, H. *Carbohydr. Res.* **1996**, *288*, 99-108.
- [198] Anisfeld, S. T.; Lansbury, P. T., Jr. *J. Org. Chem.* **1990**, *55*, 5560-5562.
- [199] Bodanszky, M.; Martinez, J. *J. Org. Chem.* **1978**, *43*, 3071-3073.
- [200] Bodanszky, M.; Natarajan, S. *J. Org. Chem.* **1975**, *40*, 2495-2499.
- [201] Meinjohanns, E.; Meldal, M.; Paulsen, H.; Dwek, R. A.; Bock, K. *J. Chem. Soc., Perkin Trans. 1* **1998**, 549-560.
- [202] Kobayashi, K.; Tawada, E.; Akaike, T.; Usui, T. *Biochim. Biophys. Acta* **1997**, *1336*, 117-122.
- [203] Hayashi, M.; Hashimoto, S.; Noyori, R. *Chem. Lett.* **1984**, 1747-1750.
- [204] Jünnemann, J.; Thiem, J.; Pedersen, C. *Carbohydr. Res.* **1993**, *249*, 91-94.
- [205] Excoffier, G.; Gagnare, D.; Utille, J. P. *Carbohydr. Res.* **1975**, *39*, 368-373.
- [206] Kuhn, R.; Lutz, P.; MacDonald, D. L. *Chem. Ber.* **1966**, *99*, 611-617.
- [207] Roy, R.; Laferriere, C. A. *Can. J. Chem.* **1990**, *68*, 2045-2054.
- [208] Eisele, T.; Schmidt, R. R. *Liebigs Ann. Recl.* **1997**, 865-872.
- [209] Wolfrom, M. L.; Thompson, A. In *Methods in Carbohydrate Chemistry*; Whistler, R. L., Wolfrom, M. L., Eds.; 1963; Vol. 2, pp 211-215.
- [210] Jansson, K.; Ahlfors, S.; Frejd, T.; Kihlberg, J.; Magnusson, G.; Dahmen, J.; Noori, G.; Stenvall, K. *J. Org. Chem.* **1988**, *53*, 5629-5647.
- [211] Kartha, K. P. R.; Kameyama, A.; Kiso, M.; Hasegawa, A. *J. Carbohydr. Chem.* **1989**, *8*, 145-158.
- [212] Rich, J. R.; Bundle, D. R. *Org. Lett.* **2004**, *6*, 897-900.
- [213] Card, P. J. *J. Org. Chem.* **1983**, *48*, 393-395.
- [214] Lindhorst, T. K.; Braun, C.; Withers, S. G. *Carbohydr. Res.* **1995**, *268*, 93-106.
- [215] Gan, Z.; Roy, R. *Can. J. Chem.* **2002**, *80*, 908-916.

