



HAL
open science

Exploration du microbiote d'invertébrés par métagénomique fonctionnelle et caractérisation structure-fonction d'une nouvelle xylanase

Barbara Guyez

► **To cite this version:**

Barbara Guyez. Exploration du microbiote d'invertébrés par métagénomique fonctionnelle et caractérisation structure-fonction d'une nouvelle xylanase. Zoologie des invertébrés. Université Paul Sabatier - Toulouse III, 2016. Français. NNT : 2016TOU30366 . tel-01677278

HAL Id: tel-01677278

<https://theses.hal.science/tel-01677278>

Submitted on 8 Jan 2018

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 Fédérale



Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :

GUYEZ Barbara

le 6 décembre 2016

Titre :

Exploration du microbiote d'invertébrés par métagénomique fonctionnelle et caractérisation structure-fonction d'une nouvelle xylanase

École doctorale et discipline ou spécialité :

ED BSB : Biologie structurale et fonctionnelle

Unité de recherche :

IPBS UMR 5089 - LISBP UMR INSA/CNRS UMR 5504 INSA/INRA UMR 792

Directeur/trice(s) de Thèse :

DUMON Claire : Directrice de thèse

TRANIER Samuel : Co-directeur de thèse

Jury :

AGHAJARI Nushin : Directrice de recherche, IBCP UMS 3760

BONNIN Estelle : Ingénieure de recherche, INRA Nantes UR 1268

PAËS Gabriel : Chargé de recherche, UMR FARE-CREA

GIRBAL-NEUHAUSER Élisabeth : Professeure des universités, LBAE EA 4565

MOUREY Lionel : Directeur de recherche, IPBS UMR 5089

A ma mamie Odette

Remerciements

Tout d'abord, je voudrais remercier mes deux directeurs de Thèse Claire Dumon et Samuel Tranier pour m'avoir fait confiance et de m'avoir encadrée tout au long de ces trois années de thèse.

Je remercie les membres de mon jury d'avoir accepté de juger mon travail.

Ensuite, je voudrais remercier Lionel Mourey de m'avoir accueillie dans son équipe de Biophysique Structurale lors de mes stages de première et deuxième années de master ainsi que pour cette thèse.

Je remercie aussi Magalie Remaud-Siméon et Mike O'Donohue de m'avoir accueillie au sein de l'équipe Catalyse et Ingénierie Moléculaires et Enzymatiques et plus particulièrement dans le groupe des Lignocellulases.

Un immense merci aux membres de l'équipe de Biophysique Structurale qui m'ont accueillie à bras ouvert dès mon premier stage et qui ont tout fait pour que je me sente bien tout de suite. Merci Cécile B de m'avoir mis le pied à l'étrier, sans ces stages je ne serai peut-être pas là aujourd'hui. Merci Éric pour toute l'aide que tu m'as apporté et le savoir que tu m'as donné lors de ce premier stage. Merci aussi d'avoir continué à me suivre et à me prodiguer tes conseils jusqu'aux derniers moments de ma thèse. Merci Sylviane !!! Tu as toujours été de bon conseil, tu es toujours là pour rendre service. Merci à ma super voisine de bureau Suzana, sans toi ça n'aurait forcément pas été pareil !! Toutes nos discussions plus ou moins scientifiques ont été que du bonheur. On se sera aussi bien marrée !!! Julian, tu nous as rejoints au bureau et ton humour et ta gentillesse ont fait que nous t'avons adopté tout de suite. Virginie, tu as toujours été là dans les bons moments mais aussi et surtout dans les moins bons. Tu as été d'une grande écoute et de bons conseils. Sabine, quel bonheur de travailler à tes côtés !!! Merci pour tout ce que tu m'as appris. Je n'oublierai jamais cette faucheuse tendance que tu avais à me taper dessus ! Merci aussi à Laurent, Valérie, Cécile C, Julie, Audrey, JD, Florence, Coralie, Alex, Stéphane.... Et ceux que j'ai eu le plaisir de croiser mais le temps nous a manqué pour mieux se connaître : Monique, Karine,

Fred, Romain, Luis, Dany, Ploï, Minh Chau. Un merci tout spécial à Loïc qui a partagé le stage de M2 avec moi. On en aura eu des fous rires et je garde un bagage remplis de souvenir. J'en viens donc logiquement à mes autres potes de M2 et de stage, Stéphan, Anto et Sab ainsi que Mounette, Aïcha et Richard, cette année de master restera inoubliable.

A l'IPBS toujours, je voudrais remercier Fabrice Dumas, directeur du M2 BSF ainsi que Marie Renault et Maryelle Tropis qui m'ont encadrée pour mon atelier 2. J'ai beaucoup aimé travailler avec vous. Un merci aussi à Fabienne Bardou qui a été ma tutrice de DCCE. Et dans le désordre, je remercie Catherine Tardin, Sandrine Uttenweiller, Thomas et Barbara, Manuelle Ducoux, Moinecha, Amscha, Anthony, Anne, Yvane, Nathalie P, Amar, Carole P ...

Je voudrais aussi remercier l'ensemble de l'équipe pédagogique de Biochimie de la fac et plus particulièrement Anne et Karine sans ça serait plus compliqué de faire les TPs !

Cap maintenant sur le LISBP. Un grand merci à tous les lignos. Merci à Greg pour m'avoir données quelques infos utiles pour commencer ma thèse et pour continuer le travail que tu avais commencé. Merci Julien, j'ai été ravie de te retrouver, voisin de bureau au tout début et voisin de paillasse tout le long, après avoir fait tout notre parcours licence et master ensemble. Merci à Thomas et Pablo pour leur accueil et leur aide. Merci aussi Eleni et Haiyang pour ces moments partagés dans le labo et le bureau. Merci à Franck d'avoir été un super stagiaire ! J'ai vraiment apprécié de travailler avec toi. Merci aussi à Claire R. Victor, tu es fou et surtout ne change pas. Merci aussi à Régis, Cédric et Sophie.

Des mercis tous spéciaux à mes comparses de toujours. Louise, on a été voisines de paillasse et ce n'est pas toujours de tout repos !!! Je n'ai plus l'occasion de te dire « Arrête de râler Louise » Ces presque 3 ans à tes côtés sont inoubliables. Merci pour nos délires et nos discussions. Lisa, ... ah là là !! Tu as été la première à m'accueillir. Nos discussions et surtout nos samedis à jouer m'ont fait le plus grand bien. Marie H, on s'est très vite comprises !! J'ai beaucoup aimé tous les moments passés ensemble à discuter, à rigoler, à râler aussi ! Merci aussi à Virginie, nos discussions pas toujours scientifiques ont été géniales. Je crois qu'on aurait pu rester des heures à parler bouquins. Mawina !! Que de bons moments passés !!! Ta douceur et ta gentillesse font du bien. Angie, quel bonheur et quelle fraîcheur !!!! On s'est bien marré au bureau et c'était souvent grâce à toi. Tu as trouvé une nouvelle tirelire ?

Merci à Gianluca pour son expertise en purif et son humour, nos discussions étaient très agréables. Et merci à Nelly et Laurence de m'avoir formée sur l'HPAEC-PAC et pour leur expertise. Une grand merci à Sandrine et Delphine d'avoir aussi gentiment et rapidement répondu à toutes mes questions en bioinfo.

Merci à mes collègues de bureau, Marie G, Emilie, Coco, Popo, Hélène, Maher, Thibault, Yvonne, Marlène C, JR, pour ces bons moments. Merci pour nos craquages du vendredi aprem et pour m'avoir suivi dans mes délires chansonniers et mes hallucinations auditives !!!

Merci Nathalie, la Bretagne nous rapprochera toujours ! On a bien rigolé, je t'ai pas mal taquiné, tu l'as toujours bien pris et tu me l'as bien rendu.

Merci à 2 super filles de l'EAD15 pour les bons moments au labo et en dehors : Gwen et Anne-Laure.

Merci à tous ceux qui ont partagé les jeudis bière ! N'arrêtez pas, n'oubliez pas de continuer à relâcher de temps en temps.

Merci à tous ceux qui ont partagé mon quotidien : Marlène V, les Marions, Betty, Vinciane, les Manons, Alex, Yannick, Charles, Marc, Emilie, Laure, Clément, Maéva, Moumou, Etienne, Akli, Vincent..... Et désolée à tous ceux que j'ai oubliés.

Les mercis plus personnels maintenant. Merci à mes parents d'avoir toujours été là, d'avoir fait ce que je suis. Merci à mes grand-mères.

Merci à ma famille, oncles, tantes, cousins et cousines qui m'apportent tant et plus particulièrement à Michel et Elisabeth pour leur relecture.

Merci à ma « belle-famille ».

Merci à mes amis. Etienne F, mon presque frère, on se connaît depuis tellement longtemps ! Lucie et JB et Pierre et Ingrid, du bonheur partagé depuis de très nombreuses années et pour longtemps encore. Etienne G, nos délires sont des souvenirs inoxydables. Maïté et Jérôme, la fac de médecine ça peut faire naître de jolies amitiés. Marie-Ange, en grattant un peu on se rend compte que tu es au moins aussi folle que moi !

Enfin, le dernier merci et non des moindres pour celui qui partage ma vie depuis un peu plus de 4 belles années. Merci de m'avoir soutenu sans faille, merci de m'avoir apporté ce réconfort dont j'avais besoin. Merci de m'avoir supporté dans les moments moins faciles de la thèse. Merci d'être là et d'être toi. Merci Yann.

Résumé

La paroi végétale est une structure complexe composée principalement de polysaccharides (cellulose, hémicellulose et pectine), de lignine et de protéines. Elle est impliquée dans de nombreuses fonctions essentielles à la vie de la cellule végétale. De plus, les constituants de cette paroi, que sont les polysaccharides et la lignine, représentent la plus grande source de carbone renouvelable de la planète. Ceci en fait des cibles de choix notamment pour la production d'énergies « vertes ». Toutefois, l'utilisation des polysaccharides tels que les hémicelluloses constituant la paroi végétale reste, à l'heure actuelle, limitée du fait de la difficulté à les dégrader. Ces dernières années, un effort important a été mis en œuvre pour identifier et caractériser de nouvelles enzymes, telles que les glycosides hydrolases, permettant de dégrader efficacement la biomasse végétale.

Dans le but de découvrir de nouvelles enzymes impliquées dans la dégradation de la biomasse végétale, des chercheurs de l'équipe « Catalyse et Ingénierie Moléculaire Enzymatiques » du LISBP ont décidé d'explorer le métagénome d'organismes connus pour dégrader la biomasse végétale. Deux espèces animales ont fait l'objet d'analyses : tout d'abord les termites qui sont considérés comme les champions de la dégradation de la biomasse végétales et souvent comparés à des bioréacteurs, et le ver de terre. Des banques métagénomiques de trois espèces différentes de termites ainsi qu'une banque métagénomique de ver de terre ont ainsi été créées.

Dans ces travaux de thèse deux des banques métagénomiques de termites, celle de *Nasutitermes corniger* et celle de *Termes hispaniolae*, ont fait l'objet d'une étude afin de comparer le potentiel hémicellulolytique de ces deux espèces. Après sélection de nombreux clones positifs sur substrats chromogéniques de chacune des deux banques, séquençage puis annotation taxonomique et fonctionnelle, un grand nombre d'enzymes et principalement des glycosides hydrolases, a pu être identifié. Les résultats montrent que le métagénome de *Nasutitermes corniger* présente majoritairement des enzymes à activité endoglycosidase alors que le métagénome de *Termes hispaniolae* possède plutôt des enzymes à activité exoglycosidase. Toutes les activités trouvées dans chacune des espèces de termite sont en bonne corrélation avec l'alimentation du termite. De plus, nous avons observé que le microbiote intestinal des deux termites ne possèdent pas les mêmes

embranchements bactériens majoritaires et nous avons pu voir que le microbiote de *Termes hispaniolae* est plus diversifié ce qui corrèle aussi avec l'alimentation des deux termites.

D'autre part, dans la banque métagénomique du ver de terre, l'annotation fonctionnelle a révélé une enzyme intéressante. Il s'agit d'une enzyme annotée par B. Henrissat (responsable de la base de données CAZy) comme étant une glycoside hydrolase putative mais n'appartenant à aucune des 135 familles de glycosides hydrolases existantes. Cette enzyme putative, appelée GH* présente des similitudes avec les GH de la famille 5 sans pour autant appartenir à cette famille du fait notamment de l'absence du résidu catalytique nucléophile conservé. Une étude structurale et fonctionnelle de GH* a donc été menée. Les expériences ont permis de prouver que GH* est une endo-xylanase ayant une préférence pour les arabinoxylanes et les xylooligosaccharides de degré de polymérisation d'au moins 5 ou 6. La structure tridimensionnelle de GH* à 1,6Å de résolution a été obtenue par cristallographie des rayons X par remplacement moléculaire à l'aide d'une GH5. Cette structure a permis de confirmer l'identité du résidu acide/base identifié par alignement de séquences et d'émettre une hypothèse sur l'identité du résidu nucléophile. Enfin des mutants de GH* pour ces deux résidus ont été obtenus et ont confirmé leur implication dans l'activité de l'enzyme.

Summary

Plant cell wall is a complex structure surrounding plant cells mainly composed by polysaccharides (cellulose, hemicellulose and pectin), lignin and proteins. The plant wall maintains and imposes the size and shape of cells. It is also important for exchanges between cells and extra cellular medium. The polysaccharides of this cell wall are the largest renewable carbon source on the earth, which makes them good targets to produce green energies. Because plant cell wall is difficult to degrade, its use for biofuels for is still limited. However, some organisms are able to efficiently degrade this biomass. Exploring the diversity of the living world to discover new effective biocatalysts has grown considerably last years, because of the emergence of metagenomics.

In this context and to discover new enzymes involved in the degradation of plant biomass, the team « Catalyse et Ingénierie Moléculaire Enzymatiques » of LISBP decided to explore metagenome of organisms known to degrade plant biomass. Two animal families were chosen for metagenomics analysis, the termite and earthworm. Metagenomics banks of three different species of termite and one metagenomics bank of an earthworm were created.

In this thesis project, two of the three metagenomics banks of termites, the one from *Nasutitermes corniger* and the other one from *Termes hispaniolae*, were studied to compare the hemicellulolytic potential of these two species. After selection of many positive clones on chromogenic substrates of both banks, sequencing, taxonomic and functional annotations, a large number of enzymes and mainly glycoside hydrolases, could be identified. The results obtained shown that the trends observed during functional screens were maintained. Indeed, it appears that *Nasutitermes corniger* has a majority of endoglycosidases while *Termes hispaniolae* has mainly exoglycosidases. Thereby, families of enzymes highlighted allowed correlating their hydrolytic activities with the diet of these species. Furthermore, we observed that the intestinal microbiota of each termite is different. Indeed, both termites do not have the same majority bacterial phyla and the microbiota of *Termes hispaniolae* is more diverse than the one of *Nasutitermes corniger*.

On the other hand, functional annotation of the metagenomics bank of the earthworm revealed an enzyme annotated as a glycoside hydrolase no belonging to any of the 135

glycoside hydrolase existing families. This enzyme, named GH*, seems to be close to GH5 but does not show the nucleophilic catalyst residue perfectly conserved in this glycoside hydrolase family. A functional and structural study of GH* was then done. We have shown that GH* is an endo-xylanase which prefers arabinoxylans and xylooligosaccharides having a polymerization degree greater than 5. In addition, we determined the crystal structure of GH* at 1.6Å resolution. This 3D structure has confirmed the presence of the acid/base residue identified by sequence alignment and allowed us to hypothesize about the identity of the nucleophilic residue. Finally, mutants of GH* for these two residues were obtained and confirmed their involvement in the activity of the enzyme. We were able to progress in the understanding of structure/function relationships of this protein.

Table des Matières

Introduction générale	1
I. Préambule.....	1
II. La paroi végétale et ses constituants.....	4
A. La paroi végétale	4
B. La cellulose	6
C. Les hémicelluloses.....	7
1. Les xyloglucanes	10
2. Les xylanes	10
3. Les glucomannanes.....	11
4. Les β -(1 \rightarrow 3,1 \rightarrow 4)-glucanes.....	11
D. Les pectines	11
E. La lignine.....	13
III. Dégradation des polysaccharides de la paroi végétale	15
A. La base de données CAZy	16
B. Les glycosides hydrolases.....	16
C. Carbohydate estérases (CE)	24
D. Enzymes à activité auxiliaire (AA).....	26
E. Modules de liaison aux glucides (CBM).....	28
F. Exemple de la dégradation totale d'un arabinoxylane	30
1. Les endo- β -1,4-xylanases	30
2. Les β -xylosidases.....	33
3. Les α -L-arabinofuranosidases	35
4. Les α -glucuronidases	37
5. Les carbohydate estérases (CEs):	40
IV. Les systèmes de dégradation de la biomasse végétale, des réservoirs naturels d'enzymes.....	42
A. Le microbiote intestinal humain	42
B. Le rumen bovin.....	44
C. Les champignons	46
D. Les vers de terre	47

E.	Les termites	50
V.	L'exploration de la diversité.....	51
A.	La métagénomique fonctionnelle	51
B.	La métagénomique.....	53
C.	La métagénomique 16S.....	53
D.	Le RNA seq ou « transcriptome sequencing »	54
E.	Comparaison de ces méthodes	55
VI.	Projet de thèse.....	56
I.	Annotation des banques métagénomiques de termites	59
A.	Assemblages puis sélection des contigs.....	59
B.	Recherche des cadres ouverts de lecture	59
C.	Annotation fonctionnelle et taxonomique	59
II.	Etude structurale et fonctionnelle de GH*	60
A.	Sous clonage.....	60
B.	Recherche des meilleures conditions pour l'expression de GH*	63
C.	Purification	64
1.	Chromatographie d'affinité sur ions métalliques immobilisés (IMAC)	64
2.	Chromatographie d'exclusion-diffusion	65
D.	Fluorimétrie différentielle à balayage.....	66
E.	Recherche de l'activité de GH*	69
1.	Activité sur pNP-substrats	69
2.	Activité de phosphoryse sur oligosaccharides	70
3.	Activité sur polysaccharides	71
F.	Chromatographie échangeuse d'anion à haute performance couplée à un système de détection ampérométrique pulsée (HPAEC-PAD)	72
G.	Détermination des paramètres optimaux.....	73
H.	Cristallisation	74
I.	Congélation des cristaux et collecte de données.....	76
J.	Détermination de la structure de GH*	76
I.	Introduction	79
II.	Résultats de l'annotation fonctionnelle.....	83
III.	Comparaison des familles de GH retrouvées dans les deux banques	84

A.	Familles retrouvées dans les deux banques	84
B.	Familles présentes dans une seule banque	87
1.	Familles identifiées majoritairement chez <i>N. corniger</i> (termite xylophage) (tableau 8).....	87
2.	Familles identifiées majoritairement chez <i>T. hispaniolae</i> (termite humivore et xylophage) (tableau 8)	88
C.	Liens avec le mode de vie des termites	90
D.	Comparaison avec d'autres espèces	91
IV.	Comparaison des autres enzymes participant à la dégradation des polysaccharides	93
A.	Les estérases	93
B.	Les polysaccharides lyases	94
C.	Domaines de liaison aux glucides (CBM).....	95
V.	Comparaison des microbiotes	103
VI.	« Polysaccharide utilization locus » (PUL).....	105
VII.	Conclusion – Discussion	108
I.	Introduction :	111
II.	Détermination des conditions optimales pour l'expression de GH*	115
A.	Sous-clonage	115
B.	Expression	116
III.	Optimisation des conditions de purification	118
A.	Chromatographie d'affinité sur ions métalliques immobilisés.....	118
B.	Chromatographie d'exclusion-diffusion.....	121
IV.	Stabilité de GH*	124
V.	Détermination de l'activité de GH*	125
A.	Criblage des pNP-glycosides.....	126
B.	Recherche d'activité de phosphorolyse sur oligosaccharides	128
C.	Criblage sur polysaccharides	128
VI.	Conditions retenues pour l'étude fonctionnelle de GH*	131
A.	Détermination du pH optimal	131
B.	Détermination de la température optimale	132
VII.	Analyse des produits d'hydrolyse du xylane de hêtre et des arabinoxylanes de blé (WAX) et de seigle (RAX).....	133
VIII.	Analyse des produits d'hydrolyse de xylooligosaccharides par GH*	136

IX. Conclusion.....	138
I. Introduction	143
II. Rappels sur GH*	143
A. Connaissances générales.....	143
B. Activité.....	144
III. Conditions de cristallisation.....	144
IV. Résolution de la structure.....	148
A. Données de diffraction.....	148
B. Obtention de la structure.....	149
V. Analyse de la structure de GH*	152
VI. Mécanisme catalytique	161
VII. Conclusions et perspectives.....	163
Conclusion générale.....	170

Liste des abréviations

AA : Auxiliary Activity

ADN : Acide DéoxyriboNucléique

ADNc : Acide DéoxyriboNucléique complémentaire

ADNr : Acide DéoxyriboNucléique ribosomique

ARN : Acide RiboNucléique

ARNm : Acide RiboNucléique messenger

ARNr : Acide RiboNucléique ribosomique

AZCL : AZurine CrossLinked

AZO : abréviation du groupement fonctionnel R-N=N-R'

BLAST : Basic Local Alignement Search Tool

BET : Bromure d'Ethidium

CAZy : Carbohyrate Active enZymes

CBM : Carbohyrate Binding Module

CCM : Chromatographie sur Couche Mince

CE : Carbohyrate Estérase

CESA : CELLulose SynthAse

CIMES : équipe Catalyse et Ingénierie Moléculaire Enzymatiques

CSL : Cellulose Synthase Like

COGs : Clusters of Orthologous Groups of proteins

DNS : acide DiNitroSalicylique

DP : Degré de Polymérisation

DSF : Fluorimétrie Différentielle à Balayage (Differential Scanning Fluorimetry)

ESTHER : ESTerases and alpha/beta-Hydrolase Enzymes and Relatives

GH : Glycoside Hydrolase

GT : Glycoside Transférase

HPAEC-PAD : Chromatographie échangeuse d'anion à haute performance couplée à un système de détection ampérométrique pulsée (High Performance Anion Exchange Chromatography – Pulsed Amperometric Detection)

ICEO : Ingénierie et Criblage d'Enzymes Originales

IMAC : Immobilized Metal Ion Affinity Chromatography

IPTG : IsoPropyl β -D-1-ThioGalactopyranoside

Kana : Kanamycine

Kb : kilo base

kDa : kilo Dalton

LB : milieu Luria-Bertani

LPMO : Lytic Polysaccharide Mono-Oxygénases

NR : Non Réducteur

ORF : Open Reading Frame

PDB : Protein Data Bank

PEG : PolyEthylène Glycol

PL : Polysaccharide Lyase

PUL : Polysaccharide Utilisation Locus

pNP : para-NitroPhényl

qPCR : quantitative Polymerase Chain Reaction

R : Réducteur

RAX : Arabinoxylane de seigle (Rye ArabinoXylan)

RMN : Résonance Magnétique Nucléaire

rpm : rotations par minute

SAD : Single Anomalous Diffraction

SDS-PAGE : Sodium DodécylSulfate – PolyAcrylamide Gel Electrophoresis

Sus : Strach utilization system

Tm : température de fusion (melting Temperature)

WAX : Arabinoxylane de blé (Wheat ArabinoXylan)

X1 : Xylose

X2 : Xylobiose

X3 : Xylotriose

X4 : Xylotétraose

X5 : Xylopentaose

X6 : Xylohexaose

Introduction générale

I. Préambule

Les glucides sont des composés organiques possédant un groupement carbonyle (cétone ou aldéhyde) et au moins deux groupements hydroxyles selon la définition de l'Union Internationale de Chimie Pure et Appliquée (IUPAC). Lorsque ces composés sont modifiés par des réductions, des oxydations ou par l'ajout de groupements fonctionnels, l'IUPAC conserve le terme de « glucide ». Toujours selon l'IUPAC, un oligosaccharide se compose de deux à dix unités glucidiques liées par des liaisons osidiques et au-delà de dix unités glucidiques on parle de polysaccharide. Un polysaccharide peut avoir une structure linéaire ou ramifiée. Un polysaccharide ramifié possède une chaîne principale linéaire et sur certains résidus monosaccharidiques de celle-ci sont branchés des oligosaccharides allant d'une à plusieurs unités.

Les glycoconjugués sont des molécules composées d'une partie glucidique liée de façon covalente à une partie non glucidique qui est soit protéique (glycoprotéine et protéoglycane) soit lipidique (glycolipide) (source IUPAC) et jouent de nombreux rôles biologiques (Varki 1993).

Dans les cellules végétales ainsi que chez les bactéries, les polysaccharides et les glycoconjugués sont présents dans la paroi cellulaire et ont une importance majeure dans le maintien de la forme des cellules (Schleifer et Kandler 1972 ; Somerville et al. 2004 ; Cosgrove 2005 ; Vollmer, Blanot, et Pedro 2008).

Les glycoconjugués et les polysaccharides présents à la surface de nombreuses cellules sont les cibles de la reconnaissance hôte/pathogène. Par exemple, des bactéries pathogènes comme *Helicobacter pylori* se fixent sur la partie glucidique de glycoconjugués présents à la surface des cellules cibles (Varki 1993 ; Finlay et Falkow 1997). De la même manière, les virus se lient à des protéoglycans et glucosaminoglycans des cellules hôtes par l'intermédiaire de glycoprotéines présentes à leur surface (Varki 1993 ; Barth et al. 2003).

Un autre exemple illustrant l'importance du rôle biologique des glycoconjugués est constitué par les glycoprotéines de surface des globules rouges définissant les antigènes des groupes

sanguins (système ABO). La partie déterminant le groupe sanguin est un noyau tétrasaccharidique L-fucosyl- α -1,2-D-galactosyl- β -1,4-N-acétylglucosamine- β -1,4-D-galactose. Enfin sur le D-galactose terminal, il y a :

- Une N-acétylgalactosamine pour le groupe A,
- Un D-galactose pour le groupe B,
- Rien pour le groupe O.

Pour le groupe AB on retrouve à la fois les antigènes des groupes A et B (Subramanian et al. 2012 ; Hara et al. 2013).

Les glycosides hydrolases (GH) sont des enzymes produites par tous les organismes vivants, car elles catalysent l'hydrolyse des liaisons O-, N- ou S-glycosidiques entre deux sucres ou entre un sucre et son glycoconjugué.

Les GH sont, par exemple, sécrétées par des pathogènes de plantes afin de dégrader les polysaccharides, qui sont les principaux constituants des parois végétales, et ainsi pouvoir infecter les plantes (Vorwerk, Somerville, et Somerville 2004).

Les GH sont aussi produites par de nombreux organismes et notamment les animaux pour la défense antibactérienne (lysozyme). Chez l'homme on retrouve du lysozyme dans de nombreuses sécrétions (salive, mucus...) ainsi que dans certaines cellules de l'immunité (macrophages et polynucléaires neutrophiles) (Yoshimura, Toibana, et Nakahama 1988 ; Peters et al. 1989). En effet, le lysozyme est capable d'hydrolyser la liaison entre l'acide N-acétyl-muramique et la N-acétyl-glucosamine qui sont des constituants du peptidoglycane des parois bactériennes (Höltje 1996).

Les GH sont également largement utilisées dans de nombreux secteurs d'activité et principalement les industries agroalimentaires, papetières et les détergents.

En œnologie, il est nécessaire de libérer les composés aromatiques contenus dans les parois végétales. Ces parois sont principalement composées de polysaccharides (cellulose, hémicellulose et pectines) et l'utilisation d'enzymes capables de les dégrader va permettre la déconstruction de la paroi et la libération des composés d'intérêt. La dégradation des polysaccharides des parois végétales en composés de plus petite taille va fluidifier le jus de raisin et ainsi permettre une meilleure filtration (Capounova et Drdak 2002 ; Guérin et al. 2009 ; Rio Segade et al. 2015). D'autre part, la déconstruction des polysaccharides des parois

végétales permet d'éviter la formation d'amas permettant donc une clarification plus rapide du jus (Capounova et Drdak 2002 ; Guérin et al. 2009).

Le processus de fabrication de la bière, connu depuis des millénaires, ne nécessite que de l'eau, des céréales et des levures. Le polysaccharide de réserve des céréales, l'amidon, est un homopolymère de glucose et on utilise des α -amylases pour l'hydrolyser en glucose qui est ensuite fermenté par les levures (Kirk, Borchert, et Fuglsang 2002). D'autres GH sont utilisées lors de la fabrication de la bière pour améliorer la qualité du moult et aider à la filtration en réduisant la viscosité de la même manière que pour le vin (Kirk, Borchert, et Fuglsang 2002).

Les jus de fruits et de légumes sont aussi traités avec des GH. L'utilisation de pectinases permet de dégrader les pectines responsables de la turbidité des jus, ce qui rend possible la clarification des jus (Charnock et McCleary 2005 ; Sieiro et al. 2012).

En boulangerie, des α -amylases et xylanases sont utilisées car elles permettent une meilleure croissance de la pâte ainsi qu'une meilleure structuration de la mie de pain (Charnock et McCleary 2005 ; Whitehurst et van Oort 2010).

Une des industries qui utilisent le plus de GH est celle des lessives et détergents. Des GH sont ajoutées aux détergents en tant qu'additifs. Elles permettent de décrocher ou dégrader les tâches à basse température ou de diminuer les quantités de détergents chimiques (Bisgaard-Frantzen et al. 1999 ; Kirk, Borchert, et Fuglsang 2002 ; Charnock et McCleary 2005).

L'industrie papetière utilise des xylanases dans les processus de blanchiment du papier comme prétraitement. En effet la lignine, responsable de la couleur brune de pâte kraft, est liée aux hémicelluloses et notamment aux xylanes dans les parois végétales. En dégradant les xylanes, l'extraction des lignines est rendue plus facile, ce qui permet de diminuer les traitements à base de chlore (Paice, Bernier, et Jurasek 1988 ; Ragauskas, Poll et Cesternino 1993 ; Charnock et McCleary 2005).

Une autre application industrielle, la production de biocarburant de deuxième génération, nécessite notamment des cellulases pour produire du bioéthanol à partir de cellulose (Sun et Cheng 2002).

Comme nous avons pu le voir, les glycosides hydrolases sont très largement utilisées dans notre environnement quotidien.

La Découverte de nouvelles GH et la meilleure compréhension de leur fonctionnement permettra d'optimiser leur utilisation.

Dans le cadre de ma thèse, je me suis intéressée à l'action des GH conduisant à la déconstruction de la biomasse végétale dans le but d'approfondir notre compréhension des mécanismes mis en jeu et de les exploiter afin d'optimiser la valorisation de la biomasse végétale, notamment pour la production de biocarburants.

II. La paroi végétale et ses constituants

A. La paroi végétale

La paroi qui entoure la membrane plasmique des cellules végétales permet à la cellule de croître et d'avoir une certaine plasticité grâce à sa structure et sa complexité, détaillée plus bas dans ce chapitre (Chundawat et al. 2011). La paroi constitue également une structure rigide autour des cellules ce qui permet le maintien de leur forme. Cette structure rigide est possible grâce à la pression osmotique qui crée une force de tension sur la paroi (Somerville et al. 2004 ; Cosgrove 2005). La paroi végétale permet aussi de résister aux infections par des agents pathogènes. En effet, en réponse à des pathogènes, la cellule végétale produit des polysaccharides résistants aux attaques enzymatiques tels que la callose, qui est un β -1,3-glucane (Aist 1976 ; Vorwerk, Somerville, et Somerville 2004 ; Underwood 2012). Les cellules végétales produisent aussi des protéines, comme des chitinases, qui se retrouveront dans la paroi végétale et pourront dégrader la chitine présente dans les parois des champignons et ainsi opposer une résistance à ceux-ci (Vorwerk, Somerville, et Somerville 2004). La paroi des jeunes cellules doit aussi être poreuse afin de laisser passer l'eau, les nutriments et les hormones nécessaires à la vie des cellules (Burton, Gidley, et Fincher 2010).

Cette paroi est essentiellement composée de polysaccharides (pour plus de 80%), de protéines glycosylées et de lignines (10 à 25%) (figure 1). Elle se divise en deux grandes parties distinctes : la paroi primaire d'une part et la paroi secondaire d'autre part. La paroi primaire est la première à être synthétisée et est composée de cellulose, d'hémicelluloses et de pectines (figure 1A) (Cosgrove 2005). La paroi secondaire est synthétisée dans un deuxième temps et vient s'intercaler entre la membrane plasmique et la paroi primaire. Elle

se compose de cellulose, d'hémicelluloses et de lignines et est plus épaisse et rigide que la paroi primaire (Somerville et al. 2004). La paroi secondaire se divise en trois parties appelées S1, S2 et S3. La partie S1 est la plus externe située juste sous la paroi primaire et se distingue des deux autres par la présence d'hélices de microfibrilles de cellulose. La partie S2 est la partie centrale et la partie S3 est la plus interne. Ces deux parties ne possèdent pas d'hélices de microfibrilles de cellulose (figure 1B) (Pérez et al. 2002).

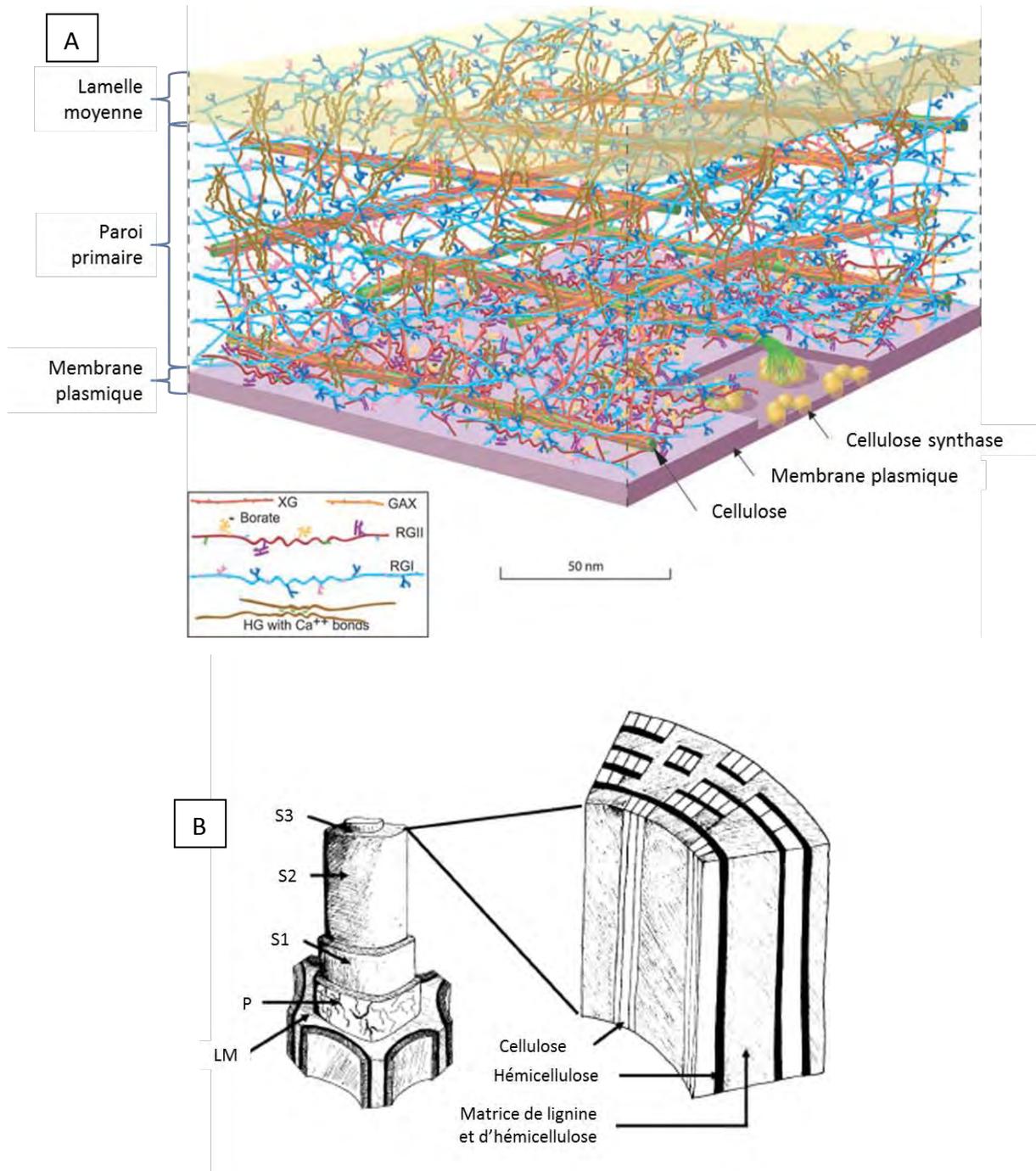


Figure 1 : Représentation schématique de la paroi primaire (A) et détail de la structure de la paroi (B). Lamelle moyenne (LM), paroi primaire (P) et paroi secondaire (S1, S2 et S3) (Somerville et al. 2004) (Pérez et al. 2002).

Les polysaccharides sont donc les constituants principaux de la paroi et représentent la plus grande source renouvelable de carbone de la planète. Parmi ceux-ci, la cellulose est le polysaccharide majeur de la paroi végétale (40 à 60 %) devant les polysaccharides dits « non-cellulosiques » que sont les hémicelluloses (20 à 40 %) et les pectines (environ 35 %). Les protéines présentes dans la paroi sont, pour certaines, impliquées dans le remodelage de celle-ci (Somerville et al. 2004 ; Cosgrove 2005). Autour de la paroi se trouve la lamelle moyenne, riche en pectines et impliquée dans l'adhésion et les jonctions cellulaires (Mohnen 2008).

B. La cellulose

La cellulose est le polymère le plus abondant sur Terre. Sa structure et sa biosynthèse sont bien caractérisées. Les microfibrilles de cellulose sont synthétisées dans la membrane plasmique parallèlement aux microtubules du cytosquelette ce qui permet à la paroi de s'adapter à la croissance des cellules végétales notamment en se réorganisant. Cette réorganisation est rendue possible par la présence de protéines dans la paroi végétale (Bashline et al. 2014). En effet, l'orientation de la cellulose est dépendante de celle des microtubules car les protéines synthétisant la cellulose sont liées aux microtubules et se déplacent le long de ceux-ci (Crowell et al. 2011 ; Chan et al. 2011). Les protéines qui fabriquent les microfibrilles de cellulose sont des celluloses synthases (CESA) et sont organisées en hexamères puis en hexamères d'hexamères, appelés rosettes, dans la membrane plasmique (figure 1A) (Somerville et al. 2004 ; Cosgrove 2005 ; Bashline et al. 2014). La cellulose est un polysaccharide linéaire composé d'unités D-glucopyranosyle liées en β -1,4 (figure 2A). La structure secondaire de la cellulose peut s'assimiler à une hélice à 2 résidus par tour, avec un pas de 10,4 Å et une translation axiale de 5,2 Å par résidu. La structure tertiaire de la cellulose est constituée d'une association parallèle des chaînes polysaccharidiques en microfibrilles de 3 à 5 nm de diamètre (Cosgrove 2005). Ces microfibrilles sont réparties tout autour de la cellule dans la paroi et sont extrêmement résistantes aux attaques mécaniques et enzymatiques en raison de leur grande stabilité due

à leur organisation cristalline (Quiroz-Castañeda et Folch-Mallol 2013) (Várnai et al. 2014). La cellulose est majoritairement sous forme cristalline dans laquelle viennent s'insérer des régions dites amorphes (5 à 20 %) (figure 2B). Les régions cristallines sont caractérisées par une structuration très organisée des microfibrilles maintenues par de nombreuses liaisons hydrogène intra et intermoléculaires ainsi que par des interactions de van der Waals (Béguin et Lemaire 1996 ; Quiroz-Castañeda et Folch-Mallol 2013). La perte de ces interactions conduit à la formation de régions amorphes moins structurées (Quiroz-Castañeda et Folch-Mallol 2013).

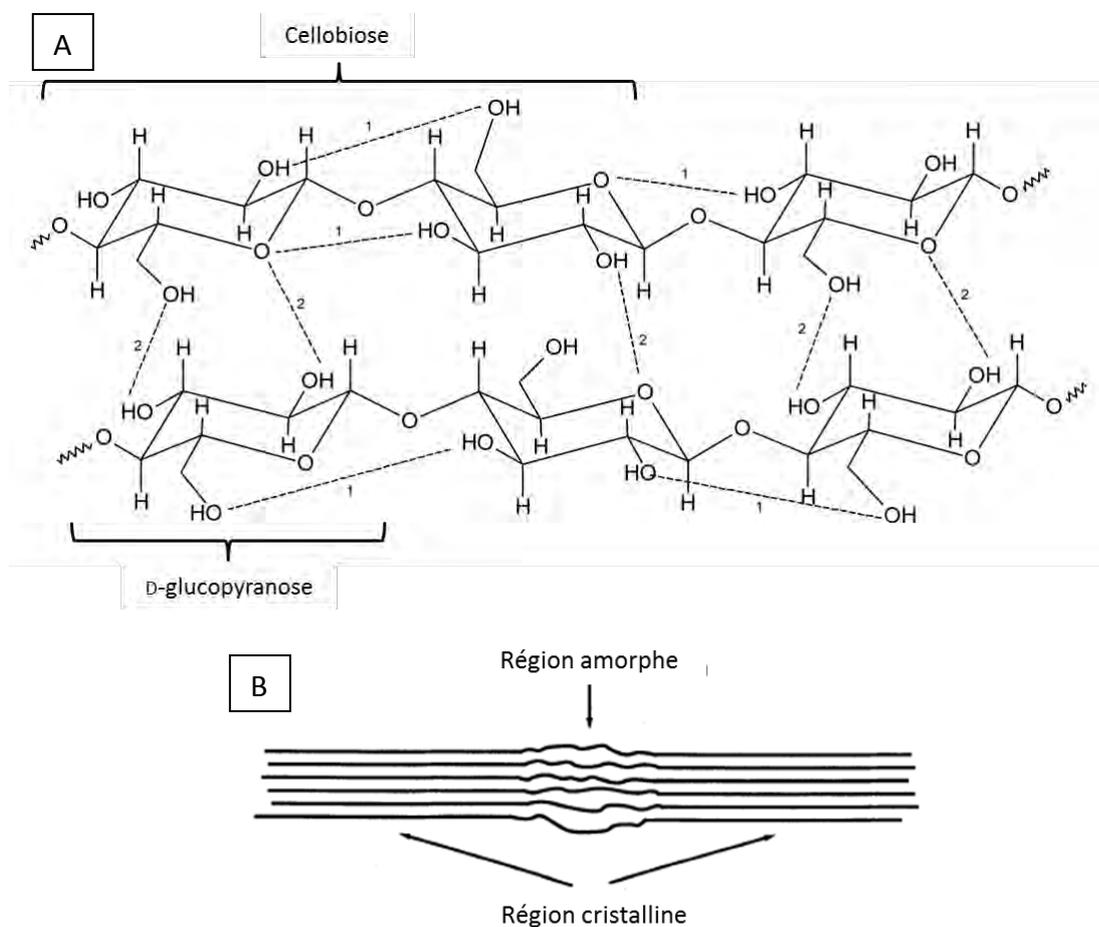
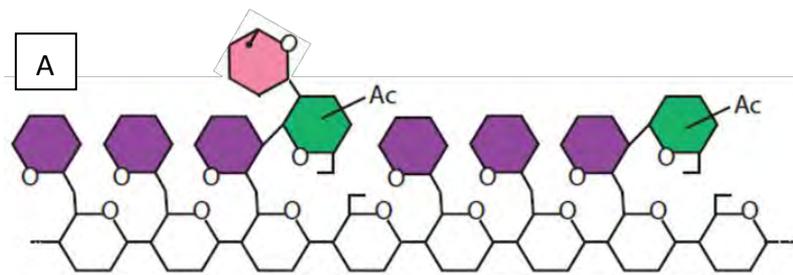


Figure 2 : Représentation détaillée de la cellulose (A). Liaisons hydrogène intra- (1) et inter-chaînes (2) permettant de maintenir l'ensemble de la structure. **Représentation schématique de la cellulose (B)** (Béguin et Lemaire 1996).

C. Les hémicelluloses

Les hémicelluloses sont un groupe hétérogène de polysaccharides ramifiés. En effet, elles présentent différents squelettes polysaccharidiques : glucanes, xylanes, mannes,

glucomannanes et β -(1 \rightarrow 3,1 \rightarrow 4)-glucanes. Des unités osidiques liées en β -1,4 et une position équatoriale des hydroxyles portés par les carbones en positions C1 et C4 (Scheller et Ulvskov 2010) font partie des caractéristiques communes aux différentes structures qui constituent les hémicelluloses. Les ramifications peuvent être constituées de différentes unités osidiques : du glucose, du galactose, de l'arabinose, du mannose, du xylose, du rhamnose, des acides glucuroniques... Ces polysaccharides peuvent aussi être estérifiés, avec formation de groupements acétyl-esters et féruloyl-esters, sur la chaîne principale ou sur les ramifications. Ces polysaccharides sont composés de monosaccharides neutres permettant leur liaison à la cellulose par l'intermédiaire d'un réseau de liaisons hydrogène, les chaînes d'hémicelluloses s'enroulant autour des fibres de cellulose (Cosgrove 2005 ; Bashline et al. 2014). Les hémicelluloses peuvent aussi se lier de façon covalente aux pectines (Cosgrove 2005 ; Bashline et al. 2014). Cette association permet la formation d'un réseau de polysaccharides dans la paroi et de diminuer la résistance mécanique de celle-ci en empêchant l'agrégation de la cellulose, facilitant ainsi l'expansion de la paroi (Somerville et al. 2004). Les hémicelluloses sont synthétisées dans l'appareil de Golgi par des enzymes dénommées « celluloses synthases like » (CSL) (Driouich, Faye, et Staehelin 1993) et sont ensuite sécrétées par le biais d'un trafic de vésicules allant de l'appareil de Golgi à la membrane (Driouich et al. 2012 ; Bashline et al. 2014). Les hémicelluloses les plus communes sont le xyloglucane, les glucurono(arabino)xylanes et les glucomannanes dans des proportions différentes en fonction du type de plante ainsi que de l'organe (tige, graine, fruit ...) (Somerville et al. 2004 ; Cosgrove 2005 ; Scheller et Ulvskov 2010 ; Bashline et al. 2014). Les hémicelluloses sont moins bien caractérisées et moins valorisées que la cellulose. Les hémicelluloses étant plus complexes que la cellulose, elles nécessitent un plus grand nombre et une plus grande variété d'enzymes pour totalement les hydrolyser (Cosgrove 2005).



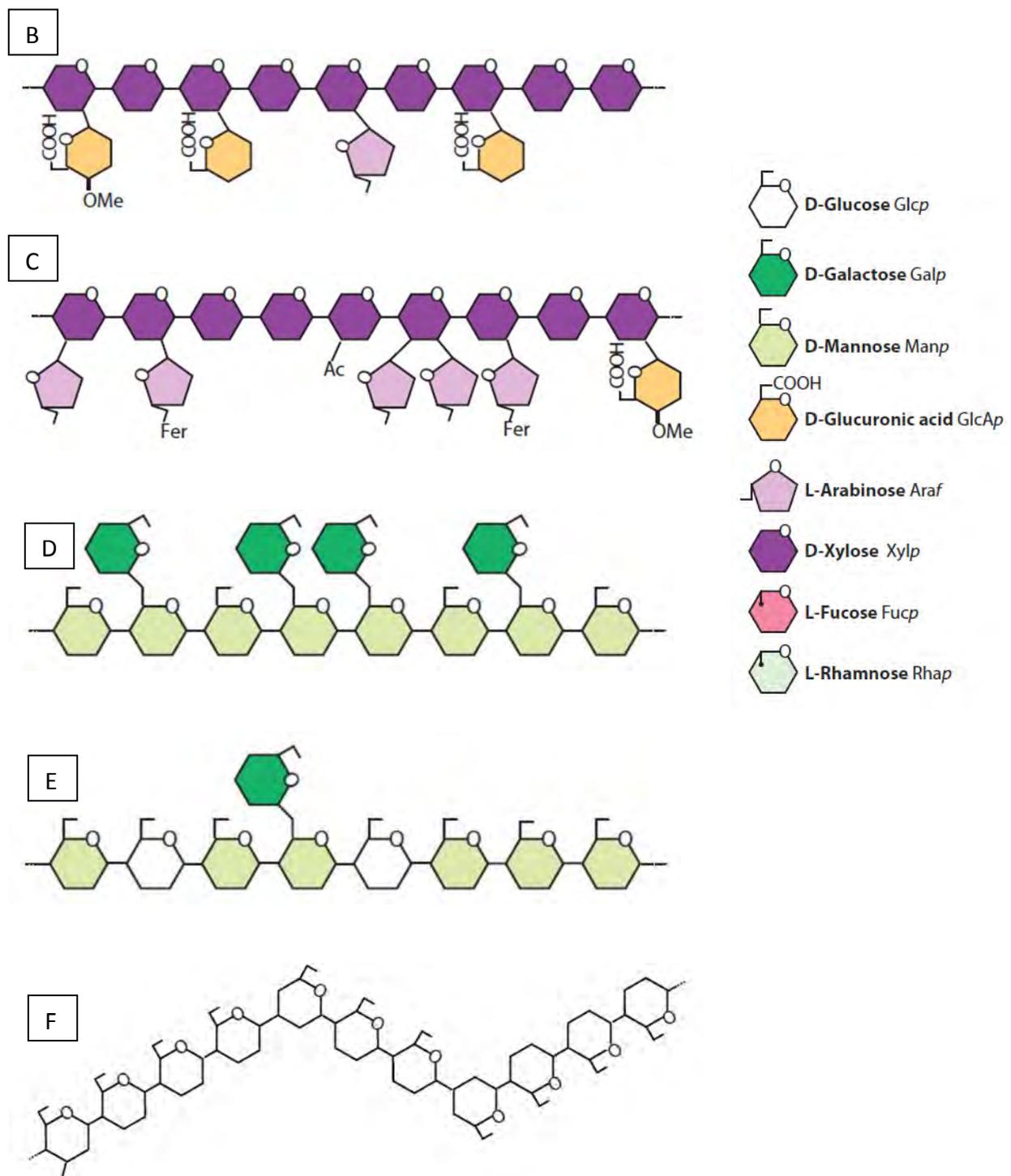


Figure 3 : Représentation schématique des différentes hémicelluloses retrouvées dans les parois végétales. Xyloglucane (A), glucuronoxylane (B), glucuronoarabinoxylane (C), galactomannane (D), galactoglucomannane (E) et β -(1 \rightarrow 3,1 \rightarrow 4)-glucane (F). Les estérifications des chaînes principales et des ramifications sont représentées comme suit : acétylations (Ac), féruloylations (Fer), O-méthylation (OMe) (Scheller et Ulvskov 2010).

Quantité de polysaccharides dans la paroi (% de matière sèche)						
Polysaccharide	Angiospermes		Graminées		Conifères	
	Paroi primaire	Paroi secondaire	Paroi primaire	Paroi secondaire	Paroi primaire	Paroi secondaire
Xyloglucane	20 – 25	Mineur	2 – 5	Mineur	10	-
Glucuronoxylane	-	20 – 30	-	-	-	-
Glucuronoarabinoxylane	5	-	20 – 40	40 – 50	2	5 – 15
(Gluco)Mannane	3 – 5	2 – 5	2	0 – 5	-	-
Galactoglucomannane	-	0 – 3	-	-	+	10 – 30
β -(1→3,1→4)-glucane	Absent	Absent	2 – 15	Mineure	Absent	Absent

Tableau 1 : Pourcentages (en masse) des différentes hémicelluloses dans les parois primaires et secondaires des plantes. Quantités mineures ou nulles (-) et quantités existantes mais avec des données non répertoriées (+) (d'après Scheller et Ulvskov 2010).