- [216] Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673-4680.
- [217] Coutinho, P. M.; Dowd, M. K.; Reilly, P. J. *Carbohydr. Res.* **1997**, *297*, 309-324.
- [218] Coutinho, P. M.; Dowd, M. K.; Reilly, P. J. *Proteins* **1997**, *28*, 162-173.
- [219] Coutinho, P. M.; Dowd, M. K.; Reilly, P. J. *Ind. Eng. Chem. Res.* **1998**, *37*, 2148-2157.
- [220] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639-1662.
- [221] Vriend, G. *J. Mol. Graph.* **1990**, *8*, 52-56.
- [222] Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1-42.
- [223] Mazeau, K.; Perez, S.; Rinaudo, M. *J. Carbohydr. Chem.* **2000**, *19*, 1269-1284.
- [224] Ramakrishnan, B.; Qasba Pradman, K. *J. Biol. Chem.* **2002**, *277*, 20833-20839.
- [225] Marcus, S. L.; Polakowski, R.; Seto, N. O. L.; Leinala, E.; Borisova, S.; Blancher, A.; Roubinet, F.; Evans, S. V.; Palcic, M. M. *J. Biol. Chem.* **2003**, *278*, 12403-12405.
- [226] Shimizu, M.; Togo, H.; Yokoyama, M. *Synthesis* **1998**, 799-822.
- [227] Mellor, J.; Dobson, M. J.; Roberts, N. A.; Tuite, M. F.; Emtage, J. S.; White, S.; Lowe, P. A.; Patel, T.; Kingsman, A. J.; Kingsman, S. M. *Gene* **1983**, *24*, 1-14.
- [228] Rietsch, A.; Beckwith, J. *Annu. Rev. Genet.* **1998**, *32*, 163-184.
- [229] Makrides, S. C. *Microbiol. Rev.* **1996**, *60*, 512-538.
- [230] Penninga, D.; van der Veen, B. A.; Knegtel, R. M. A.; van Hijum, S. A. F. T.; Rozeboom, H. J.; Kalk, K. H.; Dijkstra, B. W.; Dijkhuizen, L. *J. Biol. Chem.* **1996**, *271*, 32777-32784.
- [231] Christensen, T.; Woeldike, H.; Boel, E.; Mortensen, S. B.; Hjortshoej, K.; Thim, L.; Hansen, M. T. *Bio/Technology* **1988**, *6*, 1419-1422.
- [232] Penttila, M. E.; Andre, L.; Saloheimo, M.; Lehtovaara, P.; Knowles, J. K. C. *Yeast* **1987**, *3*, 175-185.
- [233] Nyffeler, P. T.; Liang, C.-H.; Koeller, K. M.; Wong, C.-H. *J. Am. Chem. Soc.* **2002**, *124*, 10773-10778.
- [234] Vauzeilles, B.; Dausse, B.; Palmier, S.; Beau, J. M. *Tetrahedron Lett.* **2001**, *42*, 7567-7570.

- [235] Armarego, W. L. F.; Perrin, D. D. *Purification of Laboratory Chemicals*; 4th ed.; Pergamon: Oxford, UK, 1997.
- [236] Gomori, G. In *Methods in Enzymology*; Colowick, S. P., Kaplan, N. O., Eds.; 1955; Vol. 1, pp 138-146.
- [237] Armand, S.; Drouillard, S.; Schuelein, M.; Henrissat, B.; Driguez, H. *J. Biol. Chem.* **1997**, 272, 2709-2713.
- [238] Baggett, N.; Marsden, B. J. *Carbohydr. Res.* **1982**, 110, 11-18.
- [239] Sharma, M. N.; Eby, R. *Carbohydr. Res.* **1984**, 127, 201-210.
- [240] Hasegawa, A.; Nakamura, J.; Kiso, M. *J. Carbohydr. Chem.* **1986**, 5, 11-19.
- [241] Hasegawa, A.; Morita, M.; Ito, Y.; Ishida, H.; Kiso, M. *J. Carbohydr. Chem.* **1990**, 9, 369-392.
- [242] Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning : A Laboratory Manual*; 2nd ed.; Cold Spring Harbor Laboratory Press: New York, USA, 1989.
- [243] Bradford, M. M. *Anal. Biochem.* **1976**, 72, 248-254.
- [244] Laemmli, U. K. *Nature* **1970**, 227, 680-685.
- [245] Bron, S. In *Molecular Biological Methods for Bacillus*; Harwood, C. R., Cutting, S. M., Eds.; 1990, pp 146-147.
- [246] Lin, H.; Tao, H.; Cornish, V. W. *J. Am. Chem. Soc.* **2004**, 126, 15051-15059.