1. Les xyloglucanes

Les xyloglucanes sont retrouvés dans toutes les espèces de plantes terrestres. Ce sont les hémicelluloses les plus abondantes (20 à 25%) dans la paroi primaire des angiospermes (tableau 1). La chaîne principale est composée de D-glucopyranoses liés en β -1,4. Les ramifications sont majoritairement constituées de résidus D-xylopyranosides liés en α -1,6. Ces résidus peuvent eux-mêmes être ramifiés par des β -D-galactopyranosides ou des L-arabinofuranosides liés en α -1,2. Les résidus β -D-galactopyranosides peuvent être ramifiés par des L-fucopyranosides liés en α -1,2. De plus, les résidus β -D-glucopyranoses peut être acétylés (figure 3A) (Scheller et Ulvskov 2010).

2. Les xylanes

Les xylanes sont répartis en sous-groupes possédant tous une chaîne principale de D-xylopyranoses liés en β -1,4. Ces hémicelluloses sont retrouvées dans les parois primaires et/ou secondaires de tous les végétaux (tableau 1) (Kulkarni, Shendye, et Rao 1999 ; Scheller et Ulvskov 2010). Les différents xylanes possèdent des ramifications de résidus d'acides glucuroniques liés en α -1,2- pouvant être méthylés en position 4 (Scheller et Ulvskov 2010). Les xylanes possédant majoritairement ce type de ramifications sont appelés

glucuronoxylanes et sont les hémicelluloses les plus représentées dans les parois secondaires des angiospermes (figure 3B) (Scheller et Ulvskov 2010). Les autres unités osidiques composant les ramifications des xylanes sont des résidus L-arabinofuranosides liés en α -1,2 et/ou α -1,3. Ces xylanes sont alors appelés (glucurono)arabinoxylanes selon la proportion des 2 résidus branchés (figure 3C). Certains arabinoxylanes ne possèdent pas de résidus acides glucuroniques dans leurs ramifications. Le motif de branchement des décorations (notamment les résidus L-arabinofuranosides) est fonction du type de plante (Scheller et Ulvskov 2010). Ces polysaccharides sont retrouvés en grande quantité dans les parois primaires et secondaires des graminées (20 à 40 % et 40 à 50 %, respectivement) et sont aussi présents dans les parois secondaires des conifères (5 à 15 %) (tableau 1). Les xylanes sont estérifiés à différents degrés, selon le type de plante, principalement sur la position 3 des D-xylopyranoses de la chaîne principale. Les résidus L-arabinofuranosyls peuvent aussi être féruloylés sur leur carbone en position 5 (Scheller et Ulvskov 2010).

3. Les glucomannanes

Les (gluco)mannanes ont une chaîne principale composée uniquement de D-mannopyranoses pour les mannanes (figure 3D) ou d'une alternance de D-mannopyranoses et de D-glucopyranoses pour les glucomannanes (figure 3E) liés en β -1,4. Ces 2 types de polysaccharides sont ramifiés par des résidus D-galactopyranosides liés en α -1,6 sur les D-mannopyranoses. Les (gluco)mannanes sont souvent acétylés et sont majoritairement retrouvés dans les parois secondaires (10 à 30 % chez les conifères) (tableau 1). Ce sont les hémicelluloses principales des algues vertes (Scheller et Ulvskov 2010).

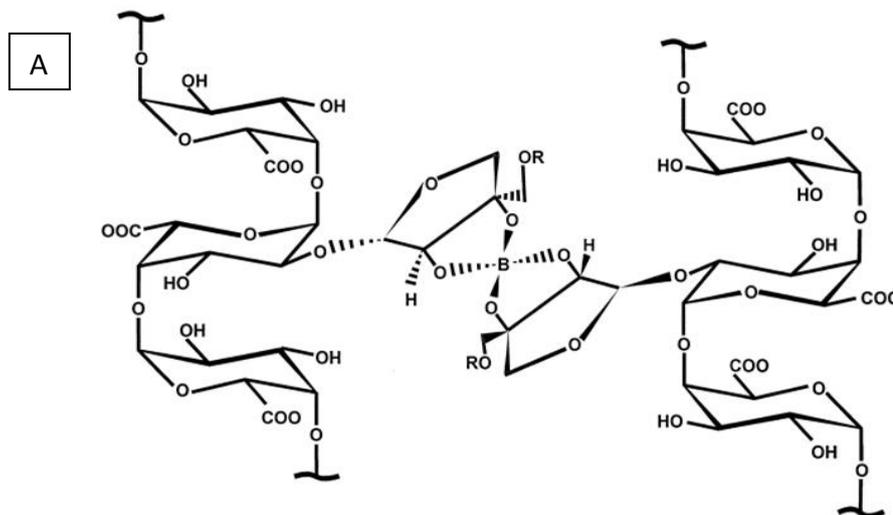
4. Les β -(1→3,1→4)-glucanes

Les β -(1→3,1→4)-glucanes sont des polymères de D-glucopyranoses liés en β -1,4 et en β -1,3 intercalés de façon régulière tous les 3 ou 4 D-glucopyranoses (figure 3F). Ces polysaccharides sont principalement retrouvés dans les parois primaires des graminées (2 à 15 %) (tableau 1) (Scheller et Ulvskov 2010).

D. Les pectines

Les pectines sont les polysaccharides les plus complexes et hétérogènes (Cosgrove 2005 ; Bashline et al. 2014). Les pectines sont une famille de polysaccharides riches en acides

galacturoniques liés de manière covalente sur les positions O-1 et O-4 des unités osidiques (Willats et al. 2001 ; Mohnen 2008 ; Voragen et al. 2009). Les principales pectines sont l'homogalacturonane (le plus abondant), le rhamnogalacturonane I, le rhamnogalacturonane II, l'apiogalacturonane et le xylogalacturonane. L'homogalacturonane est le seul de ces polysaccharides à être linéaire, les autres sont ramifiés par différentes unités osidiques comme présenté en figure 4. Le rhamnogalacturonane II peut s'organiser en dimères par l'intermédiaire de diesters de borate entre deux résidus apiosyl (figure 4A) ou en interagissant avec des ions calcium (Willats et al. 2001 ; Ridley, O'Neill, et Mohnen 2001 ; (Voragen et al. 2009). Les rhamnogalacturonanes I et II sont liés de façon covalente par des liaisons glycosidiques avec les autres pectines et notamment l'homogalacturonane ainsi qu'à d'autres polysaccharides de la paroi (Willats et al. 2001 ; Voragen et al. 2009). Tous ces polysaccharides sont constitués de domaines distincts liés entre eux de façon covalente (figure 4B). Les pectines sont estérifiées par des groupements acétyl (Willats et al. 2001 ; Mohnen 2008 ; Voragen et al. 2009). Comme les hémicelluloses, les pectines sont synthétisées dans l'appareil de Golgi par des glycosyltransférases membranaires puis sécrétées dans la paroi par l'intermédiaire de vésicules allant de l'appareil de Golgi à la membrane (Driouich, Faye, et Staehelin 1993 ; Mohnen 2008 ; Driouich et al. 2012).



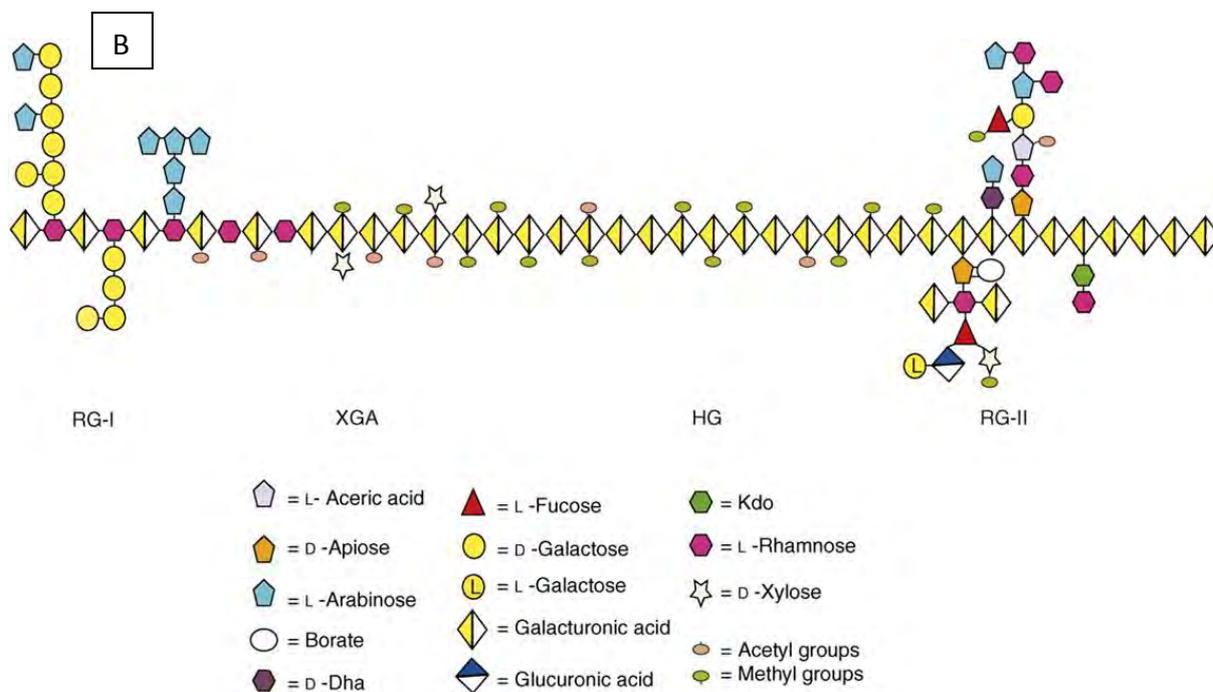


Figure 4 : A. Représentation d'un diester de borate entre 2 résidus apiosyl. La lettre R représente un oligosaccharide lié au résidu apiosyl (Ridley, O'Neill, et Mohnen 2001). **B. Représentation schématique des pectines : rhamnogalacturonane I et II (RG I et RG II), xylogalacturonane (XGA) et homogalacturonane (HG).** Acide 3-deoxyD-lyxo-heptulosarique (Dha) et acide 3-deoxy-D-manno-octolosonique (Kdo) (Mohnen 2008).

E. La lignine

La lignine est le biopolymère le plus abondant sur Terre après la cellulose et la chitine (Somerville et al. 2004 ; Cosgrove 2005). Elle joue un rôle crucial dans l'intégrité structurale de la paroi ainsi que dans la raideur et la force des tiges des plantes. Ce polymère joue aussi un rôle dans l'étanchéité et la protection face aux pathogènes des cellules végétales (Cosgrove 2005). La lignine est composée d'hétéropolymères complexes, aromatiques et racémiques, dérivés principalement de monomères d'alcools hydroxycinnamiques, appelés monolignols : alcool *p*-coumarylique (alcool 4-hydroxycinnamique), alcool coniférylique (alcool 4-hydroxy-3-méthoxy-cinnamique) et alcool sinapylique (alcool 4-hydroxy-3,5-diméthoxy-cinnamique) (figure 5) (Ishii 1997). La composition en ces différents monomères est dépendante de l'espèce, du type cellulaire et de l'environnement. La lignification se fait par réaction radicalaire à partir d'un dimère de monolignol (Boerjan, Ralph, et Baucher 2003). Les monolignols sont synthétisés dans le cytosol puis transportés jusqu'à la paroi où la lignification a lieu (Ishii 1997 ; Boerjan, Ralph, et Baucher 2003 ; Neutelings 2011). La

lignine génère une grande résistance de la paroi végétale et constitue la partie récalcitrante de la paroi végétale après hydrolyse.

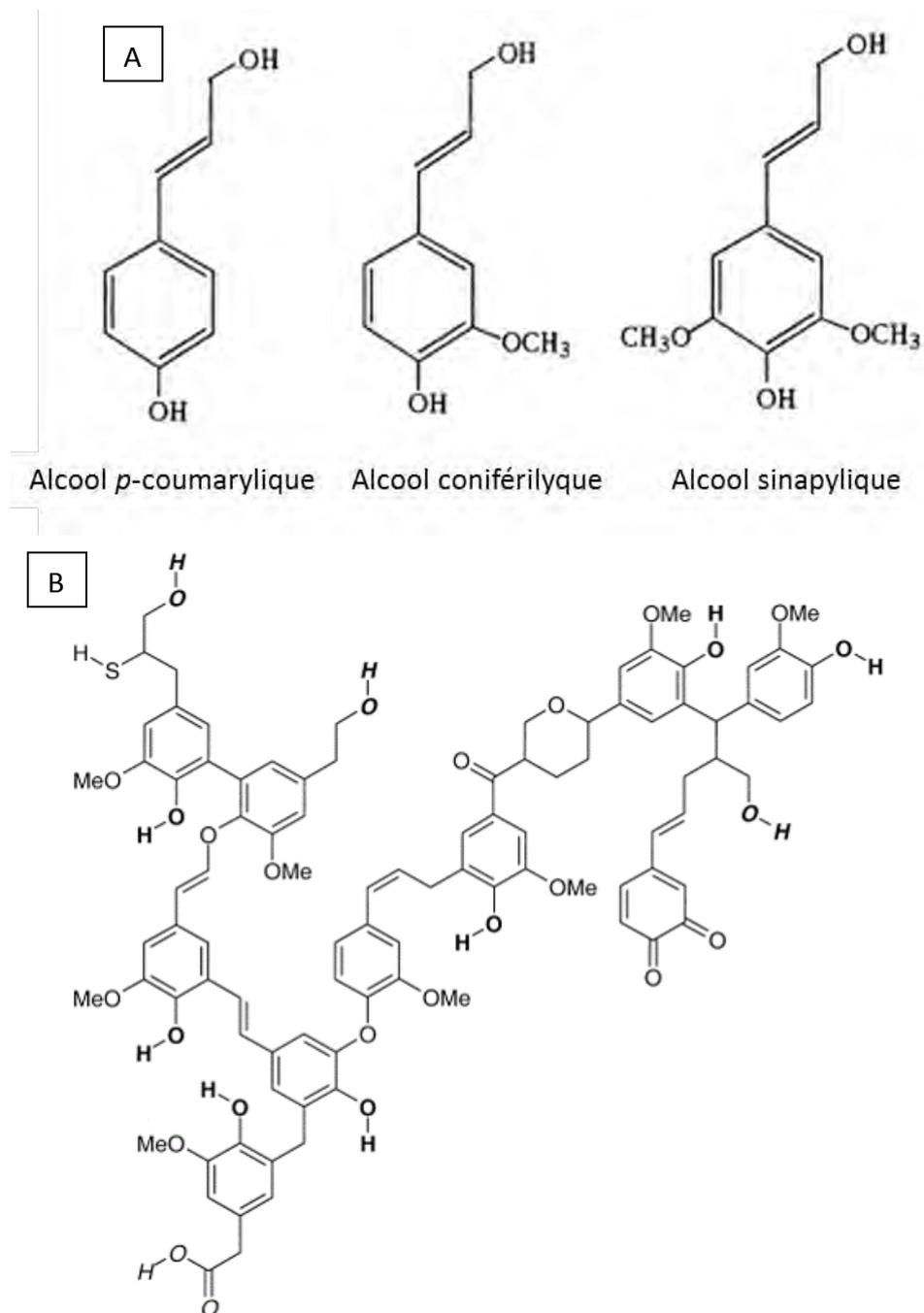


Figure 5 : Structure des 3 alcools constituant les unités de base de la lignine (A) et représentation d'une lignine (B). (Ishii 1997).

III. Dégradation des polysaccharides de la paroi végétale

Valoriser la biomasse végétale, et particulièrement les hémicelluloses, est devenu un enjeu majeur de ces dernières années notamment dans le recyclage des déchets végétaux et pour la production d'énergie afin de remplacer les énergies fossiles. En effet, la biomasse lignocellulosique représente une source d'énergie renouvelable de choix car elle permettra de diminuer l'émission de gaz à effet de serre, comme le CO₂ produit par l'utilisation et le transport des énergies fossiles (IPCC, climate change 2007). Cela permettra aussi de développer une indépendance énergétique sans entrer en compétition avec l'alimentation humaine et animale via l'utilisation des déchets de production alimentaire.

La première génération de biocarburant consiste en la fermentation des sucres de plantes à sucre (betterave ou canne à sucre) ou des produits de dégradation de l'amidon de céréales (blé, maïs) pour produire du biofuel. Le biodiesel est, quant à lui, produit à partir d'huiles végétales. Cependant cette première génération de biocarburant utilise des ressources alimentaires, ce qui pose problème.

La deuxième génération de biocarburant utilise la biomasse lignocellulosique. Actuellement, c'est très majoritairement la cellulose qui est utilisée pour produire de l'éthanol et pour cela il faut tout d'abord l'extraire de la paroi végétale (Himmel et al. 2007 ; Chundawat et al. 2011 ; Bornscheuer, Buchholz, et Seibel 2014 ; Várnai et al. 2014). Pour extraire la cellulose, deux grands types de traitements sont utilisés actuellement afin de rompre les liaisons qui permettent le maintien de la paroi : un traitement thermochimique ou un traitement chimique. Le traitement thermochimique se fait à haute température (140-200 °C) sous pression. Le traitement chimique peut se faire en conditions acide ou basique avec l'utilisation d'acide sulfurique, de soude ou d'ammoniac principalement (Himmel et al. 2007 ; Várnai et al. 2014).

Ces traitements sont polluants et coûteux en énergie il est donc nécessaire de développer l'utilisation de catalyseurs afin de diminuer l'impact des prétraitements sur l'environnement. Pour cela, les chercheurs s'intéressent aux enzymes permettant de catalyser la dégradation de la biomasse végétale.

A. La base de données CAZy

Toutes les enzymes impliquées dans la dégradation, la biosynthèse et la modification des oligosaccharides, des polysaccharides et des glycoconjugués sont référencées dans la base de données CAZy (Carbohydrate Active enZymes) (« CAZy - Home » 2016). Toutes les enzymes référencées dans la base de données CAZy sont appelées CAZymes (Lombard et al. 2014). Les CAZymes sont des enzymes potentiellement modulaires, actives sur les glucides et divisées en 5 classes :

- Les glycosides hydrolases (GH) qui hydrolysent les liaisons glycosidiques et/ou ajoutent des unités osidiques en bout de chaîne (transglycosylation) (Henrissat et Davies 1997),
- Les glycosyl transférases (GT) qui sont les principales enzymes responsables de la synthèse de glycosides (Coutinho et al. 2003),
- Les polysaccharides lyases (PL) qui réalisent des coupures des liaisons glycosidiques par β -élimination (Lombard et al. 2010),
- Les carbohydrate estérases (CE) qui catalysent des N- et O-déacétylations sur les sucres estérifiés (Jayani, Saxena, et Gupta 2005),
- Les enzymes ayant des activités auxiliaires (AA) qui sont toutes les enzymes ayant des activités d'oxydo-réduction et qui agissent de façon conjuguée avec d'autres CAZymes (Levasseur et al. 2013a).

A l'intérieur de chacune de ces classes, les CAZymes sont réparties en familles et sous-familles basées sur la similarité de séquence. L'appartenance à une famille corrèle avec un repliement et un mécanisme catalytique identiques. Cependant, des spécificités variées de substrats et de produits sont retrouvées au sein d'une même famille. Le terme module sera utilisé pour parler de la totalité ou d'une partie d'une séquence protéique possédant une activité catalytique ou étant capable de se lier sur un polysaccharide. Une protéine peut se composer d'un seul module ou posséder plusieurs modules ayant des activités ou fonctions similaires ou différentes.

B. Les glycosides hydrolases

Les glycosides hydrolases (GH) hydrolysent les liaisons glycosidiques et sont réparties en 135 familles, à ce jour, dans la base de données CAZy (Cantarel et al. 2009). Certaines de ces

familles sont regroupées en clans en fonction de leur structure 3D et de leur mécanisme (tableau 2) (Henrissat et Bairoch 1996). Il s'agit d'un niveau de classification supérieur à celui ne concernant que la séquence où le lien structure/fonction est établi. Très souvent les acides aminés catalytiques sont conservés dans un même clan. Le clan des GH-A est celui qui contient le plus grand nombre de familles. Les structures retrouvées dans tous les clans sont représentées figure 6.

Clan	Repliement	Mécanisme	Acides aminés catalytiques	Familles de GH
GH-A	TIM barrel (β/α) ₈	Rétention	Nucl : Glu Acide/base : Glu	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113 et 128
GH-B	β -jelly roll	Rétention	Nucl : Glu Acide/base : Glu	7 et 16
GH-C	β -jelly roll	Rétention	Nucl : Glu Acide/base : Glu	11 et 12
GH-D	TIM barrel (β/α) ₈	Rétention	Nucl : Asp Acide/base : Asp	27, 31 et 36
GH-E	6-fold β -propeller	Rétention	Nucl : Tyr + Glu ou Glu Acide/base : ? ou Glu	33, 34, 83 et 93
GH-F	5-fold β -propeller	Inversion	Nucl : Asp ou ? Acide/base : Glu ou ?	43 et 62
GH-G	TIM barrel (α/α) ₆	Inversion	Nucl : Glu ou ? Acide/base : Asp ou ?	37 et 63
GH-H	TIM barrel (β/α) ₈	Rétention	Nucl : Asp Acide/base : Glu	13, 70 et 77
GH-I	$\alpha + \beta$	Inversion	Nucl : Asp ou ? Acide/base : Glu ou ?	24, 46 et 80
GH-J	5-fold β -propeller	Rétention	Nucl : Asp Acide/base : Glu	32 et 68
GH-K	TIM barrel (β/α) ₈	Rétention	Nucl : carbonyle du substrat Acide/base : Glu	18, 20 et 85
GH-L	TIM barrel (α/α) ₆	Inversion	Nucl : Glu ou P ou H ₂ O Acide/base : Glu ou Asp	15, 65 et 125
GH-M	TIM barrel (α/α) ₆	Inversion	Nucl : Asp ou ? Acide/base : Glu	8 et 48
GH-N	β -helix	Inversion	Nucl : Asp Acide/base : Asp	28 et 49

Tableau 2 : Les 14 clans de GH avec les repliements caractéristiques et les familles de chaque clan.

Résidu nucléophile catalytique (Nucl), acide aspartique (Asp), acide glutamique (Glu), phosphate (P), résidu catalytique non encore identifié (?) (Henrissat et Bairoch 1996 ; « CAZy - Home » 2016 ; Alvarez, Goldbeck, Santos, et al. 2013 ; Divne et al. 1994 ; Lammerts van Bueren et al. 2012 ; Okazawa et al. 2015 ; Newstead et al. 2008 ; Cartmell et al. 2011a ; Gibson et al. 2007 ; Park et al.

2013 ; Lyu et al. 2014 ; Bujacz et al. 2011 ; Hsieh et al. 2010 ; Mizuno et al. 2004 ; Fushinobu et al. 2005 ; Abbott et Boraston 2007).

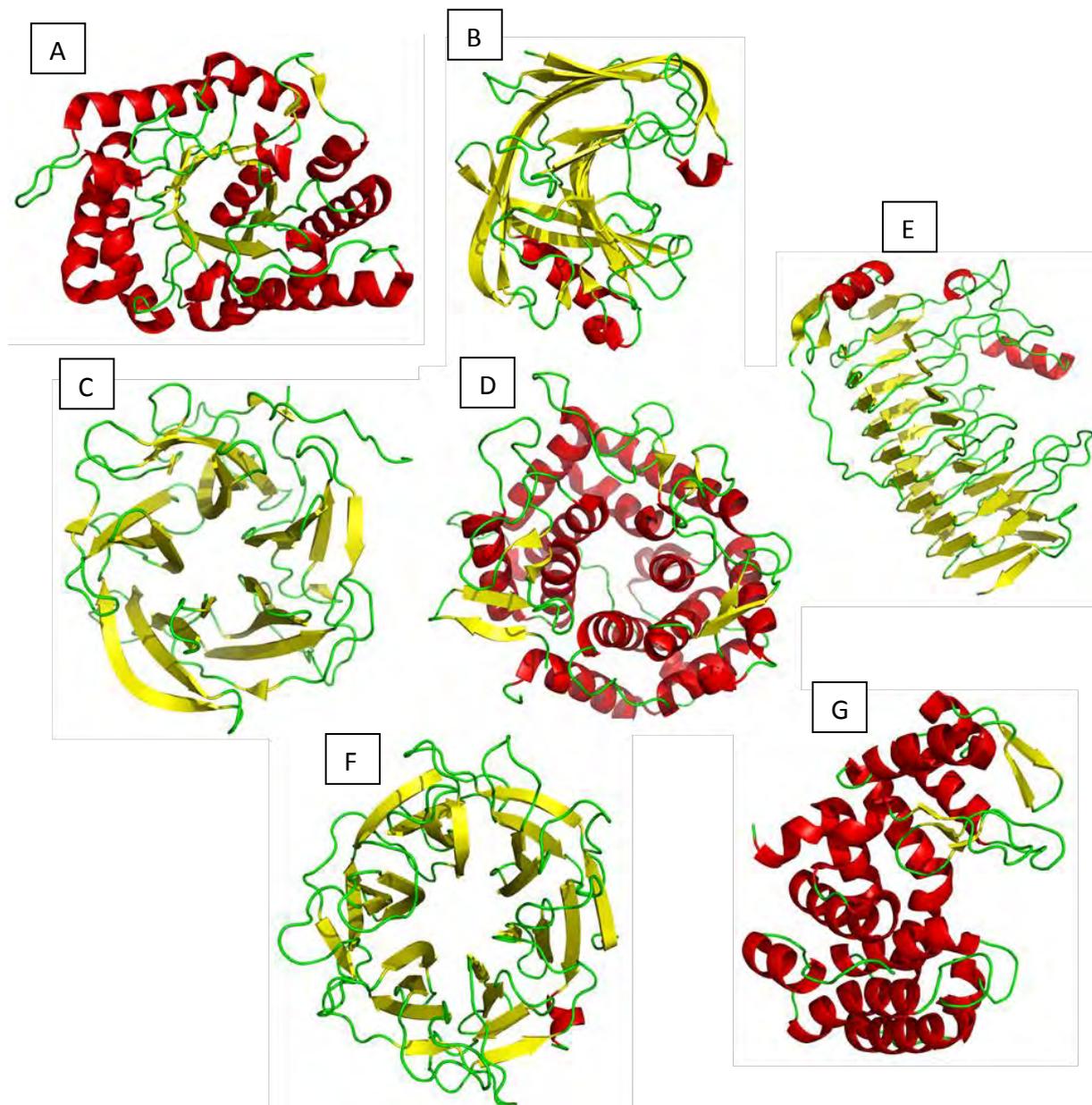


Figure 6 : Repliement de type tonneau (β/α)₈ caractéristique des clans A, D, H et K (A), β -jelly roll caractéristique des clans B et C (B), 5-fold β -propeller caractéristique des clans F et J (C), (α/α)₆ caractéristique des clans G, L et M (D), β -helix caractéristique du clan N (E), 6-fold β -propeller caractéristique du clan E (F) et $\alpha+\beta$ caractéristique du clan I (G). Les structures illustrant les repliements caractéristiques de chaque clan sont les structures A) d'une cellulase de *Fervidobacterium nodosum* appartenant à la famille GH5 (code pdb 3NCO), B) d'une endo-cellulase de *Pyrococcus furiosus* appartenant à la famille GH12 (code pdb 3WQ7), C) d'une β -xylosidase de *Bacillus halodurans* appartenant à la famille GH43 (code pdb 1YRZ), D) d'une exo-oligoxylanase de *Bacillus halodurans* appartenant à la famille GH8 (code pdb 1WU4), E) d'une polygalacturonase de

Pectobacterium carotovorum appartenant à la famille GH28 (code pdb 1BHE), F) d'une exo-arabinanase de *Fusarium graminearum* appartenant à la famille de GH93 (code pdb 2W5N) et G) d'une chitosanase de *Bacillus circulans* appartenant à la famille GH46 (code pdb 2D05). Les hélices alpha sont représentées en rouge, les brins β en jaune et les boucles en vert.

Le nom des protéines dépend de leur spécificité de substrat et plus rarement de leur mécanisme d'action. Certaines activités peuvent être retrouvées dans plusieurs familles comme par exemple l'activité endo- β 1,4-xylanase qui est présente dans les familles GH5, 8, 10, 11, 30 et 43. A l'inverse plusieurs activités différentes peuvent être retrouvées au sein d'une même famille comme par exemple la famille des GH5 (endo- β -1,4-glucanase, endo- β -1,4-xylanase, β -glucosidase, β -mannosidase, lichenase, cellodextrinase...) (Aspeborg et al. 2012). Les enzymes hydrolysant une liaison osidique à l'intérieur des chaînes polysaccharidiques sont appelées endoglycosidase, celles qui hydrolysent les liaisons osidiques par une extrémité sont dénommées exoglycosidases. Dans ce dernier cas elles auront une spécificité pour l'extrémité réductrice ou, plus fréquemment, non-réductrice du substrat (figure 7) (Davies et Henrissat 1995). D'autre part, des enzymes à activité exo qui hydrolysent les ramifications sont soit spécifiques d'une liaison (par exemple sur l'O-2 ou l'O-3 d'un résidu de la chaîne principale), soit non spécifiques et clivent alors plusieurs types de liaisons (Dumon et al. 2012 ; Lombard et al. 2014).

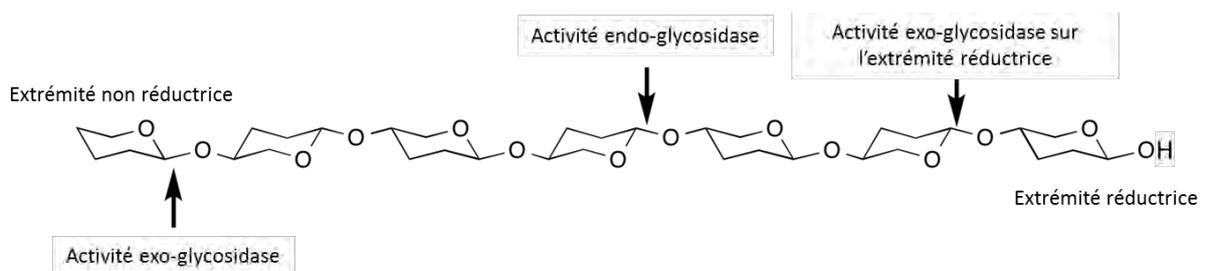


Figure 7 : Représentation schématique d'un polysaccharide et des sites de coupure pour les enzymes à activité endoglycosidase ou exoglycosidase (« CAZypedia » 2016).

Il existe une nomenclature pour définir le site catalytique des GH. Le site catalytique est ainsi divisé en sous-sites négatifs et positifs selon la nomenclature de Henrissat et Davies (Henrissat et Davies 1997). Le site de coupure se trouve entre le sous-site -1 et le sous-site +1. Dans le site catalytique des enzymes à activité endoglycosidase, les polysaccharides sont orientés avec leur extrémité réductrice du côté des sous-sites positifs et l'extrémité non-

réductrice du côté des sous-sites négatifs (figure 8A) (Alvarez, Goldbeck, Santos, et al. 2013). Chez les enzymes à activité exoglycosidase qui hydrolysent les ramifications, le résidu à hydrolyser se situe dans le sous-site -1 et la chaîne principale du polysaccharide vient se fixer dans les sous-sites positifs. L'extrémité réductrice de la chaîne principale du polysaccharide se fixe alors dans les sous-sites positifs notés +n ou +n R (pour réducteur) et l'extrémité non-réductrice se fixe dans les sous-sites positifs notés +n' ou +n NR (pour non-réducteur) (figure 8B) (Davies et Henrissat 1995).

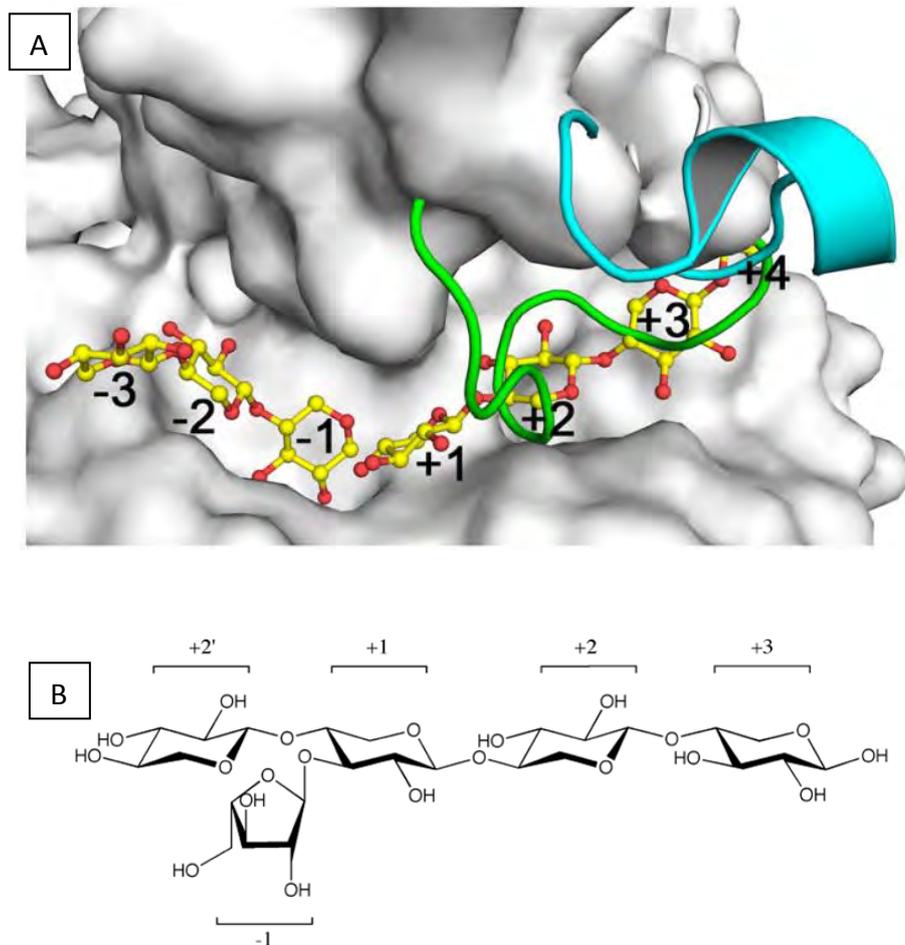


Figure 8 : A. Représentation de surface du site catalytique d'une enzyme à activité endoglycosidase en présence d'un xylooligosaccharide (Alvarez, Goldbeck, dos Santos, et al. 2013). L'hydrolyse a eu lieu entre les sous-sites -1 et +1. Les structures illustrant les sous-sites proviennent de xylanases du métagénome de sol de canne à sucre et de *Cellvibrio mixtus* appartenant à la famille GH10 (codes pdb 4K68 et 1QUY). **B. Représentation des sous-sites catalytiques d'une arabinofuranosidase (GH51 de *Thermobacillus xylanilyticus*) en présence de xylo-tétraose ramifié par un L-arabinofuranose en α -1,3- (XA³XX)** (Paës et al. 2008 ; Fauré et al. 2009).

Il existe deux grands types de mécanismes d'hydrolyse de la liaison osidique décrits à ce jour : un mécanisme dit de rétention de la configuration absolue du carbone anomérique et un mécanisme d'inversion de la configuration absolue du carbone anomérique.

Le mécanisme réactionnel de rétention de la configuration est le suivant : dans un premier temps le résidu jouant le rôle de nucléophile attaque le centre anomérique de la liaison glycosidique devant être clivée afin de déplacer la partie non réductrice et former un intermédiaire covalent glycosyl-enzyme. Pendant ce temps le résidu acide/base joue le rôle de catalyseur acide et protone l'oxygène de la liaison glycosidique à cliver. Dans un deuxième temps la liaison de l'intermédiaire glycosyl-enzyme est hydrolysée par une molécule d'eau déprotonée par le résidu acide/base agissant comme un catalyseur basique pour libérer un produit de même configuration que le substrat. Si le groupement hydroxyle d'un sucre remplace la molécule d'eau précédente alors la partie osidique du glycosyl-enzyme peut se lier à ce dernier en formant une nouvelle liaison osidique. On parle alors de transglycosylation (figure 9A) (Davies et Henrissat 1995).

Dans un mécanisme réactionnel fonctionnant par inversion de la configuration, le résidu donneur de proton va transférer un proton sur l'oxygène de la liaison glycosidique. Dans le même temps une molécule d'eau, activée par l'autre résidu catalytique par arrachement d'un de ces atomes d'hydrogènes, vient attaquer le carbone anomérique. Ces deux actions simultanées provoquent la rupture de la liaison glycosidique, la formation d'un ion pseudo-oxocarbénium et entraînent la libération d'un produit dont la configuration absolue du carbone anomérique est l'inverse de celle de ce même carbone dans le substrat (figure 9B) (Davies et Henrissat 1995).

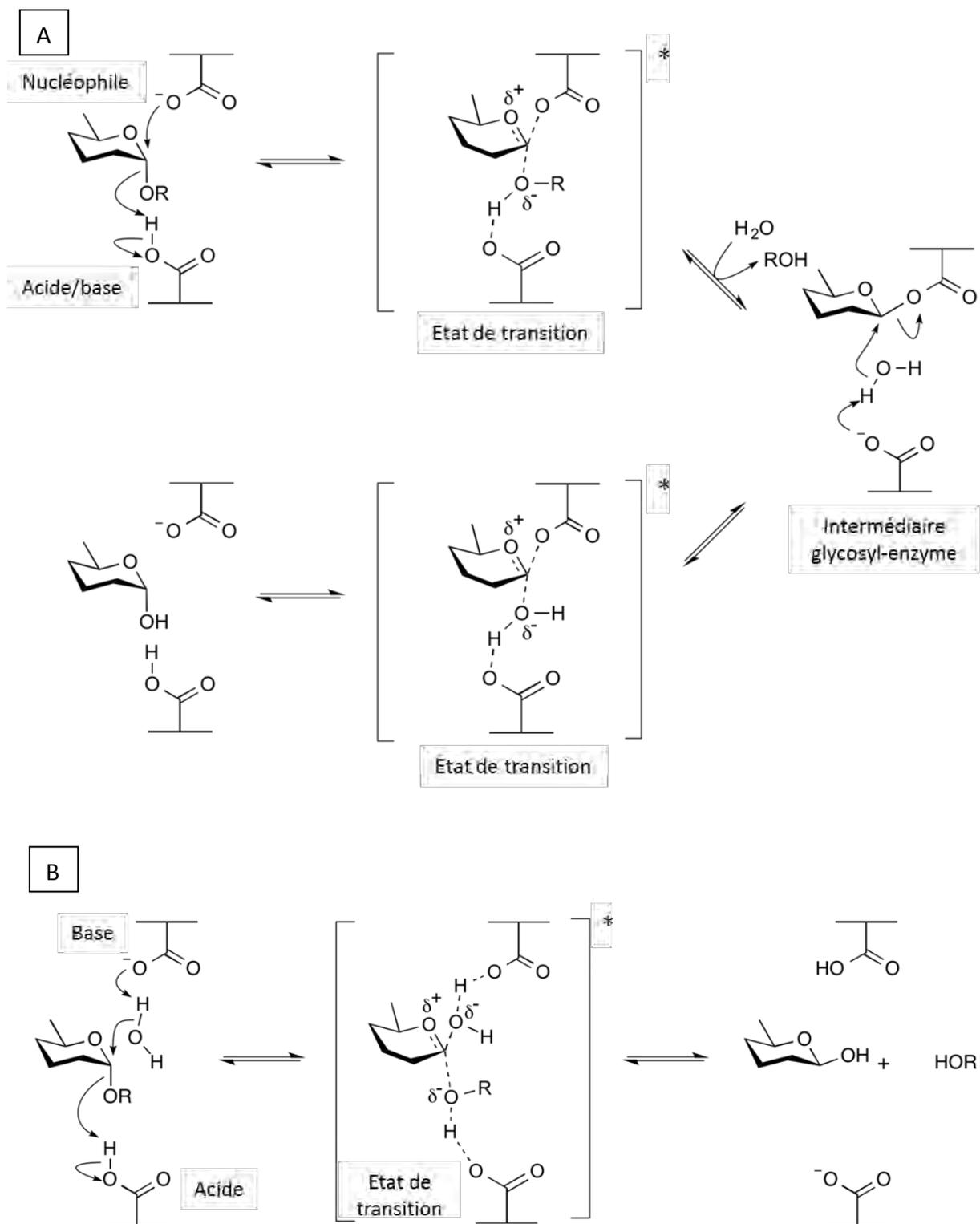


Figure 9 : Mécanisme réactionnel d'hydrolyse de la liaison glycosidique par un mécanisme de rétention (A) ou d'inversion (B) de la configuration absolue du carbone anomérique dans le cas d'une anomérie α (Rye et Withers 2000). Côté réducteur du polysaccharide (R), ion pseudo-oxocarbénium (*).

Au cours de ma thèse, je me suis particulièrement intéressée à deux familles de GH parmi les 135 présentes dans la base de données CAZy à ce jour. Il s'agit des familles GH5 et GH10 (la famille GH10 sera détaillée dans la partie 2.f.1).

La famille GH5 est une des plus grandes familles de GH, elle possède à ce jour le plus grand nombre de modules caractérisés et regroupe un grand nombre d'activités (tableau 3). Cette famille appartient au clan des GH-A et est divisée en sous-familles, 53 ont été répertoriées à ce jour. Ces sous-familles ont été créées sur la base de la relation entre la séquence et la spécificité catalytique afin de faciliter la prédiction de fonction de nouvelles enzymes (Aspeborg et al. 2012). Les GH5 ont un repliement en tonneau $(\beta/\alpha)_8$ et ont pour résidus catalytiques deux acides glutamiques, l'un jouant le rôle d'acide/base et l'autre celui de nucléophile. Ces deux résidus sont espacés de 5,5 Å environ (Davies et Henrissat 1995). Tous ces éléments sont caractéristiques du clan des GH-A. Leur activité maintient la configuration absolue du carbone anomérique par double inversion ce qui confère à certaines enzymes une activité de transglycosylation. En effet, dans un mécanisme par rétention de la configuration, si la molécule d'eau présente dans la figure 9A lors de la troisième étape est remplacée par un résidu osidique il se passe une réaction de transglycosylation (Davies et Henrissat 1995 ; Dilokpimol et al. 2011a ; Couturier et al. 2013 ; Lombard et al. 2014).

Activités endo	Activités exo
Endo- β -1,4-glucanase / cellulase (EC 3.2.1.4)	β -glucosidase (EC 3.2.1.21)
Endo- β -1,4-xylanase (EC 3.2.1.8)	β -mannosidase (EC 3.2.1.25)
Licheninase (EC 3.2.1.73)	β -glucosylceramidase (EC 3.2.1.45)
Glucane endo- β -1,6-glucosidase (EC 3.2.1.75)	Glucane β -1,3-glucosidase (EC 3.2.1.58)
Mannane endo- β -1,4-mannosidase (EC 3.2.1.78)	Exo- β -1,4-glucanase / cellodextrinase (EC 3.2.1.74)
Endoglycoceramidase (EC 3.2.1.123)	Cellulose β -1,4-cellobiosidase (EC 3.2.1.91)
Chitosanase (EC 3.2.1.132)	Stéryle β -glucosidase (EC 3.2.1.104)
Xyloglucane-spécifique endo- β -1,4-glucanase (EC 3.2.1.151)	β -primeverosidase (EC 3.2.1.149)
Endo- β -1,6-galactanase (EC 3.2.1.164)	Hespéridine 6-O- α -L-rhamnosyl- β -glucosidase (EC 3.2.1.168)
β -1,3-mannanase (EC 3.2.1.-)	
Arabinoxylane-spécifique endo- β -1,4-xylanase (EC 3.2.1.-)	
Mannan transglycosylase (EC 3.2.1.-)	

Tableau 3 : Activités enzymatiques retrouvées dans la famille GH5.

La famille des GH10 appartient aussi au clan des GH-A et présente majoritairement une activité endo- β -1,4-xylanase (Alvarez, Goldbeck, dos Santos, et al. 2013). Comme les GH5 et toutes les enzymes du même clan, les GH10 ont une structure en tonneau $(\beta/\alpha)_8$. Elles ont

également un mécanisme d'action avec rétention de la configuration absolue du carbone anomérique (Henrissat et Davies 1997). Les résidus catalytiques sont, comme pour les GH5, deux acides glutamiques, l'un jouant le rôle d'acide/base et l'autre de nucléophile.

C. Carbohydate estérases (CE)

Les carbohydate estérases ont pour fonction d'éliminer les estérifications qui pourraient gêner l'action des glycosides hydrolases. Les CE catalysent les O ou N-desestérifications de polysaccharides. On considère deux classes de substrats : ceux dont le sucre joue le rôle d'acide et ceux dont le sucre joue le rôle d'alcool. Dans la base de données CAZy, il existe 16 familles de CE et les enzymes y sont réparties en fonction de la similarité de séquence (Cantarel et al. 2009). Toutes les CE d'une même famille ont le même repliement. Le motif catalytique le plus retrouvé est constitué de la triade Sérine-Histidine-Aspartate. Le site catalytique des CE est lui aussi divisé en sous-sites catalytiques. La fonction ester à hydrolyser est située dans le sous-site 0. L'extrémité réductrice de l'oligo- ou du polysaccharide est du côté des sous-sites positifs et l'extrémité non réductrice du côté des sous-sites négatifs (figure 10) (Biely 2012).

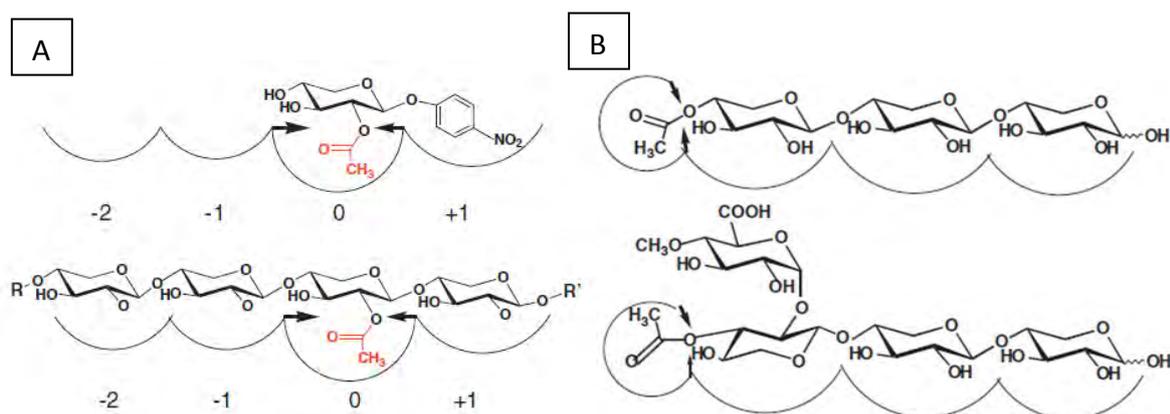


Figure 10 : Schéma montrant l'action d'une acétylxylane estérase. A. Hydrolyse d'un groupement acétyl en milieu de chaîne. L'hydrolyse se fait dans le sous-site 0 et est symbolisée par les flèches. Hydrolyse du groupement acétyl d'un 2-O-acetyl-4-nitrophenyl β-D-xylopyranoside (en haut) et du groupement acétyle d'un actéylglucuronoxylane (en bas). **B. Hydrolyse d'un groupement acétyle en bout de chaîne.** Le substrat est un xylotriose acétylé en position 4 (en haut) ou un acide aldotétrauronique (en bas) (Biely 2012).

Les structures 3D retrouvées présentent des repliements de type $(\alpha/\beta/\alpha)$ sandwich, tonneau $(\beta/\alpha)_7$, hélice β , tonneau $(\beta/\alpha)_8$, sandwich à 2 couches, ou de type α/β (figure 11) (Biely 2012).

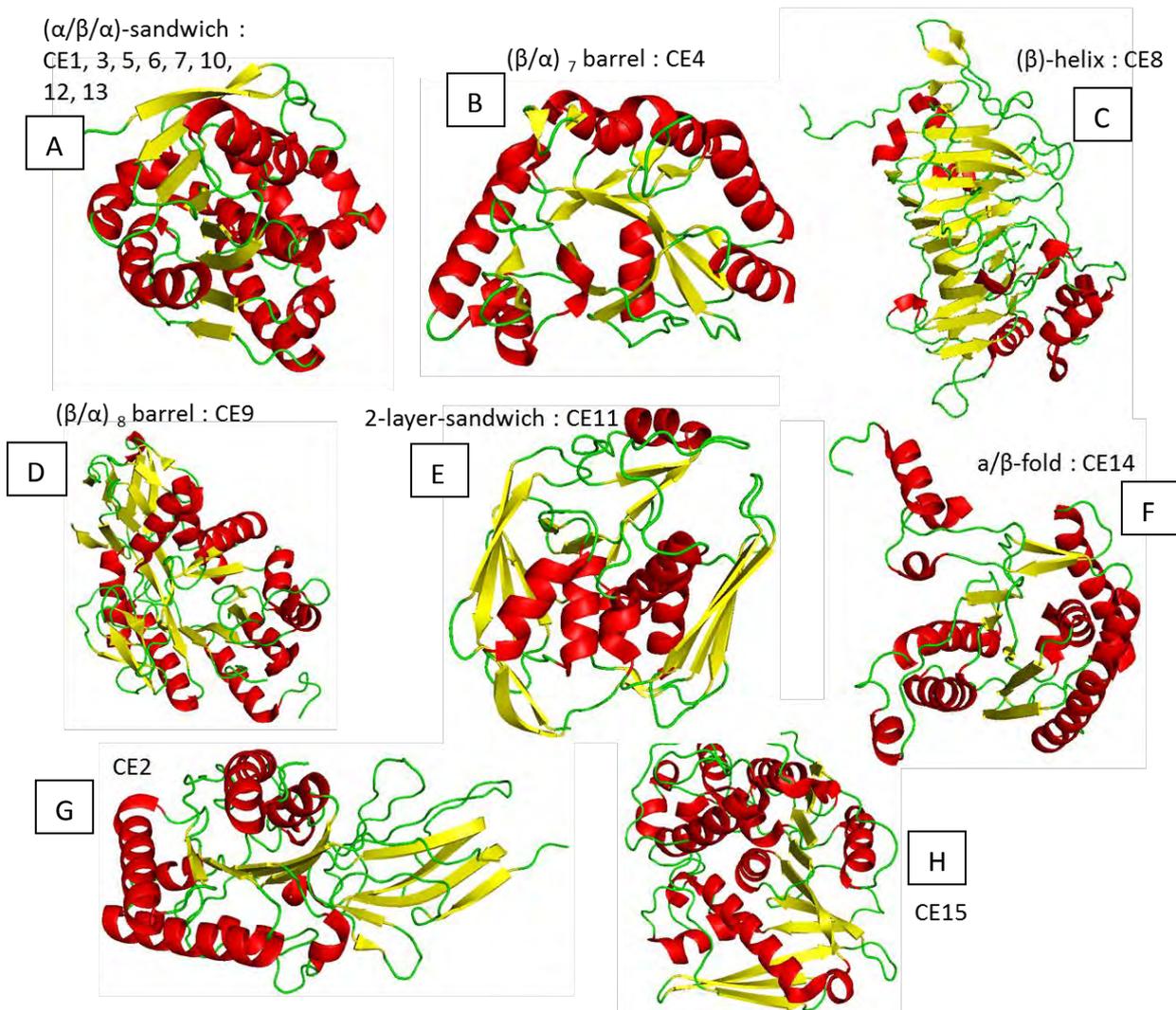


Figure 11 : Repliements des Carbohydrate Esterases : $(\alpha/\beta/\alpha)$ sandwich (A), tonneau $(\beta/\alpha)_7$ (B), hélice β (C), tonneau $(\beta/\alpha)_8$ (D), sandwich à deux couches (E), repliement α/β (F) et les structures caractéristiques des CE2 (G) et 15 (H) qui n'ont pas de nom. Les structures illustrant les repliements caractéristiques des CE sont les structures A) d'une mycolyltransférase de *Mycobacterium tuberculosis* appartenant à la famille CE1 (code pdb 1R88), B) d'une peptidoglycane déacétylase de *Ammonifex degensii* appartenant à la famille CE4 (code pdb 4WCJ), C) d'une pectine méthylestérase de *Escherichia coli* appartenant à la famille CE8 (code pdb 3GRH), D) d'une N-acétylglucosamine déacétylase de *Escherichia coli* appartenant à la famille CE9 (code pdb 1YRR), E) d'une N-acétylglucosamine déacétylase de *Aquifex aeolicus* appartenant à la famille CE11 (code pdb 1BHE), F)

d'une diacylchitobiose déacétylase de *Pyrococcus horikoshii* appartenant à la famille de CE14 (code pdb 3WE7), G) d'une acétylxylane estérase de *Cellvibrio japonicus* appartenant à la famille de CE2 (code pdb 2WAA) et H) d'une méthylglucuronyl transférase de *Thermothelomyces thermophila* appartenant à la famille de CE15 (code pdb 4G4G). Les hélices α sont représentées en rouge, les feuillets β en jaune et les boucles en vert.