ANNEXES

ANNEXES

Liste des abréviations

A	Ac	acétyle
	ADN	acide désoxyribonucléique
	ADNc	ADN complémentaire
	All	allyle
	Ara	arabinose
	ARN	acide ribonucléique
	ARNt	ARN de transfert
	ATP	adénosine triphosphate
B	BHK	"Baby Hamster Kidney"
	Bn	benzyle
	Boc	<i>tert</i> -butyloxycarbonyle
	BOP	hexafluorophosphate de benzotriazol-1-yl-oxy-tris(diméthylamino)phosphonium
C	CAZy	"Carbohydrate-Active enZYmes"
	CBD	"Cellulose Binding Domain"
	CCM	chromatographie sur couche mince
	CDG	"Congenital Disorder of Glycosylation"
	CHO	"Chinese Hamster Ovary"
	CMC	carboxyméthylcellulose
	CMP	cytidine monophosphate
	COSY	"COrelated SpectroscopY"
	CTP	cytidine triphosphate
D	d	doublet
	DAST	trifluorure de diéthylaminosulfure
	DCC	<i>N,N'</i> -dicyclohexylcarbodiimide
	DCI	"Desorption Chemical Ionization"
	dd	doublet de doublet
	DEPT	"Distortionless Enhancement by Polarization Transfert"
	DMAP	diméthylaminopyridine
	DMF	<i>N,N</i> -diméthylformamide
	DNA	"DeoxyriboNucleic Acid"
	DNS	acide 3,5-dinitrosalicylique
	Dol	dolichol
	DP	degré de polymérisation
	E	EPO
éq.		équivalent
ES		électrospray

	Et	éthyle
F	<i>f</i>	furanose
	FAB	"Fast Atom Bombardment"
	FRET	"Fluorescence Resonance Energy Transfer"
	Fuc	fucose
	FucNAc	<i>N</i> -acétylfucosamine
G	Gal	galactose
	GalNAc	<i>N</i> -acétylgalactosamine
	GH	glycoside hydrolase
	Glc	glucose
	GlcNAc	<i>N</i> -acétylglucosamine
	GP	glycoprotéine
	GPI	glycosylphosphatidylinositol
H	HCA	"Hydrophobic Cluster Analysis"
	HMQC	"Heteronuclear Multiple Quantum Correlation"
	HOBt	<i>N</i> -hydroxybenzotriazole
	HPLC	"High Performance Liquid Chromatography"
	Hyl	hydroxylysine
	Hyp	hydroxyproline
I	IPTG	isopropyl- β -D-thiogalactopyranoside
	IUB-MB	"International Union of Biochemistry and Molecular Biology"
K	kb	kilobase
L	Le ^x	Lewis x
	Le ^y	Lewis y
M	m	multiplet
	M	mole par litre
	MALDI	"Matrix Assisted Laser Desorption Ionisation"
	Man	mannose
	ManNAc	<i>N</i> -acétylmannosamine
	Me	méthyle
N	NCBI	"National Center for Biotechnology Information"
	Neu5Ac	acide <i>N</i> -acétylneuraminique
	Neu5Gc	acide <i>N</i> -glycolylneuraminique
O	<i>o</i> NP	<i>ortho</i> -nitrophényle
	OST	oligosaccharyltransférase
P	P	phosphate
	PAGE	"PolyAcrylamide Gel Electrophoresis"
	pb	paire de bases
	PCR	"Polymerase Chain Reaction"
	PEG	polyéthylène glycol

	PEP	phosphoénolpyruvate
	Ph	phényle
	<i>p</i> NP	<i>para</i> -nitrophényle
	ppm	partie par million
Q	q	quadruplet
	quant.	quantitatif
R	RE	réticulum endoplasmique
	REL	réticulum endoplasmique lisse
	RER	réticulum endoplasmique rugueux
	RMN	résonance magnétique nucléaire
	RX	rayons X
S	s	singulet
	SAB	sérum albumine bovine
	SDS	sodium dodécyl sulfate
	SE	(triméthylsilyl)éthyle
	sLe ^a	sialyl Lewis a
	sLe ^x	sialyl Lewis x
	SM	spectrométrie de masse
T	t	triplet
	TBAHS	hydrogénosulfate de tétrabutylammonium
	TEMED	<i>N,N,N',N'</i> -tétraméthyléthylènediamine
	Tf	trifluorométhane sulfonyle
	TFA	acide trifluoroacétique
	Ts	<i>para</i> -toluène sulfonyle
U	UDP	uridine diphosphate
X	Xyl	xylose
Z	Z	benzyloxycarbonyle