D. Enzymes à activité auxiliaire (AA)

Cette classe d'enzymes, créée plus récemment, regroupe les enzymes qui catalysent des réactions d'oxydo-réduction ou des réactions radicalaires et sont réparties en 16 familles à ce jour. Parmi les activités retrouvées au sein des AA, un type d'activité est particulièrement étudié ces dernières années car il permet la dégradation de polysaccharides récalcitrants tels que la cellulose et la chitine (Vaaje-Kolstad et al. 2010 ; Forsberg et al. 2011). Il s'agit des « lytic polysaccharide mono-oxygénases » (LPMO) qui sont classées dans les familles d'AA 9 et 10. Ces enzymes sont impliquées dans la dégradation de la cellulose et de la chitine cristalline (Vaaje-Kolstad et al. 2010 ; Forsberg et al. 2011).

Les LPMO de la famille AA9 sont des polysaccharides mono-oxygénases dépendantes du cuivre. Elles possèdent 2 résidus histidine impliqués dans un centre-cuivre de type 2 qui consiste en un atome de cuivre maintenu par 4 liaisons de coordination sur un même plan par trois azotes de deux histidines (figure 12A) et l'oxygène d'une molécule d'eau (Pujol 2010 ; Frandsen et al. 2016). L'histidine N-terminale impliquée dans la coordination du cuivre est méthylée (sur l'azote du cycle imidazole ne participant pas à la coordination du cuivre, figure 12A) mais le rôle de cette méthylation reste encore inconnu (Li et al. 2012). Le clivage oxydatif de la cellulose, catalysé par les AA9, nécessite la présence d'agents réducteurs. La réaction se fait par oxydation du carbone anomérique du D-glucopyranose mais peut aussi avoir lieu sur les carbones 4 et 6. Les LPMO de la famille AA9 ont principalement été identifiées dans des génomes de champignons (figure 12A) (Westereng et al. 2011 ; Quinlan et al. 2011 ; Levasseur et al. 2013a ; Frandsen et al. 2016).

Les LPMO de la famille AA10 sont aussi des mono-oxygénases dépendantes du cuivre qui utilisent une molécule de dioxygène et des histidines mais, contrairement aux AA9, oscillent entre un centre cuivre I où la coordination du cuivre est sous forme trigonale et un centre cuivre II où la coordination du cuivre est plane (Forsberg et al. 2016). Le clivage de cellulose ou de chitine cristalline par des LPMO de la famille AA10 produit des acides aldoniques.

Contrairement aux enzymes de la famille AA9, les LPMO de la famille AA10 n'ont qu'un seul résidu histidine conservé, situé du côté de l'extrémité N-terminale de la protéine. Ce résidu est aussi impliqué dans le maintien de l'atome de cuivre (Forsberg et al. 2016). Les AA10 ne sont pas spécifiques d'un règne mais sont particulièrement retrouvées chez les bactéries (figure 12B) (Aachmann et al. 2012 ; Levasseur et al. 2013a ; Wu et al. 2013 ; Forsberg et al. 2016).

Les enzymes de ces 2 familles sont souvent multimodulaires. Les AA9 sont retrouvées associées à des CBM1 et 18, tandis que les AA10 sont associées à des CBM2, 5, 10 et 12 (Levasseur et al. 2013a).

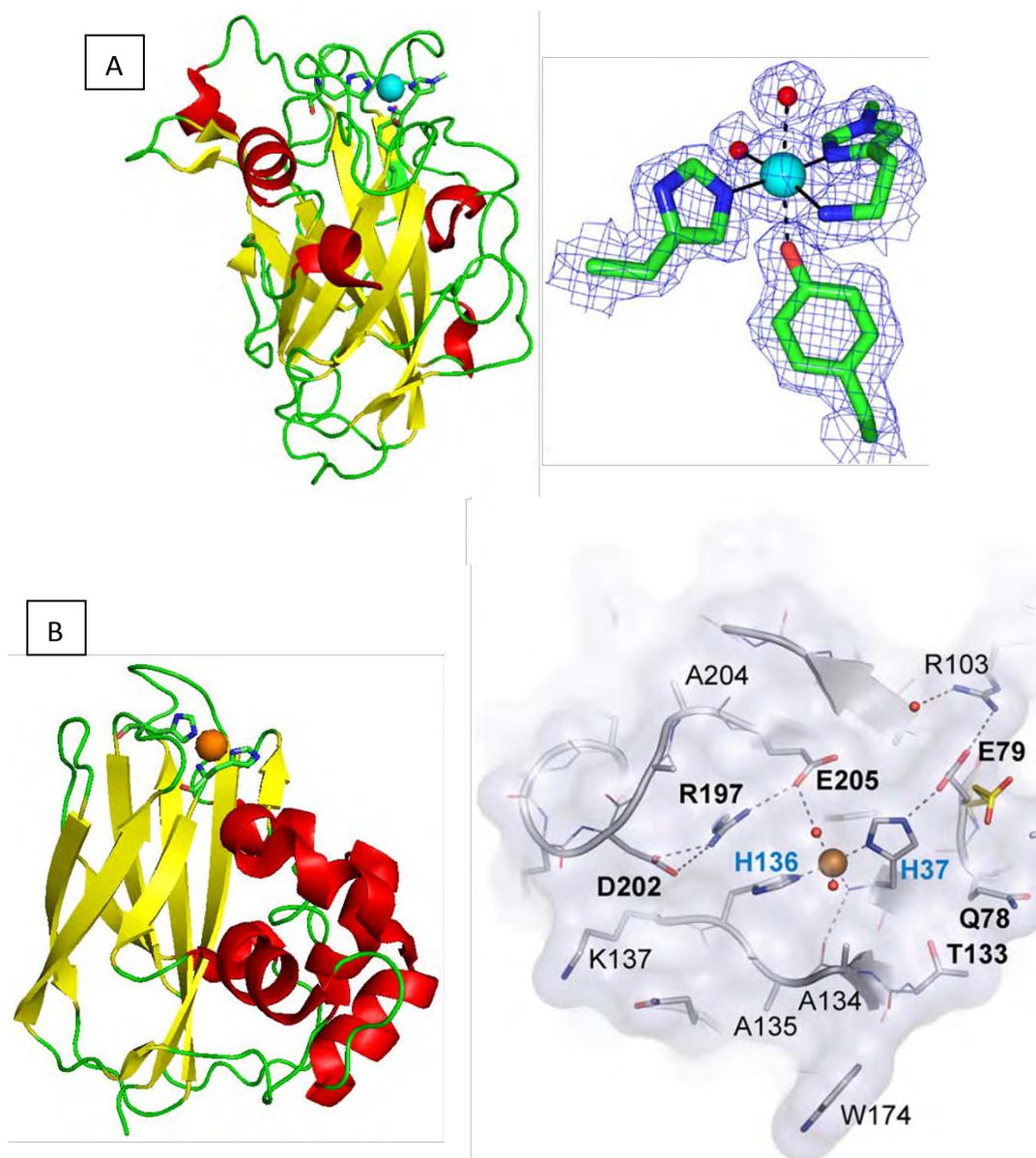


Figure 12 : Structures de LPMO des familles AA9 (A) et AA10 (B) avec les zooms respectifs sur la coordination du cuivre. Les structures illustrant les repliements caractéristiques des LPMO sont les

structures A) d'une LPMO de *Lentinus similis* appartenant à la famille AA9 (code pdb 5ACH) et B) d'une LPMO de *Cellvibrio japonicus* appartenant à la famille AA10 (code pdb 5FJQ). Les hélices α sont représentées en rouge, les brins β en jaune et les boucles en vert. L'ion cuivre II est représenté par une sphère cyan en (A) et l'ion cuivre I en orange en (B) (Frandsen et al. 2016).

E. Modules de liaison aux glucides (CBM)

Les CBM (Carbohydrate Binding Module) sont des modules généralement associés aux GH qui dégradent des substrats difficilement accessibles. Ils permettent à l'enzyme de se fixer au substrat, d'augmenter la concentration de substrat autour de l'enzyme et ainsi d'accroître de deux à cinq fois leur activité (Guillén, Sánchez, et Rodríguez-Sanoja 2009 ; Cuskin et al. 2012). Certains CBM ont la capacité de pouvoir perturber l'organisation et les propriétés physico-chimiques de certains polysaccharides (Guillén, Sánchez, et Rodríguez-Sanoja 2009). Par exemple, certains CBM sont capables de diminuer la cristallinité de la cellulose en modifiant le réseau de liaisons hydrogène qui maintiennent la cellulose ce qui va faciliter l'action des GH (Guillén, Sánchez, et Rodríguez-Sanoja 2009 ; Kang et al. 2013). Tout comme les GH, les CBM sont classés en familles en fonction de leur séquence en acides aminés. Il existe à ce jour 81 familles de CBM. Différents repliements sont retrouvés comme illustré sur la figure 13 (Boraston et al. 2004 ; Guillén, Sánchez, et Rodríguez-Sanoja 2009). Les CBM ne présentant pas d'activité catalytique propre sont très souvent associés à un domaine catalytique, principalement des GH. Les séquences codant les CBM sont présents isolés ou en tandem sur des fragments d'ADN comme dans le cas des CBM impliqués dans les cellulosomes (complexes multienzymatiques ancrés à la surface des bactéries qui les produisent, permettant d'avoir une grande concentration de protéines caractérisées principalement pour l'hydrolyse de la cellulose, détaillé en section IV. B.) (Bayer et al. 2004 ; Flint et al. 2008 ; Cantarel et al. 2009).

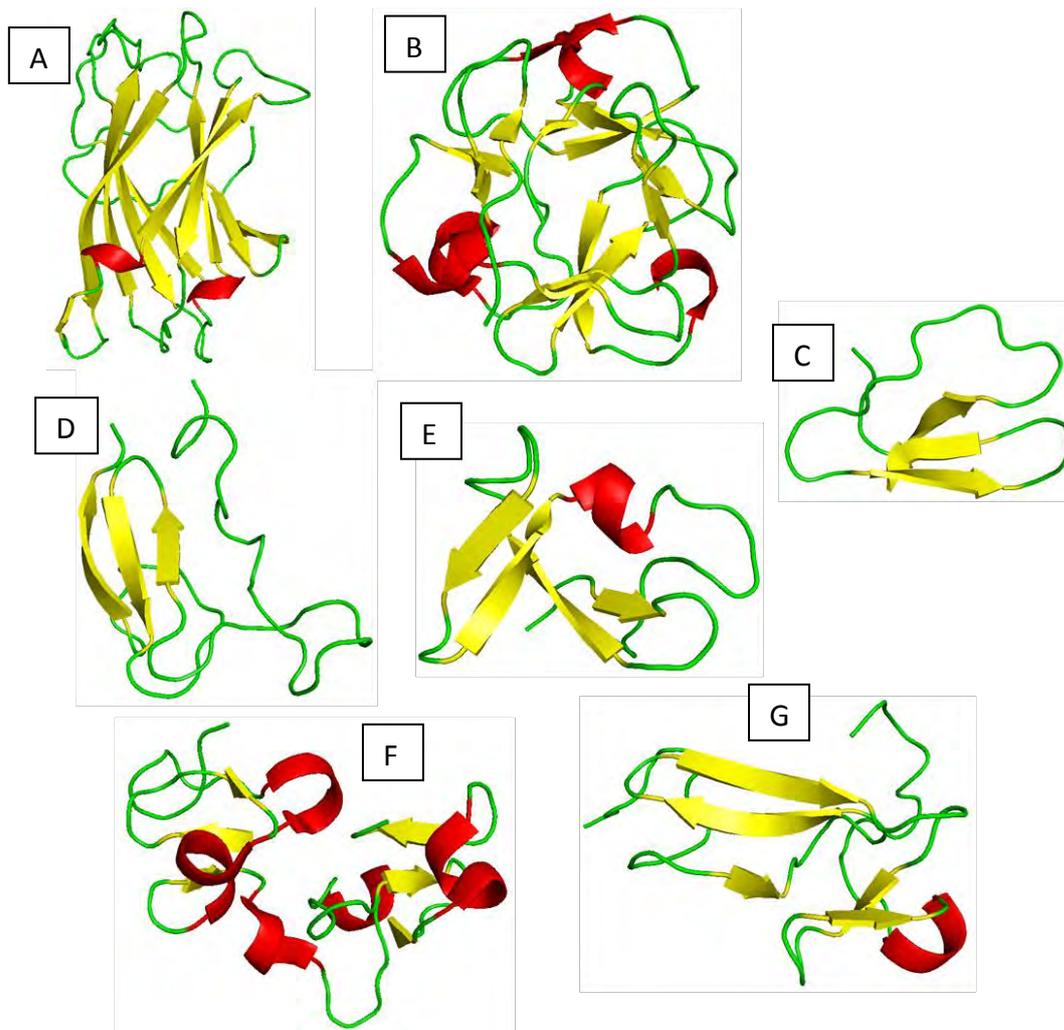


Figure 13 : Différents repliements de CBM : sandwich β : repliement 1 des CBM (A), trèfle β : repliement 2 des CBM (B), nœud cystéine : repliement 3 des CBM (C), unique : repliement 4 des CBM (D), repliement OB : repliement 5 des CBM (E), repliement hévéine : repliement 6 des CBM (F) et unique contenant des repliement hévéine : repliement 7 des CBM (G). Les structures illustrant les repliements caractéristiques des CBM sont les structures A) d'un CBM se fixant à la cellulose de *Ruminiclostridium thermocellum* appartenant à la famille CBM3 (code pdb 1NBC), B) d'un CBM se fixant au xylane de *Streptomyces lividans* appartenant à la famille CBM13 (code pdb 1MC9), C) d'un CBM se fixant à la cellulose de *Trichoderma reesei* appartenant à la famille CBM1 (code pdb 1CBH), D) d'un CBM se fixant à la chitine de *Dickeya dadantii* appartenant à la famille CBM5 (code pdb 1AIW), E) d'un CBM se fixant à la cellulose de *Cellvibrio japonicus* appartenant à la famille CBM10 (code pdb 1E8R), F) d'un CBM se fixant à la chitine de *Urtica dioica* appartenant à la famille CBM18 (code pdb 1EN2) et G) d'un CBM se fixant à la chitine de *Tachypleus tridentatus* appartenant à la famille CBM14 (code pdb 1DCQ). Les hélices α sont représentées en rouge, les brins β en jaune et les boucles en vert.

F. Exemple de la dégradation totale d'un arabinoxylane

La dégradation totale d'un arabinoxylane nécessite un ensemble d'enzymes ayant des activités diverses.

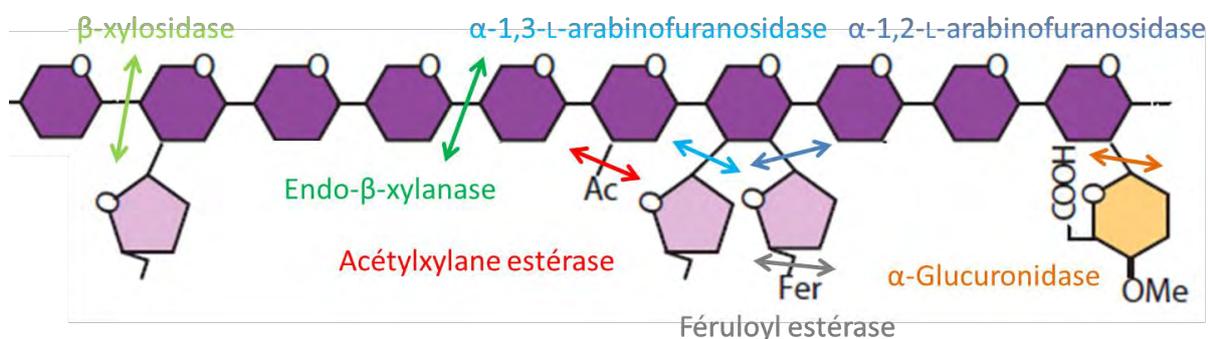


Figure 14 : Représentation schématique de la dégradation d'un arabinoxylane. Les résidus de D-xylopyranoses liés en β-1,4 sont représentés en violet, les ramifications α-L-arabinofuranosides sont représentées en mauves, les ramifications d'acide glucuronique sont représentées en orange, les groupements acétyles sont notés « Ac », les groupements féruloyle sont notés « Fer ». Les flèches représentent les coupures des liaisons par des enzymes (d'après Scheller et Ulvskov 2010).

La chaîne principale d'un arabinoxylane est composée de D-xylopyranoses liés en β-1,4. Afin de la dégrader totalement, il faut 2 types d'activités : des endo-β-1,4-xylanases (GH5, 8, 10, 11, 30, 43, 51 et 98) qui vont hydrolyser la chaîne principale en libérant des xylooligosaccharides (figure 14, flèches en vert foncé) et des β-xylosidases (GH3, 30, 39, 43, 52 et 120) qui sont des exoglycosidases, qui vont hydrolyser, soit les xylooligosaccharides libérés par les endo-β-1,4-xylanases soit, plus rarement, directement la chaîne principale d'un arabinoxylane (figure 14, flèches en vert clair). Les 2 principales familles présentant une activité endo-β-1,4-xylanase sont les GH10 et 11 (« CAZy - Home » 2016). Les principales familles de β-xylosidases (exoglycosidases) sont les familles GH39 et 43.

1. Les endo-β-1,4-xylanases

a) Les GH10

Les GH10 sont principalement des endo-β-1,4-xylanases (figure 14, flèche en vert foncé). Toutes les enzymes de cette famille présentent un repliement en tonneau (β/α)₈ caractéristique du clan GH-A (figures 6A et 15A). Les GH10 ont un mécanisme d'action avec rétention de la configuration (Henrissat et Davies 1997). Les résidus catalytiques sont deux acides glutamiques, l'un jouant le rôle d'acide/base et l'autre de nucléophile, et sont

localisés près de l'extrémité C-terminale des brins $\beta 4$ et $\beta 7$ (Jenkins et al. 1995). Le site catalytique est un sillon ouvert dans lequel se trouvent 4 à 7 sous-sites catalytiques (5 ou 6 le plus souvent) (figure 15B) (Pell et al. 2004 ; Alvarez, Goldbeck, dos Santos, et al. 2013). Les sous-sites -2, -1 et +1 sont les plus conservés. Le sous-site -2 présente une très forte affinité pour les D-xyloses ainsi que pour les groupements féruloyl (Vardakou et al. 2005). Du côté réducteur, le polysaccharide est maintenu principalement par des interactions de van der Waals s'établissant entre le polysaccharide et des résidus aromatiques tels que des tryptophanes. Du côté non-réducteur, il est maintenu par des liaisons hydrogène s'établissant avec des acides glutamiques et aspartiques, des histidines, des asparagines, des lysines et des tryptophanes (figure 15C et D). Les résidus impliqués dans la fixation et le maintien du substrat sont eux aussi très conservés (Alvarez, Goldbeck, dos Santos, et al. 2013). Les substrats accommodés peuvent être des xylanes aussi bien linéaires que ramifiés ou des xylooligosaccharides. De plus, de nombreuses GH10 sont associées à des CBM et leur caractérisation montre un effet important de la présence des CBM dans l'activité enzymatique (Pollet, Delcour, et Courtin 2010).

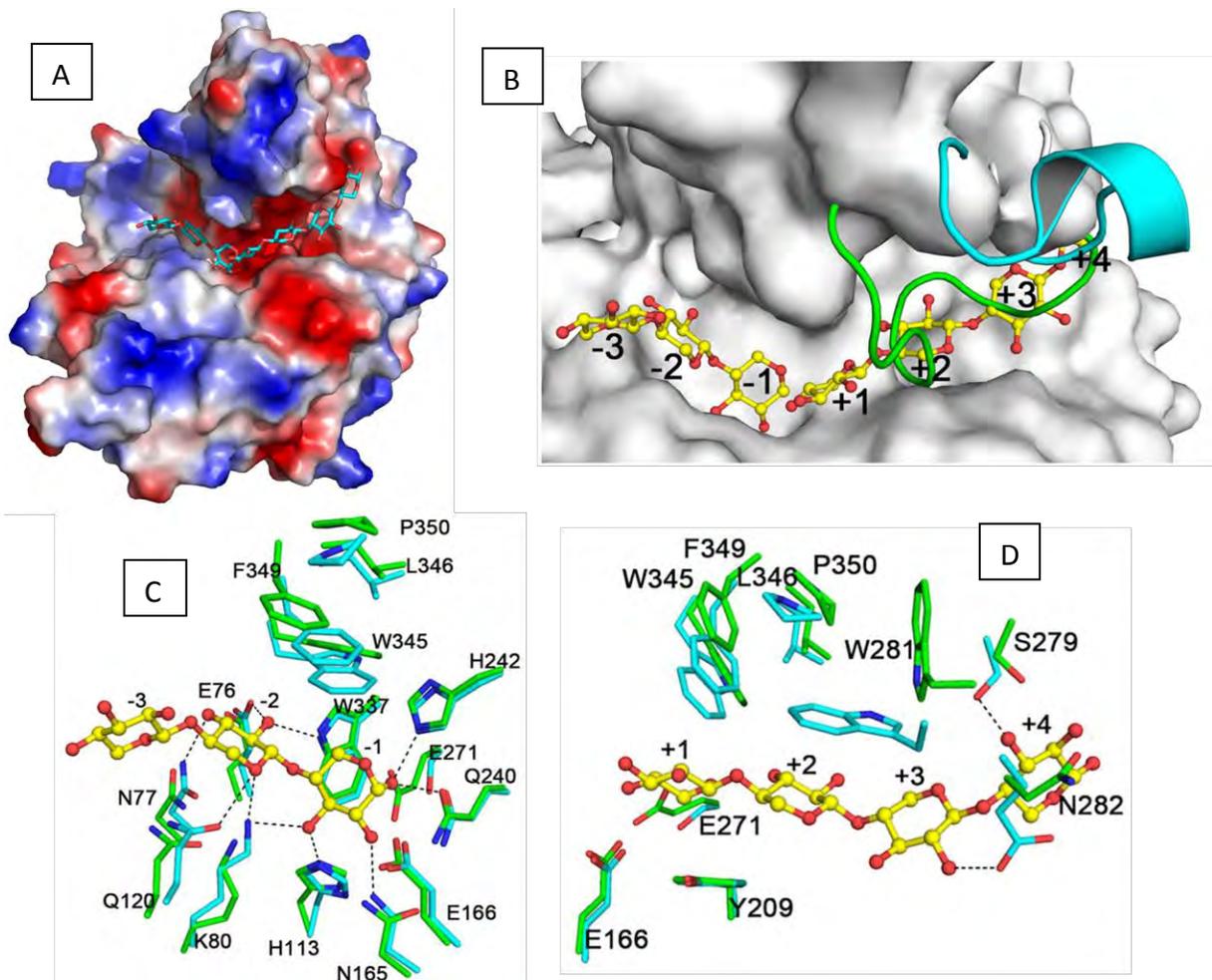


Figure 15 : A. Représentation de surface d'une GH10 en présence du substrat, un xylooligosaccharide. Les zones électronégatives sont en rouge et les zones électropositives en bleu. **B. Zoom sur le site catalytique. Superposition des résidus impliqués dans le maintien du substrat de 2 GH10 dans les sous-sites négatifs (C) et positifs (D).** Les structures illustrant les sous-sites proviennent de xylanases du métagénome de sol de canne à sucre et de *Cellvibrio mixtus* appartenant à la famille GH10 (codes pdb 4K68 et 1QUY) (Alvarez, Goldbeck, dos Santos, et al. 2013).

L'autre principale famille de xylanases est la famille des GH11.

b) Les GH11

Les GH11 qui ont été caractérisées sont, contrairement aux GH10, presque exclusivement des endo- β -1,4-xylanases qui sont souvent séparées en 2 catégories, les xylanases acides et les xylanases basiques, en fonction de leur pH optimal (autour de pH 2 et pH 11 respectivement) (Joshi et al. 2000 ; Sapag et al. 2002 ; Pollet, Delcour, et Courtin 2010). Elles appartiennent au clan GH-C et adoptent un repliement en β -jelly roll caractéristique de ce clan. Leur repliement est assimilable à la forme d'une main droite, la paume formée par la partie vrillée du feuillet β et l'hélice α , le pouce formé par boucle entre les brins β 7 et β 8 et les doigts formés par les boucles connectant les brins β faisant face au pouce (figures 6B et 16A) (Henrissat et Davies 1997 ; Paës et al. 2007 ; Lammerts van Bueren et al. 2012). Ces enzymes fonctionnent avec un mécanisme de rétention de la configuration absolue du carbone anomérique pour hydrolyser leurs substrats. Les résidus catalytiques sont 2 acides glutamiques, l'un jouant le rôle d'acide/base et l'autre de nucléophile (Henrissat et Davies 1997 ; Gruber et al. 1998). Les résidus catalytiques sont situés dans la partie creuse de la paume. Les résidus autour du nucléophile sont plus conservés que ceux autour de l'acide/base. Les sous-sites catalytiques sont au nombre de 5 ou 6 et sont très conservés (figure 16B). Les résidus D-xylosides sont particulièrement bien fixés dans les sous-sites -2, -1 et +1, alors que dans les autres sous-sites, le maintien des résidus est plus lâche (Gruber et al. 1998 ; Pollet et al. 2010). Dans le sous-site -2, les résidus sont stabilisés par des liaisons hydrogène notamment avec une tyrosine. Les résidus d'acides aminés du pouce participent au maintien de la partie non-réductrice du polysaccharide. Le maintien du substrat peut aussi se faire *via* les groupements féruloyles accommodés dans les sous-sites -3 et +2 (Vardakou et al. 2008 ; Pollet, Delcour, et Courtin 2010). La position de la boucle

constituant le pouce détermine la largeur de la poche catalytique (Paës et al. 2007 ; Paës, Berrin, et Beaugrand 2012 ; Song et al. 2014).

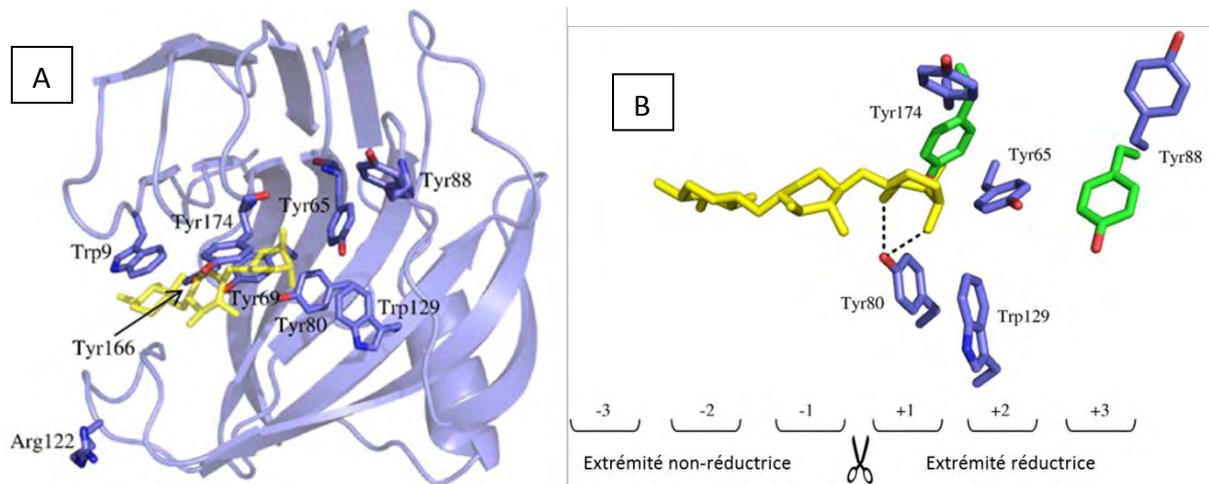


Figure 16 : A. Représentation en cartoon en présence d'un substrat d'une GH11. B. Représentation des sous-sites catalytiques. La structure illustrant les sous-sites provient d'une xylanase de *Bacillus subtilis* appartenant à la famille GH11 (code pdb 2QZ3). Le substrat, un xylotriose, est représenté en jaune (Pollet et al. 2010).

2. Les β -xylosidases

Les xylosidases sont décrites dans les familles GH 3, 30, 39, 43, 52 et 120. Les GH39 appartiennent au clan des GH-A et comme toutes les enzymes de ce clan, elles adoptent un repliement caractéristique en tonneau (β/α)₈. Ce sont majoritairement des β -xylosidases (figure 14 flèches en vert clair), toutefois on retrouve dans cette famille quelques α -L-iduronidases. Les enzymes de la famille GH39 ont un mécanisme d'action avec rétention de configuration absolue du carbone anomérique et les résidus catalytiques sont deux acides glutamiques, l'un jouant le rôle d'acide/base et l'autre de nucléophile (Henrissat et Davies 1997). Les GH39 libèrent du D-xylopyranose à partir de xylooligosaccharides, d'arabinoxxylooligosaccharides ou d'arabinoxylane (Czjzek et al. 2005 ; Lagaert et al. 2014). Par contre, ces enzymes ne présentent qu'une très faible activité hydrolytique sur du xylobiose (Adelsberger et al. 2004 ; Wagschal et al. 2005). En revanche, le mécanisme catalytique de rétention de la configuration absolue du carbone anomérique par double inversion, confère une activité transglycosidase aux enzymes de cette famille. Elles sont donc capables de former du xylobiose ou du xylotriose à partir de xyloses (Smaali, Rémond, et O'Donohue 2006 ; Muzard et al. 2009). Les GH39 possèdent 3 domaines structuraux

distincts, mais à ce jour seul le domaine catalytique a été caractérisé. Le site catalytique contient au moins 3 sous-sites (+2, +1 et -1) (figure 17) (Czjzek et al. 2005).

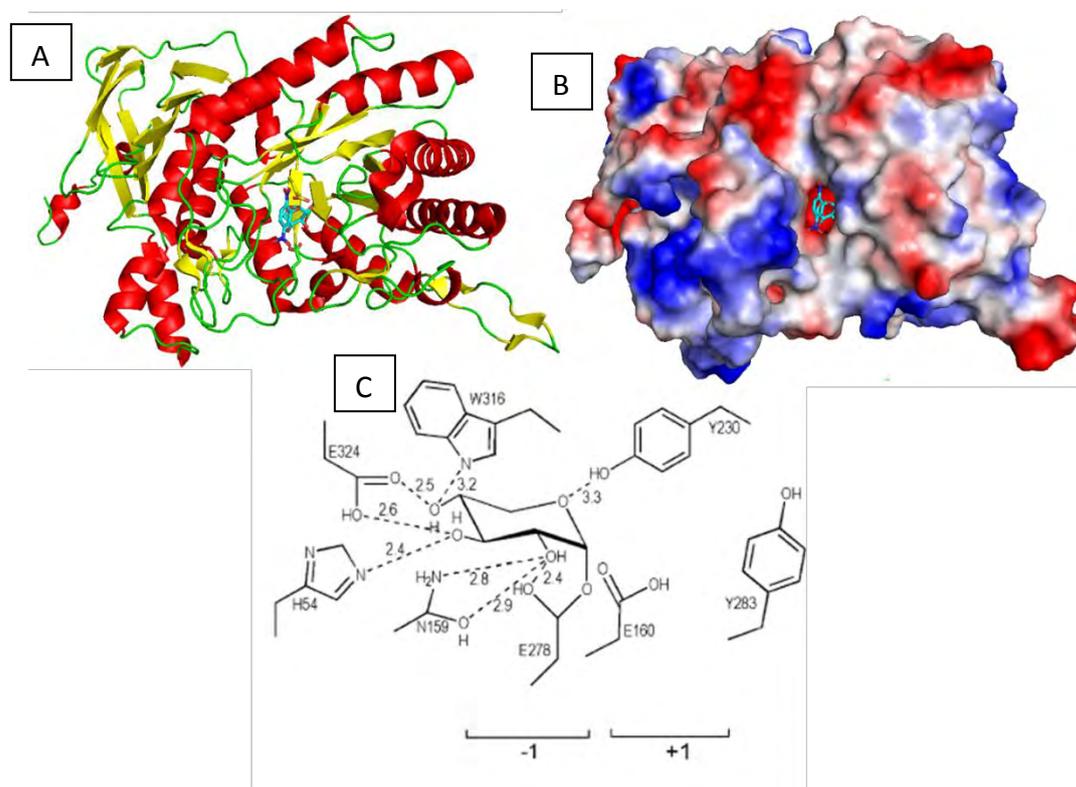


Figure 17 : Représentation en cartoon (A) et en surface (B) d'une GH39 en présence de son substrat. La structure illustrant les sous-sites provient d'une β -xylosidase de *Geobacillus stearothermophilus* appartenant à la famille GH39 (code pdb 2BFG). Les hélices α sont représentées en rouge, les brins β en jaune et les boucles en vert. Le substrat, un dinitrophényl β -xyloside, est représenté en bleu. **Représentation des acides aminés impliqués dans la réaction, vue avec l'intermédiaire covalent xylosyl-enzyme (C).** Les interactions sont représentées par des pointillés (Czjzek et al. 2005).

Pour dégrader totalement un arabinoxylane il faut retirer les ramifications qui, par ailleurs, peuvent gêner l'activité des endo- β -1,4-xylanases et des β -xylosidases. Cela nécessite l'action conjointe d' α -L-arabinofuranosidases (GH3, 43, 51, 54, 62) qui vont être spécifiques ou non des liaisons α -1,2 ou α -1,3 (figure 14, flèches bleues foncées et bleues claires respectivement) et d' α -glucuronidases (GH4, 67 et 115) (figure 14, flèches en orange). Les 2 familles d' α -L-arabinofuranosidases les mieux caractérisées sont les GH43 et 51. Les 2 principales familles d' α -glucuronidases sont les GH67 et 115. Cette dernière famille hydrolyse les acides 4-o-méthyl α -glucuroniques.

3. Les α -L-arabinofuranosidases

a) Les GH43

Les GH43 sont principalement des α -L-arabinofuranosidases mais on retrouve aussi des β -xylosidases dans cette famille. Beaucoup d'enzymes de cette famille sont bi-fonctionnelles et possèdent ces 2 activités (Lagaert et al. 2014). Les GH43 appartiennent au clan GH-F et adoptent le repliement en 5-fold β -propeller (Henrissat et Davies 1997). La famille des GH43 est divisée en 37 sous-familles à ce jour (Mewis et al. 2016). Ces enzymes hydrolysent principalement les L-arabinofuranosides liés en α -1,2 (figure 14, flèches en bleu foncé) ou les doubles substitutions (L-arabinofuranosides liés en α -1,2 et en α -1,3) (figure 14, flèches en bleu foncé et bleu clair respectivement) (Cartmell et al. 2011a). Ces enzymes ont un mécanisme d'action avec inversion de la configuration du carbone anomérique. Les résidus catalytiques sont un acide aspartique qui joue le rôle de nucléophile/base et un acide glutamique qui joue le rôle de donneur de protons (Davies et Henrissat 1995). Trois résidus sont parfaitement conservés au sein de la famille des GH43, il s'agit de deux acides aspartiques et d'un acide glutamique. Un des acides aspartiques joue le rôle de modulateur de pK_a de l'acide catalytique pour maintenir une valeur élevée de pK_a de ce dernier (Cartmell et al. 2011a ; McKee et al. 2012). Les deux autres résidus conservés sont les résidus catalytiques. Les GH43 libèrent du L-arabinose à partir d'arabinoxylane, d'arabinane et d'arabinoooligosaccharides et du D-xylose à partir de xylane et de xyloooligosaccharides. Le site catalytique des GH43 possède jusqu'à 7 sous-sites (-1, +1, +2NR*, +2NR, +3NR, +2R et +3R). Le maintien du substrat se fait majoritairement par des interactions de van der Waals avec des résidus tels que des tryptophanes ou des phénylalanines et par des liaisons hydrogène avec des tyrosines ou des histidines (figure 18). Enfin, les GH43 sont très souvent associées à un CBM (Cartmell et al. 2011a ; McKee et al. 2012 ; Lagaert et al. 2014 ; Mewis et al. 2016).

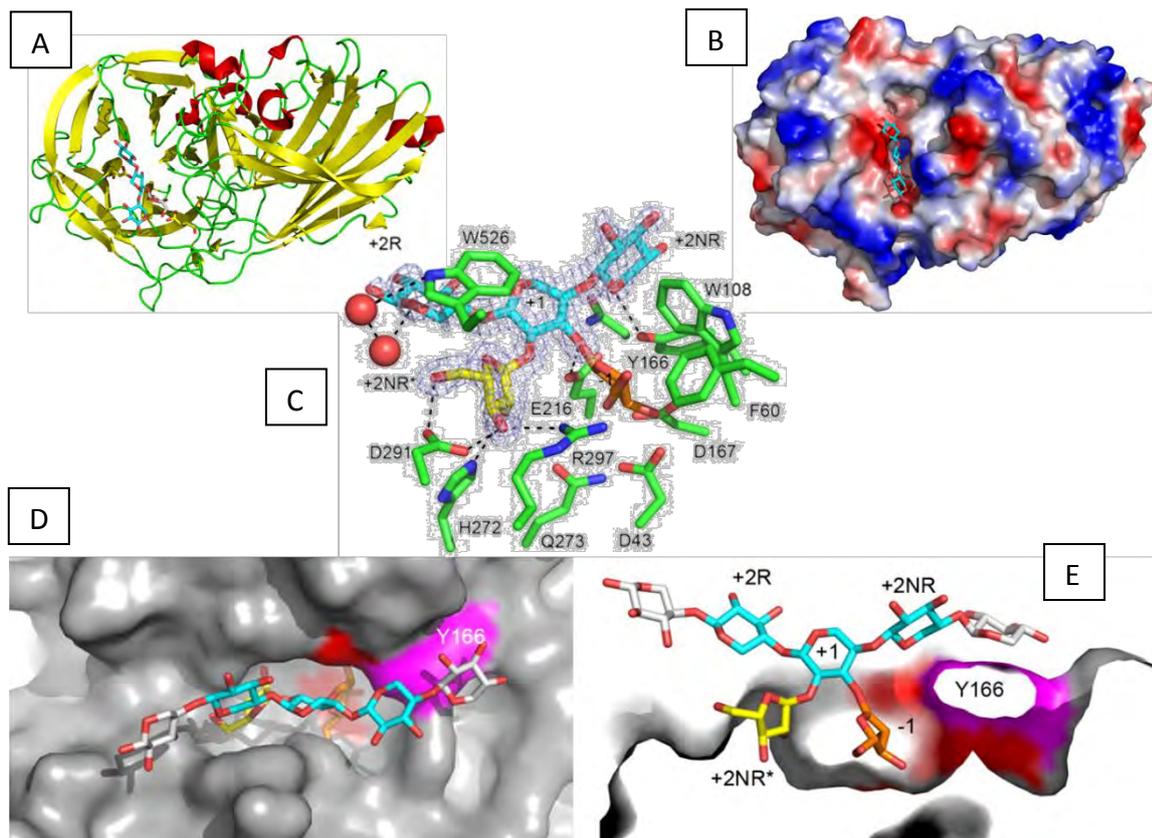


Figure 18 : Représentation en cartoon (A) et en surface (B) d'une GH43 en présence de son substrat. La structure illustrant les sous-sites provient d'une α -L-arabinofuranosidase de *Humicola insolens* appartenant à la famille GH43 (code pdb 3ZXK). Les hélices α sont représentées en rouge, les feuillets β en jaune et les boucles en vert. Le substrat, un xylopentaose doublement substitué par des L-arabinofuranosides en α -1,2 et α -1,3 (XXA²⁺³XX), est représenté en bleu. **C : Représentation des acides aminés impliqués dans la réaction.** Les interactions sont représentées par des pointillés. La carte de densité $2mF_{obs}-DF_{calc}$ est montrée autour du substrat. **D. Zoom sur le site catalytique.** **E. Représentation des sous-sites catalytiques.** Les D-xylosides sont en bleu, le résidu L-arabinofuranoside lié en α -1,3- est en jaune, le résidu L-arabinofuranoside lié en α -1,2 est en orange et les résidus d'acides aminés impliqués dans le maintien du substrat (McKee et al. 2012) sont en vert.

Parmi les différentes familles d' α -L-arabinofuranosidases citées précédemment, une autre est bien caractérisée, la famille des GH51.

b) Les GH51

Les GH51 sont très majoritairement des α -L-arabinofuranosidases mais on retrouve quelques endo-glucanases également. Cette famille de GH appartient au clan GH-A. Les

enzymes de la famille GH51 adoptent donc un repliement en tonneau (β/α)₈. En plus du domaine catalytique, un domaine supplémentaire en β -sandwich est très conservé, bien que son rôle ne soit pas encore connu. Les acides aminés catalytiques sont deux acides glutamiques, l'un jouant le rôle d'acide/base et l'autre de nucléophile et ces enzymes agissent avec un mécanisme de rétention de configuration absolue du carbone anomérique (Davies et Henrissat 1995). Les α -L-arabinofuranosidases de cette famille libèrent des α -L-arabinoses à partir d'arabinoxylanes, d'arabinoxyloligosaccharides, arabinane et arabinoooligosaccharides. Les GH51 sont actives spécifiquement sur les substitutions en α -1,2 ou α -1,3 (figure 14 flèches bleu foncé et bleu clair respectivement) et sont rarement actives sur les doubles substitutions. Le site catalytique comporte jusqu'à 5 sous-sites (-1, +1, +2, +2' et +3). De nombreux tryptophanes sont impliqués dans le maintien de la chaîne principale du substrat. Des liaisons hydrogène faites par des asparagines, des tyrosines, des acides aspartiques et glutamiques sont aussi impliquées dans la liaison au sucre (figure 19) (Taylor et al. 2006 ; Paès et al. 2008 ; Lagaert et al. 2014).

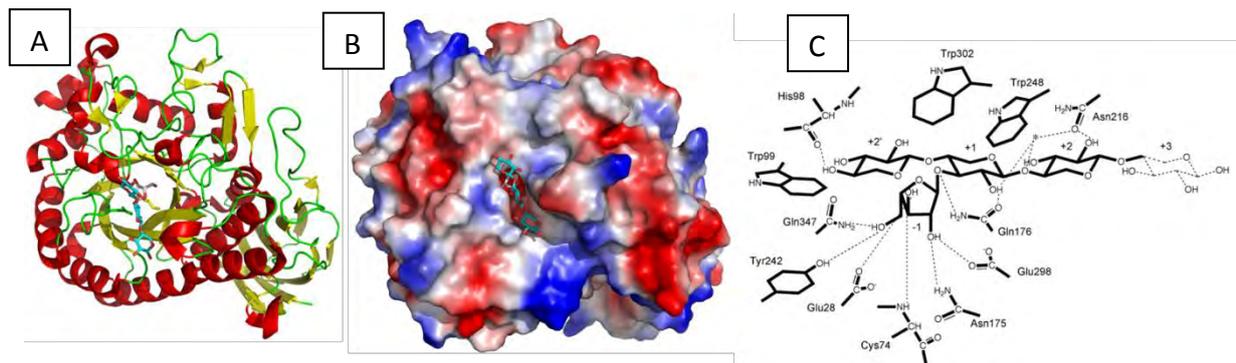


Figure 19 : Représentation en cartoon (A) et en surface (B) d'une GH51 en présence de son substrat, un XA³XX. La structure illustrant les sous-sites provient d'une α -L-arabinofuranosidase de *Thermobacillus xylanilyticus* appartenant à la famille GH51 (code pdb 2VRQ). Les hélices α sont représentées en rouge, les brins β en jaune et les boucles en vert. Le substrat est représenté en bleu. **C : Représentation des résidus impliqués dans la réaction.** Les liaisons hydrogène sont représentées par des pointillés (Paès et al. 2008).

4. Les α -glucuronidases

a) Les GH67

Les GH67 sont des α -glucuronidases (figure 14, flèches en orange). Cette famille n'est assignée à aucun clan. Les GH67 ont un repliement du domaine catalytique en tonneau

(β/α)₈, un autre domaine est présent mais sa fonction n'est pas encore connue. Deux acides glutamiques et un acide aspartique sont très conservés, un des acides glutamiques est l'acide/base. Les 2 autres résidus semblent être de bons candidats pour activer une molécule d'eau et l'un des deux serait donc le nucléophile qui n'a pas encore été identifié. Des résidus arginine, valine et tryptophane semblent être impliqués dans le maintien du substrat (figure 20) (Nurizzo et al. 2002 ; Golan et al. 2004). Les enzymes de cette famille ont un mécanisme d'inversion de la configuration absolue du carbone anomérique (Davies et Henrissat 1995). Toutes les enzymes de cette famille qui ont été caractérisées libèrent un 4-o-méthyl-glucuronate à partir d'oligosaccharides ramifiés par des 4-o-méthyl-glucuronate et n'ont pas d'activité sur polysaccharide. Les polysaccharides doivent donc être traités par d'autres enzymes au préalable pour que les GH67 puissent agir (Nurizzo et al. 2002 ; Golan et al. 2004).

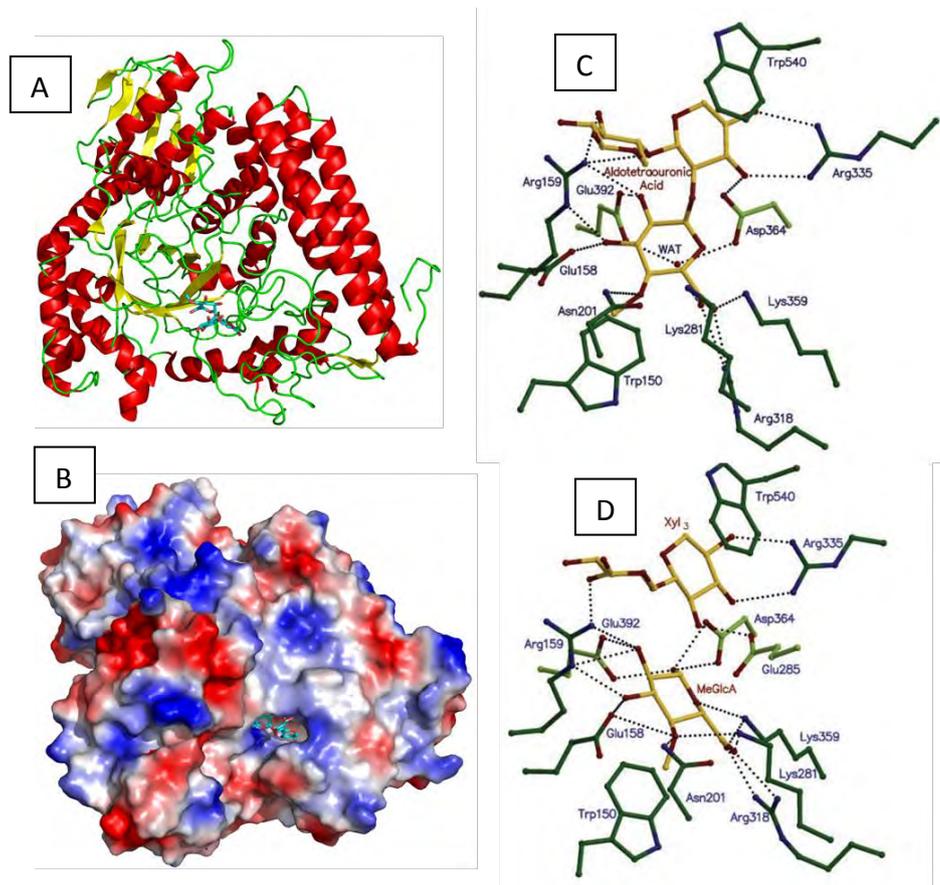


Figure 20 : Représentation en cartoon (A) et en surface (B) d'une GH67 en présence de son substrat. La structure illustrant les sous-sites provient d'une α -glucuronidase de *Geobacillus stearothermophilus* appartenant à la famille GH67 (code pdb 2VRQ). Les hélices α sont représentées en rouge, les brins β en jaune et les boucles en vert. Le substrat est représenté en bleu. **Représentation des acides aminés impliqués dans la réaction en présence du substrat (C) et des**

produits d'hydrolyse (D). Les interactions sont représentées par des pointillés. Le substrat est représenté en jaune et les résidus de la protéine en vert (Golan et al. 2004).

b) Les GH115

Les GH115 sont, comme les GH67, des α -glucuronidases (figure 14 flèches orange). Cette famille n'est assignée à aucun clan. Les GH115 ont un repliement en tonneau $(\beta/\alpha)_8$ (figure 21). Ni le résidu nucléophile, ni le résidu acide/base n'ont été identifiés à ce jour (Rogowski et al. 2014 ; Wang et al. 2016). Les enzymes de cette famille ont un mécanisme d'inversion de la configuration absolue du carbone anomérique (Davies et Henrissat 1995). Des études biochimiques laissent penser que le site catalytique de ces enzymes posséderait au moins 5 sous-sites (-1, +1, +2R, +2NR, +3R) (Rogowski et al. 2014 ; Wang et al. 2016). Les GH115 possèdent 3 domaines supplémentaires dont un en β -sandwich et un en hélice α (entourés en rouge et en orange respectivement sur la figure 21). Certains résidus ont été montrés comme étant essentiels pour l'activité, ce sont les acides glutamiques 162 et 375 et les acides aspartiques 192, 206, 332 et 478 ainsi que l'arginine 328, les tryptophanes 169, 245 et 249, la tyrosine 420 et l'histidine 422 dans le cas de l' α -glucuronidase de *Bacteroides ovatus* BoAgu115A (Rogowski et al. 2014). Toutes les enzymes de cette famille qui ont été caractérisées libèrent un 4-o-méthyl-glucuronate à partir de xylooligosaccharides ramifiés par des 4-o-méthyl-glucuronates et n'ont pas d'activité sur polysaccharides comme les GH67. Les enzymes de cette famille nécessitent donc l'action préalable d'endo-xylanases (Sunna et Antranikian 1997 ; Rogowski et al. 2014 ; Wang et al. 2016).