Code à trois lettres et à une lettre des acides aminés

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cystéine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Méthionine	Met	M
Phénylalanine	Phe	F
Proline	Pro	P
Sérine	Ser	S
Thréonine	Thr	T
Tryptophane	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Liste des Figures

Figure 1 : Rôles biologiques des protéines des organismes vivants.	- 3 -
Figure 2 : Représentation schématique d'une cellule eucaryote.....	- 5 -
Figure 3 : Liaisons sucre-protéine connues chez les procaryotes et les eucaryotes ^{a,b} [2,5]..	- 8 -
Figure 4 : Représentation des 5 types connus de glycosylation des protéines [2].	- 9 -
Figure 5 : Biosynthèse des <i>N</i> -glycoprotéines [5].	- 11 -
Figure 6 : Structure du dolichol.....	- 11 -
Figure 7 : Structure du précurseur tétradécasaccharidique des glycanes des <i>N</i> -glycoprotéines.	- 12 -
Figure 8 : Structures des trois types de <i>N</i> -glycanes rencontrés chez les mammifères.	- 14 -
Figure 9 : Structures des noyaux identifiées à ce jour dans les chaînes <i>O</i> -glycaniques de type mucine.	- 16 -
Figure 10 : Fonctions de reconnaissance des groupements glycaniques.	- 19 -
Figure 11 : Critères dans le choix du système d'expression de protéines hétérologues.	- 31 -
Figure 12 : Routes potentielles de la biosynthèse de l'acide <i>N</i> -acétylneuraminique.....	- 38 -
Figure 13 : Mécanisme d'une β -D-glycoside hydrolase agissant avec inversion de configuration.	- 44 -
Figure 14 : Mécanisme d'une β -D-glycoside hydrolase agissant avec rétention de configuration.	- 45 -
Figure 15 : Topologie du site actif des glycoside hydrolases [97].....	- 46 -
Figure 16 : Structure aux RX du complexe entre Cel7B de <i>Fusarium oxysporum</i> et le thio- cellopentaoside de méthyle (inhibiteur) [111].	- 47 -
Figure 17 : Mécanisme d'action d'une glycosynthase.	- 48 -
Figure 18 : Mécanisme d'une thioglycoligase (A) et d'une thioglycosynthase (B).....	- 53 -
Figure 19 : Exemples de principes d'ingénierie des protéines disponibles à l'heure actuelle ^a	- 60 -
Figure 20 : Approche envisagée pour le remodelage de la <i>N</i> -glycosylation des protéines recombinantes produites chez les plantes.....	- 70 -
Figure 21 : Réaction de couplage avec la glycosynthase Cel7B E197A entre le donneur 16 et l'accepteur 14	- 73 -
Figure 22 : Principe de l'approche convergente.	- 74 -
Figure 23 : Formation de l'aspartimide pendant la synthèse convergente d'un <i>N</i> -glycopeptide.	- 78 -
Figure 24 : Réaction de couplage entre le donneur 31 et l'accepteur β -D-glucopyranoside de <i>p</i> -nitrophényle.	- 81 -
Figure 25 : Superposition d'un cellononaose et d'un résidu <i>N</i> -acétylglucosaminyle dans le sous-site +1.....	- 92 -
Figure 26 : Superposition d'un résidu <i>N</i> -acétylglucosaminyle dans le sous-site +1 de Cel7B d' <i>H. insolens</i>	- 93 -
Figure 27 : Alignement de séquences d'enzymes de la famille GH7 des glycoside hydrolases.	- 95 -
Figure 28 : Cello-oligosaccharides modélisés à partir du cellononaose.	- 99 -
Figure 29 : Les différentes conformations* du groupe <i>N</i> -acétyle dans le <i>N</i> - acétylglucosaminide de méthyle [223].....	- 100 -
Figure 30 : Vues (A) et (B) d'une solution de docking du cellotriose A dans le site actif de l'enzyme Cel7B d' <i>H. insolens</i>	- 104 -