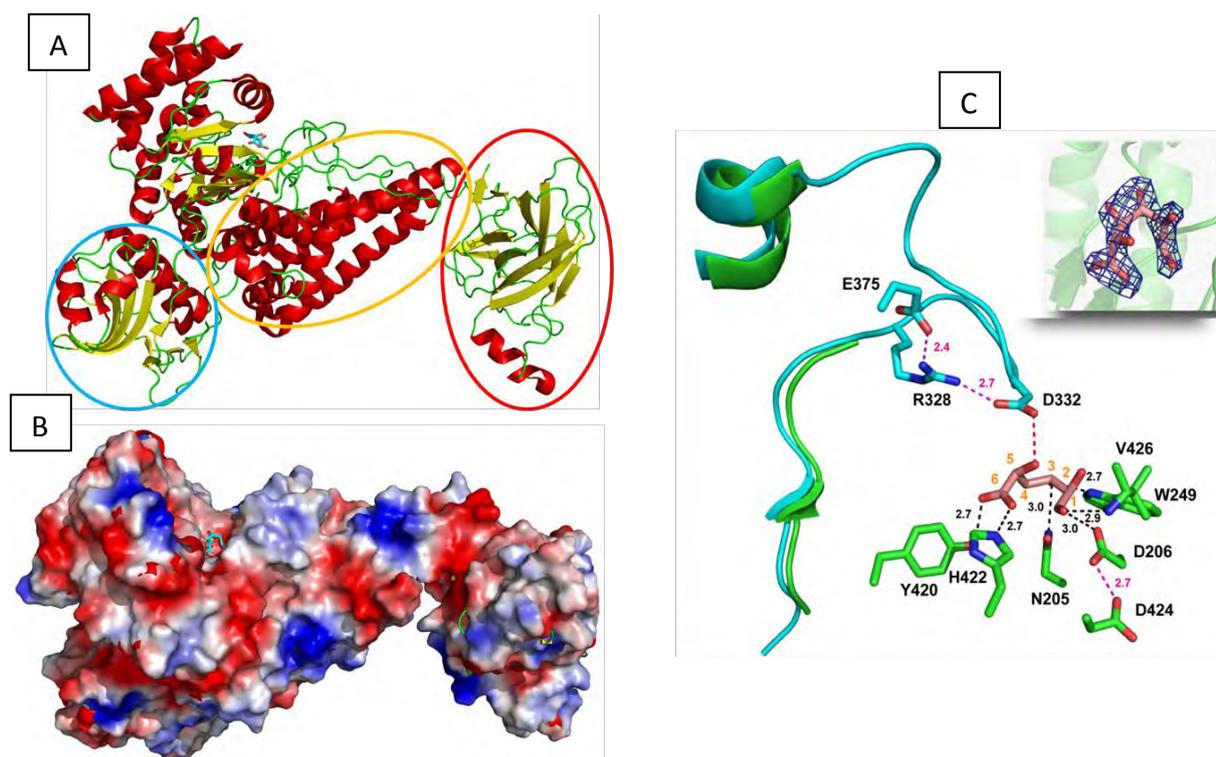


Figure 21 : Représentation en cartoon (A) et en surface (B) d'une GH115 en présence de son substrat. La structure illustrant les sous-sites provient d'une α -glucuronidase de *Bacteroides ovatus* appartenant à la famille GH115 (code pdb 4C91). Les hélices α sont représentées en rouge, les brins β en jaune et les boucles en vert. Les trois domaines supplémentaires sont entourés. Le substrat, un acide glucuronique est représenté en bleu. **C : Représentation des acides aminés impliqués dans la réaction.** Les interactions sont représentées par des pointillés. Le substrat, un acide glucuronique est représenté en saumon, l'apoprotéine en bleu et la protéine en complexe avec l'acide glucuronique en vert (Rogowski et al. 2014).

Pour dégrader totalement l'arabinoxylane présenté en figure 14, il faut donc l'association de plusieurs enzymes afin d'augmenter l'accessibilité au substrat de chacune ce qui a pour conséquence une augmentation de l'hydrolyse totale du polysaccharide mettant en évidence une action synergique de certaines de ces enzymes.

5. Les carbohydate estérases (CE):

Certaines unités osidiques peuvent présenter des estérifications gênant l'activité des glycoside hydrolases. Pour y remédier, l'action d'acétylxylane estérases (CE1, 2, 3, 4, 5, 6, 7 et 12) (figure 14, flèches rouges) et des féruloyl-estérases (CE1) (figure 14, flèches grises) est indispensable dans le cas d'un arabinoxylane.

Les acétylxylane estérases sont réparties dans plusieurs familles. Les familles des CE1, 3, 5, 7 et 12 ont un repliement en ($\alpha/\beta/\alpha$)-sandwich, la famille des CE4 a un repliement en tonneau (β/α)₇ et la famille des CE2 a un repliement sans repliement caractéristique (Chávez, Bull, et Eyzaguirre 2006). Les acétylxylane estérases possèdent soit une triade catalytique (Ser-His-Asp ou Ser-His-Glu), soit une diade catalytique (Ser-His). Ces enzymes n'ont pas de spécificité pour une acétylation en position 2 ou 3 et réalisent généralement une double déacétylation où la deuxième réaction est plus rapide que la première (Biely 2012). Les acétylxylane estérases sont souvent associées à un CBM généralement de la famille 1 (Chávez, Bull, et Eyzaguirre 2006).

Les féruloylestérases sont classées dans la famille des CE1. Ces enzymes adoptent un repliement en ($\alpha/\beta/\alpha$)-sandwich (figure 22) (« CAZy - Home » 2016). Bien que les féruloylestérases soient toutes classées dans la même famille de CAZymes elles sont toutes différentes avec notamment une grande variabilité de taille, allant de 270 à 1050 résidus, et de spécificité.



Figure 22 : Structure d'une féruloylestérase. La structure illustrant une féruloylestérase appartenant à la famille GH115 (code pdb 1gkk) provenant de *Ruminiclostridium thermocellum*. Les hélices α sont représentées en rouge, les brins β en jaune et les boucles en vert.

IV. Les systèmes de dégradation de la biomasse végétale, des réservoirs naturels d'enzymes

A. Le microbiote intestinal humain

L'Homme est omnivore et lors de la consommation de végétaux nécessaires à sa santé, il ne possède pas le bagage enzymatique permettant de dégrader les polysaccharides des parois végétales. Pour l'aider, il possède un microbiote produisant des enzymes capables de les dégrader. Les principaux embranchements bactériens retrouvés dans le microbiote humain sont les *Bacteroidetes*, les *Firmicutes* et les *Actinobactéries*. Les deux premiers embranchements cités représentent 90% des bactéries du microbiote (Eckburg et al. 2005 ; Ley, Peterson, et Gordon 2006 ; Flint et al. 2008). La plupart des bactéries de l'intestin sont actives sur polysaccharides isolés. Très peu sont actives directement sur la paroi végétale (Flint et al. 2008). Les bactéries cellulolytiques (peu de *Bacteroidetes* et beaucoup de *Firmicutes*) se fixent sur les parois végétales et les dégradent (Walker, A. W. 2006). Les produits d'hydrolyse et les polysaccharides ainsi libérés sont ensuite dégradés par les autres bactéries de l'intestin (Walker, A. W. 2006).

Bacteroides thetaiotaomicron est la bactérie la plus étudiée du microbiote intestinal humain et appartient à l'embranchement le plus abondant de celui-ci (Xu et al. 2003 ; Bjursell, Martens, et Gordon 2006 ; Flint et al. 2008 ; Koropatkin, Cameron, et Martens 2012). C'est une bactérie Gram⁻ appartenant à l'embranchement des *Bacteroidetes*. C'est une des bactéries qui possède le plus de GH parmi les bactéries dont le génome a été séquencé. Le génome de *B. thetaiotaomicron* code 226 GH et 15 PL. 61% de ces GH sont dans le périplasme, la membrane externe ou sont extracellulaires. Ces bactéries sont réparties en différents endroits de l'intestin appelés niches et se nourrissent des glucanes de la paroi des cellules intestinales ainsi que des polysaccharides de notre alimentation. Cette bactérie est connue pour dégrader un grand nombre de polysaccharides ce qui fait qu'elle participe à la digestion d'une grande partie de notre alimentation (Xu et al. 2003 ; Bjursell, Martens, et Gordon 2006 ; Flint et al. 2008 ; Koropatkin, Cameron, et Martens 2012). Afin de dégrader les polysaccharides, *B. thetaiotaomicron* possède dans son génome des groupes de gènes. Le groupe de gènes le mieux caractérisé est le système Sus (*Starch utilisation system*), qui est un système de dégradation de l'amidon. Ce groupe de gènes contient 8 gènes *susR*, *A*, *B*, *C*, *D*, *E*, *F* et *G*. Les produits des gènes *susC*, *D*, *E*, *F* et *G* sont des protéines qui sont

situées sur la membrane externe de la bactérie. SusC est transmembranaire et les 4 autres sont exposées à l'extérieur et liées à la membrane externe par une ancre *N*-acyl-S-diacylglyceryl liée de façon covalente à une cystéine de l'extrémité N-terminale de la protéine (Bos, Robert, et Tommassen 2007). Ces 5 enzymes sont responsables de la reconnaissance spécifique de l'amidon et de sa dégradation en oligosaccharides. SusC assure le transport des oligosaccharides de l'extérieur de la cellule vers le périplasma (figure 23). Les protéines SusA et SusB sont situées dans le périplasma et sont responsables de la dégradation en monosaccharides et en disaccharides, des oligosaccharides qui sont entrés dans le périplasma. Les mono et disaccharides ainsi obtenus sont internalisés dans la cellule afin d'être utilisés dans le métabolisme de la bactérie (figure 23). Enfin SusR est un senseur/régulateur situé dans la membrane plasmique qui fixe les mono et disaccharides issus de la dégradation de l'amidon du côté du périplasma, ce qui provoque l'activation de la transcription des autres protéines Sus (figure 23) (D'Elia et Salyers 1996).

D'autres systèmes de ce type ont été retrouvés pour la dégradation d'autres polysaccharides, ils sont appelés systèmes Sus like (Martens et al. 2009 ; Martens et al. 2011 ; Bolam et Koropatkin 2012). Ces systèmes ont été trouvés dans des *Bacteroidetes* et sont actifs sur gomme guar (additif alimentaire principalement composée de galactomannane) (Valentine, Arnold, et Salyers 1992), sur cellulose (Xie et al. 2007) ou encore sur arabinane (Arnal 2014).

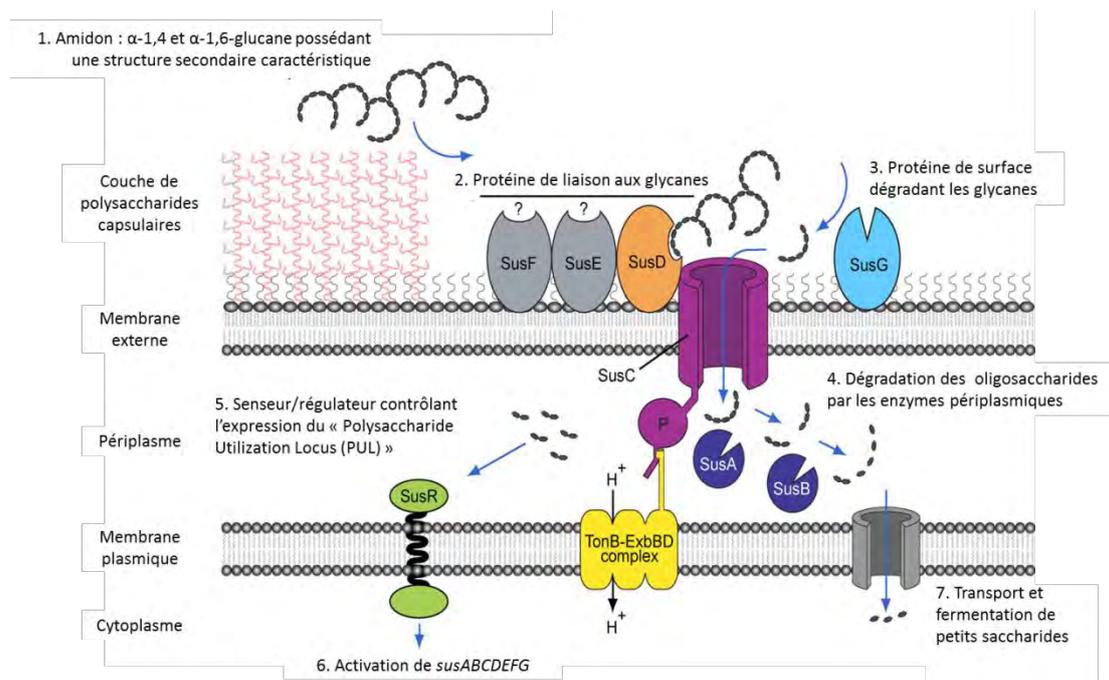


Figure 23 : Représentation schématique du système SUS (d'après Martens et al. 2009).

B. Le rumen bovin

Les mammifères herbivores ne sont pas capables de dégrader par leurs propres enzymes les polysaccharides des parois végétales issus de leur alimentation. Pour les dégrader ces mammifères possèdent un microbiote spécialisé dans le rumen (la panse) constituant un des quatre compartiments du système digestif des ruminants. Chez les bovins, de la même manière que chez les hommes, le microbiote est non fonctionnel à la naissance et se développe essentiellement pendant les 6 premiers mois de la vie du veau (Flint et al. 2008 ; Jami et al. 2013). Une fois le microbiote établi, on retrouve un large panel de microorganismes allant du procaryote à l'eucaryote complexe : archaea méthanogènes, bactéries, protozoaires et champignons. Les bactéries ruminales représentent environ 95% du microbiote du rumen (Kamra, D. N. 2005). Différents embranchements bactériens sont présents tels que les *Bacteroidetes*, les *Firmicutes* et les *Fibrobacters*. Les embranchements *Bacteroidetes* et *Firmicutes* représentent 90% des bactéries du rumen (Tajima et al. 1999 ; Flint et al. 2008). Les bactéries du rumen semblent plus efficaces contre les substrats celluloseux récalcitrants que leurs homologues du tractus digestif humain (Wedekind, Mansfield, et Montgomery 1988 ; Coen et Dehority 1970 ; Robert et Bernalier-Donadille 2003). De plus, il existe une synergie entre les bactéries cellulolytiques et les bactéries non cellulolytiques du rumen (Coen et Dehority 1970 ; Osborne et Dehority 1989 ; Flint et al. 2008). Comme dans l'intestin humain, les bactéries cellulolytiques (majoritairement des *Firmicutes*) se fixent sur les parois végétales intactes pour les dégrader. Des polysaccharides sont libérés et donc dégradés par les autres bactéries. Les produits d'hydrolyse des bactéries cellulolytiques sont aussi pris en charge par les autres bactéries du rumen (Bekele, Koike, et Kobayashi 2011).

La bactérie la plus étudiée du rumen est *Ruminococcus flavefaciens*. C'est une bactérie Gram⁺ appartenant à l'embranchement des *Firmicutes*. Cette bactérie est capable de dégrader la cellulose quelle que soit sa forme (amorphe ou cristalline) et fait partie des bactéries capables de dégrader la paroi végétale (Flint et al. 2008). Pour se fixer et dégrader la paroi, *R. flavefaciens* synthétise un complexe multienzymatique appelé cellulosome qui permet d'avoir une grande concentration de protéines sur la cellulose (Béguin et Lemaire 1996 ; Bayer et al. 1998 ; Bayer et al. 2004). Une partie des enzymes de ce complexe est organisée en groupe de gènes : le groupe de gènes *sca*. Le groupe de gènes *sca* est impliqué dans la dégradation de la cellulose et des hémicelluloses et code 5 protéines, des

scaffoldines, appelées ScaA, B, C et E ainsi que CttA. Toutes ces protéines sont liées par des systèmes dockérine/cohésine. L'ancrage du cellulosome à l'extérieur de la bactérie se fait par l'intermédiaire de ScaE qui se fixe de façon covalente à la matrice extracellulaire. Ensuite CttA, une petite protéine non catalytique, se fixe au domaine cohésine de ScaE. CttA possède des CBM qui lui permettent de se lier à la cellulose cristalline, ce qui fait que l'association ScaE-CttA permet le rapprochement de la bactérie et du substrat à dégrader. ScaB peut, elle aussi, se fixer par son domaine dockérine, présent à l'extrémité C-terminale, à la cohésine de ScaE. ScaB se fixe par son domaine cohésine sur la dockérine de la partie C-terminale de ScaA. ScaA et B sont deux grosses enzymes glycosylées portant plusieurs modules cohésine. Enfin ScaC, qui se fixe aussi sur ScaA par une cohésine, est une petite protéine adaptatrice. Des enzymes capables de dégrader la cellulose et/ou l'hémicellulose se fixent sur ScaA (figure 24) (Bayer et al. 1998 ; Bayer et al. 2004 ; Flint et al. 2008).

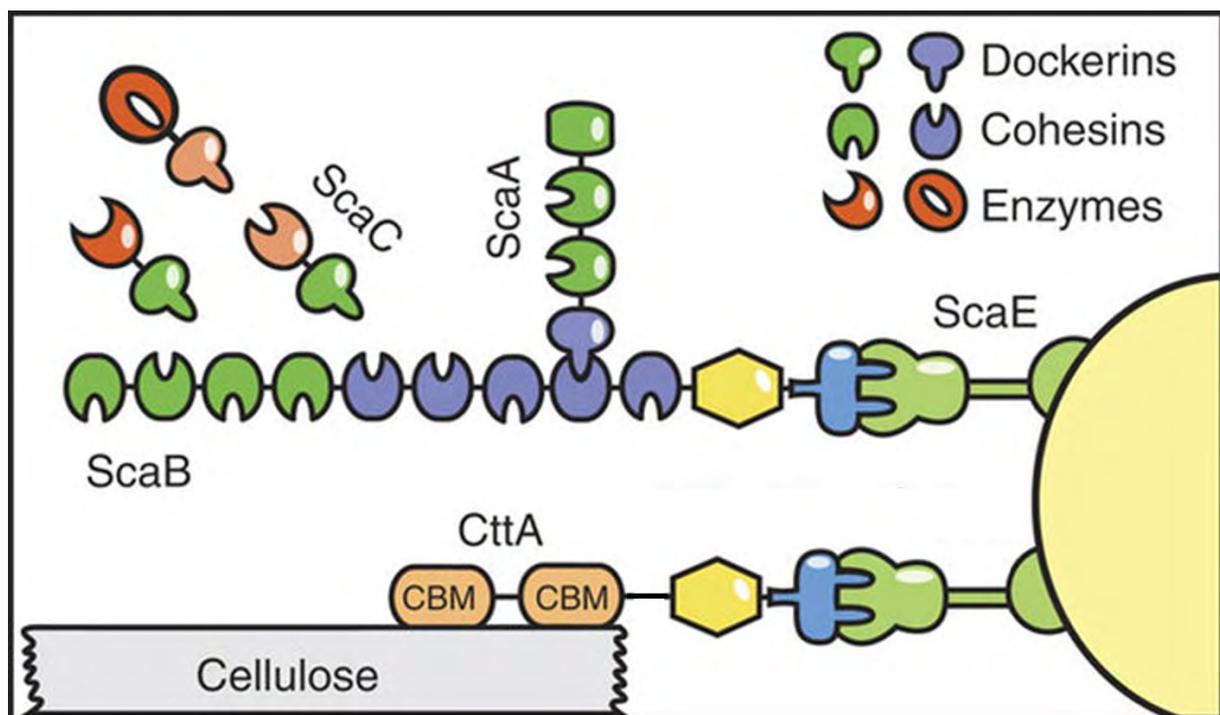


Figure 24 : Représentation schématique d'un cellulosome (Schoeler et al. 2014).

Dans le microbiote du rumen bovin d'autres microorganismes ont des fonctions précises, c'est le cas des archaea méthanogènes qui piègent les molécules de dihydrogène générées lors de la fermentation du rumen afin d'avoir un processus de fermentation continu (Kamra 2005).

Les protozoaires sont capables de dégrader principalement l'amidon, les pectines et des sucres solubles. Ces microorganismes présentent peu de cellulases et d'hémicellulases. Les protozoaires sont aussi impliqués dans le maintien du pH du rumen ainsi que dans la méthanogenèse (Kamra 2005).

Les champignons anaérobies du rumen ont longtemps été considérés comme étant des protozoaires. C'est seulement au milieu des années 1970 qu'il a été prouvé que ce sont bien des champignons appartenant à l'embranchement des *Neocallimastigomycota* (Orpin 1975 ; Davies et al. 1993). Il a été montré que ces champignons jouent un rôle majeur dans la dégradation des polysaccharides lignocellulosiques récalcitrants grâce à la production d'enzymes s'organisant en cellulosomes-like (Krause et al. 2003 ; Gruninger et al. 2014).

C. Les champignons

Les champignons utilisent 2 grands types de processus pour dégrader la biomasse végétale : le premier consiste en la production d'enzymes dans le milieu extracellulaire qui vont dépolymériser la biomasse comme par exemple des cellulases et cellobiohydrolases et le deuxième consiste en la génération d'espèces radicalaires oxydatives qui vont directement agir sur la biomasse (Cragg et al. 2015).

Il existe 3 grands types de champignons :

- Les pourritures blanches qui dégradent essentiellement la lignine mais aussi un peu la cellulose et les hémicelluloses. Ce sont des basidiomycètes comme *Phanerochaete chrysosporium*, *Schizophyllum commune* et *Trametes versicolor*. Ces pourritures blanches attaquent la lignocellulose en sécrétant de grandes quantités de cellulases ainsi que des enzymes qui vont oxyder la lignine (ligninases, manganèse peroxydases, peroxydases polyvalentes ou laccases). Certaines pourritures blanches produisent aussi de grandes quantités de LPMO (Dashtban et al. 2010 ; Cragg et al. 2015).
- Les pourritures brunes qui dégradent uniquement la cellulose et l'hémicellulose. Ce sont les basidiomycètes comme *Gloeophyllum trabeum*, *Serpula lacrymans* et *Postia placenta*. Ces pourritures brunes, qui sont issues de différentes évolutions des pourritures blanches, ne possèdent pas d'enzymes dégradant les lignines. Ces pourritures dégradent la biomasse végétale par réaction radicalaire de Fenton médiée par un chélateur. En présence de peroxyde d'hydrogène, un ion Fe(II) est oxydé en ion Fe(III) avec formation d'un anion hydroxyle et d'un radical hydroxyle. A

partir de ce radical une réaction radicalaire va se mettre en place et va permettre la dépolymérisation de la biomasse végétale sans passer par des enzymes (Arantes et Milagres 2006 ; Dashtban et al. 2010 ; Cragg et al. 2015).

- Enfin, les ascomycètes sont capables de tout dégrader (cellulose, hémicellulose et lignine) avec une grande efficacité car ils possèdent un arsenal d'enzymes pour chacune de ces grandes familles de polysaccharides comme c'est le cas pour *Aspergillus niger*, *Neurospora crassa* ou *Podospora anserina* (Karnaouri et al. 2014 ; Couturier et al. 2016). Ces champignons sécrètent une grande quantité de cellulases appartenant majoritairement aux familles de GH3, 5 et 7, des hémicellulases appartenant principalement aux familles de GH43, 3 et 11 ainsi que des peroxydases et des oxydases pour dégrader la lignine. Dans le sécrétome de ces champignons on retrouve diverses AA, dont des LPMO (AA9) et des estérases (Karnaouri et al. 2014 ; Cragg et al. 2015). *Aspergillus* et *Trichoderma* sont deux genres de champignons qui sécrètent à haute concentration de très grandes variétés d'hémicellulases qui agissent en synergie et ainsi dégradent, en mono ou disaccharides, la totalité des polysaccharides qui les entourent (Shallom et Shoham 2003).

D. Les vers de terre

Les vers de terre sont des organismes qui possèdent dans leur génome quelques séquences codant des enzymes capables de dégrader la biomasse végétale notamment des cellulases. Toutefois, elles sont peu nombreuses et ne suffisent pas à dégrader la lignocellulose, c'est pourquoi les vers ont aussi besoin de microorganismes pour dégrader totalement la biomasse végétale (Fujii, Ikeda, et Yoshida 2012). Le microbiote des vers est pour 50 % aérobie et 50 % anaérobie et le tube digestif du ver est en grande partie anaérobie. De plus, il a été montré que certains microorganismes aérobies se multiplient durant leur passage dans le ver (Krišrtúfek, Ravasz, et Pižl 1992 ; Parthasarathi et al. 2007). Dans l'étude de Fujii et collaborateurs (Fujii, Ikeda, et Yoshida 2012), deux espèces de vers ont été placées en milieu stérile pendant cinq jours. Ensuite, aucun microorganisme n'a pu être isolé à partir des vers. Cela prouve que les vers de terre n'ont pas de microbiote propre et donc que les organismes qui les aident à dégrader la lignocellulose proviennent de leur alimentation, c'est-à-dire de la terre. Dans cette même étude, Fujii et collaborateurs ont prouvé que les vers peuvent digérer la cellulose amorphe par leurs cellulases endogènes et

qu'ils ont besoin de cellulases exogènes (de microorganismes) pour digérer la cellulose cristalline (figure 25A). Les produits d'hydrolyse de la biomasse végétale sont ensuite consommés par le ver comme source d'énergie et comme substrats fermentescibles par les microorganismes anaérobies. Pendant le même temps, les microorganismes cellulolytiques aérobies sont inactivés (figure 25B). Enfin les microorganismes anaérobies sont digérés par le ver pendant que les microorganismes cellulolytiques aérobie sont excrétés (figure 25C) (Fujii, Ikeda, et Yoshida 2012).