Figure 31 : Vues (A) et (B) d'une solution de docking du trisaccharide VII dans le site actif de l'enzyme Cel7B H209A d' <i>H. insolens</i>	- 106 -
Figure 32 : (A) Agrandissement de la solution de docking du trisaccharide VII dans le site actif de l'enzyme mutée Cel7B H209A d' <i>H. insolens</i> ; (B) Superposition de la même solution de docking dans le site actif de l'enzyme native Cel7B d' <i>H. insolens</i>	- 107 -
Figure 33 : Cellononaose dans le site actif (A) de l'enzyme Cel7B d' <i>H. insolens</i> et (B) du mutant Cel7B H209A d' <i>H. insolens</i>	- 108 -
Figure 34 : Positionnement du tétrasaccharidique 34 dans le site actif de l'endoglucanase	- 109 -
Figure 35 : Analyse des endoglucanases Cel7B d' <i>H. insolens</i> et de <i>T. reesei</i> sur gel de polyacrylamide à 10% en conditions dénaturantes.....	- 111 -
Figure 36 : Analyse par HPLC de la réaction d'hydrolyse du tétrasaccharide 34 par une endoglucanase.....	- 112 -
Figure 37 : Mesure de l'activité hydrolytique des endoglucanases Cel7B vis-à-vis du tétrasaccharide 34	- 113 -
Figure 38 : Activité relative d'hydrolyse entre les deux endoglucanases Cel7B vis-à-vis du cellotétraose.....	- 113 -
Figure 39 : Les éléments de contrôle du système pET.....	- 120 -
Figure 40 : Analyse des mutants de la glycosynthase d' <i>H. insolens</i> sur gel de polyacrylamide à 10% en conditions dénaturantes.....	- 125 -
Figure 41 : Analyse par CCM du couplage enzymatique entre le donneur 16 et le cellobiose par Cel7B E196A de <i>T. reesei</i> , après 48 h d'incubation.....	- 127 -
Figure 42 : Analyse par HPLC de la réaction de couplage entre le donneur 16 et le cellobiose catalysée par une glycosynthase.....	- 128 -
Figure 43 : Mesure de l'activité glycosynthase des différents mutants de Cel7B E197A d' <i>H. insolens</i>	- 129 -
Figure 44 : Analyse par HPLC de la réaction de couplage enzymatique catalysée par la glycosynthase Cel7B E197A H209A d' <i>H. insolens</i> entre le donneur 16 et l'accepteur 38 ..	- 135 -
Figure 45 : Analyse par CCM du couplage entre le donneur de lactosyle 16 et l'accepteur 39 ...	- 137 -
Figure 46 : Spectres RMN ¹ H (A) du tétrasaccharide 42 et (B) du tétrasaccharide 43 (CDCl ₃ , 400 MHz, 298 K).....	- 140 -
Figure 47 : Spectres RMN ¹³ C (A) du tétrasaccharide 42 et (B) du tétrasaccharide 43 (CDCl ₃ , 400 MHz, 298 K).....	- 141 -
Figure 48 : Analyse par CCM du couplage entre le donneur de lactosyle 16 et l'accepteur 39 après 24 h d'incubation.....	- 143 -
Figure 49 : Fichier caractéristique regroupant les paramètres utilisés par AutoGrid.....	- 148 -
Figure 50 : Fichier caractéristique regroupant les paramètres utilisés par AutoDock.....	- 149 -

Liste des Tableaux

Tableau 1 : Exemples de modifications post-traductionnelles des protéines.....	- 7 -
Tableau 2 : Types de médicaments à base de protéines.....	- 21 -
Tableau 3 : Caractéristiques des profils de <i>N</i> -glycosylation des différents systèmes d'expression de protéines hétérologues.....	- 32 -
Tableau 4 : Modifications génétiques utiles pour améliorer la <i>N</i> -glycosylation des protéines recombinantes selon le système d'expression utilisé [82].....	- 37 -
Tableau 5 : <i>Exo</i> -glycosynthases connues.....	- 51 -
Tableau 6 : <i>Endo</i> -glycosynthases connues.....	- 52 -
Tableau 7 : Exploration de la spécificité de la glycosynthase dérivée de Cel7B d' <i>H. insolens</i> au niveau de l'accepteur [130].....	- 56 -
Tableau 8 : Exemples de propriétés des enzymes pouvant être améliorées ou modifiées..	- 59 -
Tableau 9 : Quelques exemples d'enzymes modifiées obtenues par évolution dirigée.....	- 65 -
Tableau 10 : Structures PDB disponibles à ce jour pour la famille GH7 des glycoside hydrolases.....	- 91 -
Tableau 11 : Les différents trisaccharides modélisés dans le site actif de l'enzyme.....	- 101 -
Tableau 12 : Les différentes expériences de docking réalisées.....	- 102 -
Tableau 13 : Docking des cellotrioses A et C dans le site actif de l'endoglucanase Cel7B d' <i>H.</i> <i>insolens</i> ^a	- 103 -
Tableau 14 : Docking de trisaccharides I-VIII dans le site actif de l'endoglucanase mutée Cel7B H209A d' <i>H. insolens</i> ^a	- 105 -
Tableau 15 : Liste des différentes mutations réalisées sur Cel7B d' <i>H. insolens</i>	- 116 -
Tableau 16 : Etude d'expression de Cel7B d' <i>H. insolens</i> chez <i>E. coli</i>	- 122 -
Tableau 17 : Les différentes réactions de couplage enzymatique réalisées avec les mutants de Cel7B E197A d' <i>H. insolens</i>	- 131 -
Tableau 18 : Résultats de l'analyse LC-MS du couplage enzymatique entre le donneur 16 et l'accepteur 38 par la glycosynthase Cel7B E197A H209A d' <i>H. insolens</i>	- 135 -
Tableau 19 : Structures des deux régioisomères produits par les glycosynthases dérivées de Cel7B d' <i>H. insolens</i>	- 139 -
Tableau 20 : Proportions de chaque type de liaisons formées par chaque glycosynthase au cours de la réaction de couplage entre le donneur 16 et l'accepteur 39	- 144 -
Tableau 21 : Liste des différentes souches d' <i>Escherichia coli</i> utilisées.....	- 190 -
Tableau 22 : Liste des différents vecteurs plasmidiques utilisés.....	- 190 -
Tableau 23 : Séquences des amorces pour l'insertion du site de restriction <i>NotI</i> en 3' de Cel7B d' <i>H. insolens</i>	- 193 -
Tableau 24 : Séquences des amorces pour la mutation du site de restriction <i>NcoI</i> au niveau de l'alanine 197 de Cel7B E197A d' <i>H. insolens</i>	- 194 -
Tableau 25 : Séquences des amorces utilisées pour la délétion du peptide signal de Cel7B d' <i>H.</i> <i>insolens</i>	- 194 -
Tableau 26 : Séquences des amorces utilisées pour l'insertion des sites de restriction <i>NcoI</i> et <i>NotI</i> dans pDP66K.....	- 195 -
Tableau 27 : Séquences des amorces utilisées pour la mutagenèse dirigée sur Cel7B d' <i>H.</i> <i>insolens</i>	- 196 -
Tableau 28 : Séquences des amorces utilisées pour la mutation du nucléophile de Cel7B de <i>T.</i> <i>reesei</i>	- 197 -

Tableau 29 : Composition des milieux réactionnels de couplage entre le donneur 16 et le cellobiose.....	- 203 -
--	---------

Liste des Schémas

Schéma 1 : Polymérisation enzymatique de fluorures de cellobiosyle substitués catalysée par la glycosynthase Cel7B E197A d' <i>H. insolens</i> [130].	- 57 -
Schéma 2 : Préparation du pentapeptide 10	- 75 -
Schéma 3 : Préparation de la β -D-glucosylamine 13	- 77 -
Schéma 4 : Préparation du glucopeptide 14	- 78 -
Schéma 5 : Préparation du fluorure d' α -lactosyle 16	- 79 -
Schéma 6 : Couplage enzymatique entre le donneur 16 et l'accepteur 14	- 80 -
Schéma 7 : Rétrosynthèse du composé 31	- 82 -
Schéma 8 : Préparation du précurseur sialyle 20	- 83 -
Schéma 9 : Préparation du précurseur lactoside 25	- 84 -
Schéma 10 : Thioglycosylation entre le donneur 20 et l'accepteur 25	- 85 -
Schéma 11 : Couplage enzymatique entre le donneur 31 et le β -D-glucopyranoside de <i>p</i> -nitrophényle.....	- 86 -
Schéma 12 : Synthèse du composé 34	- 110 -
Schéma 13 : Hydrolyse enzymatique du tétrasaccharide 34	- 111 -
Schéma 14 : Couplage enzymatique entre le donneur 36 et le N^I, N^{II} -diacétylchitobiose. -	114 -
Schéma 15 : Préparation du fluorure de β -lactosyle 36	- 114 -
Schéma 16 : Couplage enzymatique entre le donneur 16 et le cellobiose.	- 126 -
Schéma 17 : Préparation des disaccharides 38 et 39	- 132 -
Schéma 18 : Couplage enzymatique entre le donneur 16 et l'accepteur 39	- 136 -

Ingénierie de glycoside hydrolases pour la glycosylation des proteines recombinantes

Le contrôle de la glycosylation des protéines recombinantes présente un enjeu considérable dans le développement de la production de protéines d'intérêt thérapeutique dans des systèmes d'expression hétérologues. La glycosylation joue en effet un rôle essentiel dans leurs propriétés, notamment pharmacocinétiques. Le remodelage de la *N*-glycosylation des protéines recombinantes a été envisagé *via* l'utilisation de la glycosynthase Cel7B E197A d'*Humicola insolens*. Cette enzyme doit tout d'abord être modifiée afin de pouvoir fixer un groupement *N*-acétylglucosaminyle dans le sous-site accepteur +1. Des études de modélisation moléculaire ont mis en évidence deux acides aminés qui pourraient empêcher le positionnement d'une unité glucosidique substituée en position 2. Différents mutants ont été préparés par mutagenèse dirigée afin d'étudier leur spécificité de substrat. Leurs activités glycosynthases ont été caractérisées, montrant l'influence des mutations introduites. Même si la spécificité de substrat désirée n'a pu être obtenue, une modification de la régiosélectivité de la glycosynthase a été mise en évidence vis-à-vis d'un substrat substitué en position 2 par un groupement de type azido.

MOTS CLES : ingénierie, mutagenèse dirigée, glycoside hydrolases, glycosynthases, spécificité de substrat, glycosylation, protéines recombinantes.

Enzyme engineering of glycoside hydrolases for the glycosylation of recombinant proteins

The control of glycosylation of recombinant proteins presents a considerable stake in the development of therapeutic proteins production in heterologous expression systems. Indeed glycosylation plays a key role in their properties, in particular pharmacokinetic properties. The remodelling of *N*-glycosylation of recombinant proteins was considered *via* the use of glycosynthase Cel7B E197A from *Humicola insolens*. This enzyme must first of all be modified in order to be able to accomodate an *N*-acetylglucosaminyl group in the acceptor subsite +1. Molecular modelling studies highlighted two amino acids which could prevent from positioning a carbohydrate unit substituted in position 2. Various mutants were prepared by site-directed mutagenesis in order to study their substrate specificity. Their glycosynthase activities were characterized, showing the influence of the introduced mutations. Even if the desired substrate specificity has not been obtained, a modification of the regioselectivity of the glycosynthase was highlighted with a substrate substituted in position 2 by an azido group.

KEYWORDS : enzyme engineering, site-directed mutagenesis, glycoside hydrolases, glycosynthases, substrate specificity, glycosylation, recombinant proteins.