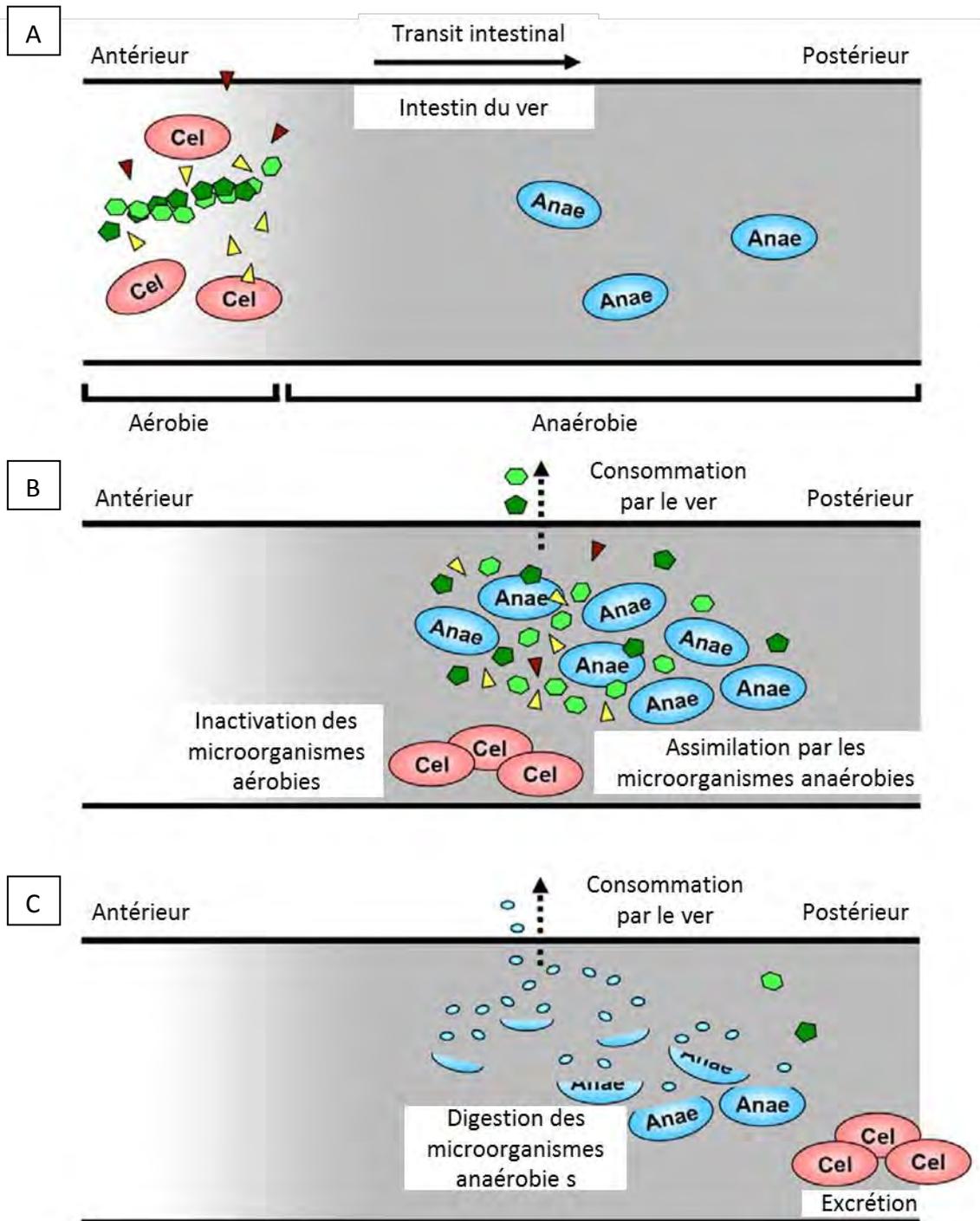


Figure 25 : Représentation schématique de la dégradation de la biomasse végétale dans le tractus digestif d'un ver (hypothèse proposée par Fujii, Ikeda, et Yoshida 2012). Les hexagones et pentagones représentent le glucose et le xylose de la biomasse lignocellulosique respectivement, les triangles représentent des enzymes lignocellulolytiques. Microorganismes cellulolytiques aérobie s (Cel) et microorganismes anaérobies (Anaë).

E. Les termites

Enfin, un autre système de dégradation efficace de la biomasse végétale est le termite. Il existe, en fait, 2 grands types de termites : les termites dits inférieurs et ceux dits supérieurs. Les termites inférieurs possèdent dans leur tube digestif des protistes flagellés contrairement aux termites supérieurs. Les termites supérieurs possèdent, quant à eux, un tube digestif plus long et très compartimenté (Y. Abe, Bignell, et Higashi 2014). Les termites ont été divisés en quatre grands groupes en fonction de leur alimentation (classification de Donovan) (Donovan, Eggleton, et Bignell 2001):

- Le groupe I ne comprend que les termites inférieurs, ceux-ci se nourrissent principalement de bois mort et d'herbe.
- Le groupe II comprend des termites dont l'alimentation est variée (bois mort, lichen, herbe, feuillage).
- Le groupe III comprend les termites se nourrissant de matière organique se situant à la surface de la litière.
- Enfin, le groupe IV comprend les termites se nourrissant de la couche inférieure, minéralisée, de la litière.

Tous les termites possèdent un microbiote composé de bactéries et d'archaea.

Le microbiote des termites inférieurs contient, en plus des bactéries et des archaea, des flagellés cellulolytiques permettant la dégradation de la cellulose et des xylanes principalement. L'embranchement bactérien le plus retrouvé dans ce microbiote est celui des *Spirochaetes* (Brune 2014). De plus, les termites inférieurs possèdent quelques endo- β -1,4-glucanases et β -glucosidases endogènes dans leurs glandes salivaires pour commencer la dégradation de la cellulose amorphe. Le reste de l'activité cellulasique provient des enzymes du microbiote intestinal et notamment des flagellés (Ohkuma 2003 ; Ni et Tokuda 2013). De la même façon, des hémicellulases endogènes sont retrouvées dans l'intestin des termites inférieurs mais la grande majorité des hémicellulases sont produites par leur microbiote (Ni et Tokuda 2013).

Chez les termites supérieurs, les embranchements les plus retrouvés sont les *Fibrobacteres*, les *Spirochaetes*, les *Bacteroidetes* et les *Firmicutes* (Hongoh et al. 2005 ; He et al. 2013). Les termites supérieurs sont divisés en trois groupes en fonction de leur alimentation. Les termites se nourrissant de bois (retrouvés dans les groupes II et III de la classification de Donovan), ceux se nourrissant d'humus (retrouvés dans les groupes II, III et IV de la

classification de Donovan) et les termites champignonnistes (retrouvés dans le groupe II de la classification de Donovan) (Brune 2014). On compare le système digestif de ces termites à un bioréacteur d'une grande efficacité. Les termites supérieurs ne possèdent que des β -glucosidases endogènes dans leurs glandes salivaires. Les endo- β -1,4-glucanases endogènes sont sécrétées directement dans la première partie du tube digestif et vont permettre, comme pour les termites inférieurs, une première dégradation de la cellulose amorphe. Des enzymes d'origine bactérienne à activité cellulolytique sont aussi retrouvées dans leur tube digestif (Ni et Tokuda 2013). Il n'a été rapporté aucune activité hémicellulase endogène chez les termites se nourrissant de bois et d'humus. En revanche les termites champignonnistes possèdent des hémicellulases endogènes et produites par les champignons. Chez les termites supérieurs, la totalité ou la grande majorité (pour les termites champignonnistes) des hémicellulases sont bactériennes (Ni et Tokuda 2013). Chez tous les termites, la lignine n'est pas dégradée mais des changements structuraux et des dépolymérisations interviennent. La lignine est seulement modifiée chimiquement (avec par exemple des décarboxylations et des oxydations) au cours de son passage dans le tube digestif des termites car seules des enzymes modifiant la lignine y sont présentes (Ohkuma 2003 ; Ni et Tokuda 2013).

V. L'exploration de la diversité

Comme mentionné dans la partie précédente, il existe de nombreux réservoirs naturels d'enzymes présentant une grande diversité d'activités. Dans le cadre de la recherche de nouveaux biocatalyseurs, notamment dans des organismes connus pour dégrader très efficacement la biomasse végétale, plusieurs techniques permettant l'exploration de leur métagénome peuvent être mises en place.

A. La métagénomique fonctionnelle

Dans l'équipe Catalyse et Ingénierie Moléculaire Enzymatiques (CIMEs) du LISBP, la métagénomique fonctionnelle est effectuée avec un criblage fonctionnel à haut débit. L'idée ici, est de prélever l'ADN total d'un écosystème dédié à la dégradation de la lignocellulose. Cet ADN est ensuite purifié puis fragmenté. Les fragments d'ADN sont ensuite purifiés par

électrophorèse en champ pulsé. Après cette étape, des fragments de grande taille sont récupérés et sont sous-clonés dans des fosmides. Des bactéries *E. coli* sont transformées avec les fosmides ce qui permet la création de banques métagénomiques (figure 26A) (Tasse et al. 2010 ; Bastien-Uluis 2012 ; Heux et al. 2015).

Une fois créées, les banques sont criblées directement avec des substrats chromogéniques. Dans l'équipe CIMEs, les substrats utilisés sont des substrats indolylés synthétisés par le Dr. Régis Fauré ou des AZO- et AZCl-polysaccharides commerciaux. Les substrats indolylés vont mimer des di ou tri-saccharides et être internalisés dans les bactéries, si la colonie possède l'activité pour dégrader le substrat celle-ci va se colorer (figure 26B). Si des colonies ont des activités pour dégrader les AZO- et AZCl-polysaccharides, il y aura apparition d'un halo coloré autour des colonies (figure 26C). Les clones positifs sont ensuite séquencés et annotés (Tasse et al. 2010 ; Bastien-Uluis 2012 ; Heux et al. 2015).

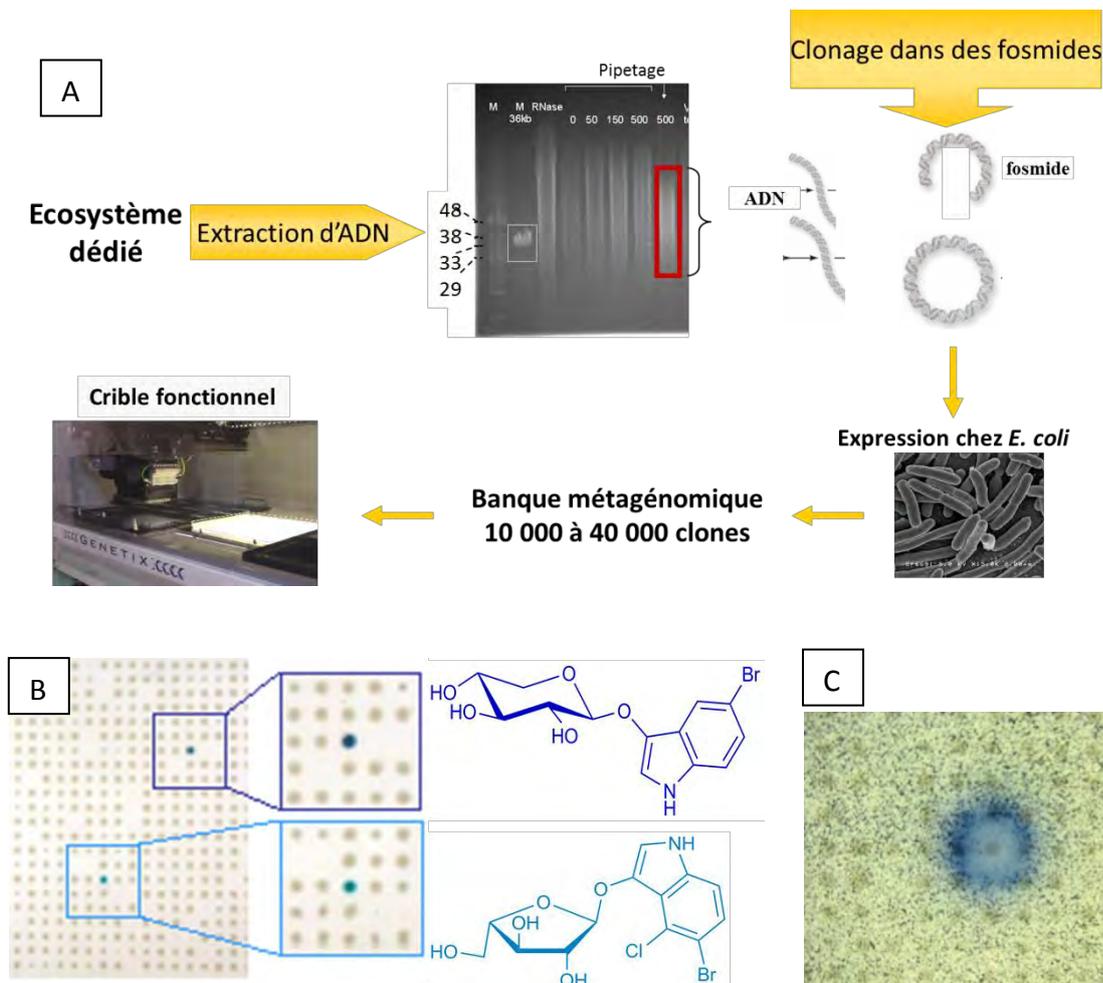


Figure 26 : A. Schéma général des étapes de la métagénomique fonctionnelle. B. Crible fonctionnel d'une banque métagénomique par des substrats indolylés : le 5-bromo-3-indolyl β-D-xylopyranoside en bleu foncé et le 5-bromo-4-chloro-3-indolyl-α-L-arabinofuranoside en bleu clair.

C. Crible fonctionnel d'une banque métagénomique par des AZO- ou AZCI-polysaccharides : apparition d'un halo bleu autour d'un clone positif (Bastien-Uluis 2012).

Cette méthode permet de cribler rapidement un grand nombre de substrats et a permis de découvrir des enzymes fonctionnelles sur un ou plusieurs des substrats testés. Dans un deuxième temps, l'annotation permet d'identifier un très grand nombre d'enzymes putatives qui peuvent être ensuite étudiées (Li et al. 2011 ; Bastien et al. 2013 ; Heux et al. 2015).

Les inconvénients de cette méthode sont qu'il est possible de négliger certaines activités qui ne sont pas criblées au départ car « on ne trouve que ce que l'on cherche » (Heux et al. 2015). D'autre part, comme les banques sont créées dans des bactéries, certaines enzymes eucaryotes peuvent ne pas être produites ou être non fonctionnelles du fait de l'absence de modifications post-traductionnelles, dès lors elles ne seront pas identifiées lors du crible fonctionnel. Enfin, il est parfois difficile de parvenir à une annotation taxonomique poussée et donc de déterminer de quels organismes proviennent les enzymes identifiées.

B. La métagénomique

La métagénomique permet l'étude d'un écosystème entier. Les étapes sont les mêmes que pour la métagénomique fonctionnelle sauf la dernière. En effet une fois les banques créées il n'y a pas d'étape de criblage fonctionnel mais directement un séquençage de celles-ci. Il peut y avoir alors une annotation fonctionnelle des gènes codant des enzymes présents dans le métagénome et une annotation taxonomique (Thomas, Gilbert, et Meyer 2012).

C. La métagénomique 16S

Cette méthode permet l'étude taxonomique d'un écosystème entier. Il ne s'agit pas de prélever l'ADN total mais les ARN ribosomaux 16S. Ces ARN possèdent des régions variables entre taxons mais conservées au sein d'un taxon. Ces régions sont encadrées par des régions flanquantes très conservées qui ont permis de dessiner des amorces générales pour la PCR (Claesson et al. 2009). Cette méthode permet de cribler à haut débit l'écosystème choisi. L'ADN total est extrait. Des PCR sont réalisées avec les amorces citées ci-dessus afin de n'amplifier que les ADN codant les ARNr 16S, appelés ADNr. Ces ADNr sont alors purifiés sur gel et les bandes d'intérêt sont excisées pour générer une banque d'amplicon de taille

voulue. Les ADNr sont quantifiés par qPCR et enfin séquencés par séquençage 454 ou illumina. Comme les régions choisies pour créer les amorces sont spécifiques d'un taxon il est donc possible, après le séquençage, de faire des alignements de la séquence d'intérêt contre celles déposées dans différentes bases de données et à partir des résultats de réaliser des annotations taxonomiques (figure 27) (Hall et al. 2008 ; Schloss et al. 2009 ; Engelbrekton et al. 2010 ; Kozich et al. 2013 ; Razzauti et al. 2015).

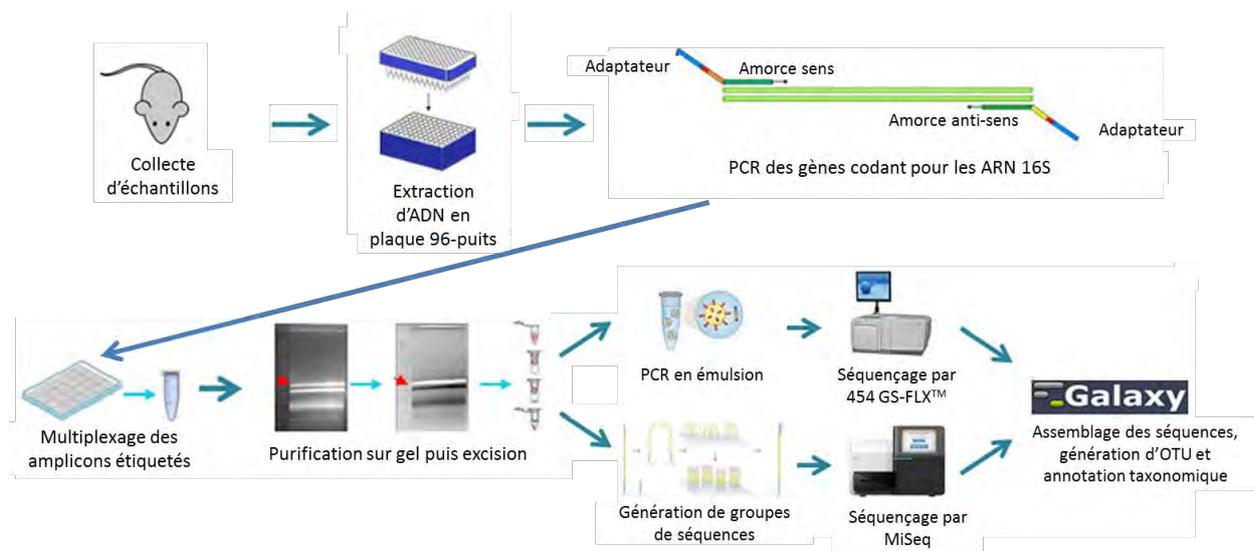


Figure 27 : Présentation des différentes étapes de la métagénomique 16S (Razzauti et al. 2015).

La différence majeure par rapport à la métagénomique fonctionnelle est qu'on ne peut pas réaliser de recherche fonctionnelle.

Par contre, grâce à l'annotation taxonomique, il est possible d'identifier les organismes et donc d'identifier, par exemple, les bactéries impliquées dans la dégradation de la paroi végétale. Une fois ces organismes identifiés, il est possible d'aller explorer plus en détail leur génome.

Un des avantages majeurs de cette technique est qu'elle permet de voir la diversité au sein de chaque population (Razzauti et al. 2015).

D. Le RNA seq ou « transcriptome sequencing »

Cette méthode permet d'analyser tous les ARNm d'une cellule. Le transcriptome d'une cellule est dynamique. L'analyse d'un transcriptome reflète donc l'état d'une cellule à un moment donné. La première étape du RNA seq consiste en la sélection d'un écosystème à

étudier. Ensuite, les ARN totaux sont extraits puis rétrotranscrits en ADNc. Ensuite, les ADNc sont amplifiés par PCR en utilisant des amorces qui se fixent au hasard sur les ADNc. Les ADNc sont ensuite séquencés puis les séquences sont nettoyées et assemblées en contigs. Enfin des alignements de séquences contre des banques de données sont effectués afin de faire de l'annotation taxonomique (figure 28) (Razzauti et al. 2015).

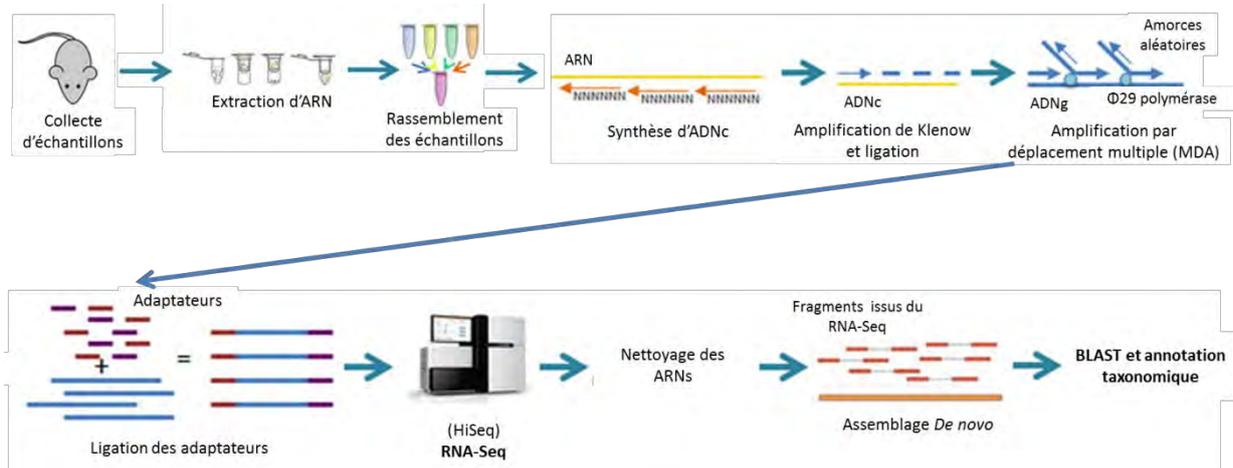


Figure 28 : Présentation des différentes étapes du RNA seq (Razzauti et al. 2015).

Comme dans le cas de la métagénomique 16S, il n'est pas possible de faire de recherche fonctionnelle avec cette technique. Par contre, le RNA seq permet d'identifier les gènes effectivement transcrits et par conséquent, les protéines réellement produites dans la cellule (Razzauti et al. 2015).

Comme la métagénomique 16S, le RNA seq permet de faire des annotations taxonomiques très poussées jusqu'à l'espèce, ce qui permet l'identification des bactéries présentes (Razzauti et al. 2015).

E. Comparaison de ces méthodes

Toutes ces méthodes se font à haut débit, ce qui implique qu'il est possible d'avoir accès à un très grand nombre d'informations en peu de temps et, de façon générale, totalement automatisée.

La métagénomique 16S et le RNA seq apportent en général les mêmes informations taxonomiques. Toutefois avec le RNA seq, il est souvent possible d'aller plus loin dans l'annotation taxonomique (Razzauti et al. 2015).

Le RNA seq présente un biais par son amplification aléatoire qui implique que les génomes les plus grands seront plus souvent amplifiés (Razzauti et al. 2015).

La métagénomique fonctionnelle, quant à elle, ne permet pas de déterminer avec précision de quel organisme proviennent les séquences annotées. Par contre c'est la méthode la plus efficace pour identifier rapidement des activités d'intérêt.

La métagénomique permet d'identifier des protéines d'intérêt mais sans aucune certitude de leurs réelles activités car contrairement à la métagénomique fonctionnelle aucun crible fonctionnel n'est effectué. L'activité supposée des séquences protéiques, uniquement basées sur l'annotation des génomes et donc sur l'existence de séquences homologues déposées dans des bases de données, est donc hypothétique alors que par métagénomique fonctionnelle l'activité du clone qui a permis d'identifier la protéine permet de penser que celle-ci est active. La métagénomique fonctionnelle permet également d'identifier d'autres protéines ayant une activité enzymatique différente de l'activité ciblée. En effet, de grands fragments d'ADN sont utilisés pour créer les banques ce qui permet d'obtenir de nombreux gènes flanquant le gène responsable de l'activité ciblée certains pouvant se révéler intéressants par leur originalité ou par leur action synergique avec la protéine d'intérêt par exemple.

VI. Projet de thèse

Dans le cadre de la recherche de nouveaux biocatalyseurs, l'équipe Catalyse et Ingénierie Moléculaire Enzymatique et plus particulièrement le groupe Lignocellulase s'est intéressé à deux organismes connus pour dégrader la biomasse végétale : le termite et le ver de terre. Pour explorer le métagénome de ces organismes, l'équipe a utilisé la technique de métagénomique fonctionnelle. Trois banques issues des trois espèces différentes de termites *Pseudacanthotermes militaris*, *Nasutitermes corniger* et *Termes hispaniolae* et une banque issue du ver de terre *Pontoscolex corethrurus* ont été créées

Ma thèse s'inscrit dans la continuité de ce travail consistant en l'annotation fonctionnelle et taxonomique de la totalité de la banque métagénomique du termite *Termes hispaniolae* et d'une partie de celle du termite *Nasutitermes corniger*. Ensuite, j'ai effectué une

comparaison du potentiel hémicellulolytique de ces deux termites et ce travail fait l'objet du premier chapitre du présent manuscrit.

D'autre part à l'issue de l'annotation fonctionnelle de la banque métagénomique du ver de terre *Pontoscolex corethrurus*, une enzyme a suscité notre intérêt, il s'agit de GH* une enzyme annotée comme étant une glycoside hydrolase putative mais n'appartenant à aucune famille de glycoside hydrolases connue à ce jour. Une étude fonctionnelle et structurale a donc été menée afin de valider sa fonction et de la caractériser. L'étude fonctionnelle de GH* est décrite dans le chapitre II de mon manuscrit de thèse. L'étude structurale est décrite dans le chapitre III.

Cette thèse a été effectuée en collaboration entre deux laboratoires de Toulouse : le Laboratoire d'Ingénierie de Systèmes Biologiques et des Procédés dans l'équipe Catalyse et Ingénierie Moléculaire Enzymatique, groupe Lignocellulase, pour la partie « métagénomique et caractérisation fonctionnelle » et l'Institut de Pharmacologie et de Biologie Structurale dans l'équipe de Biophysique Structurale pour la partie « caractérisation structurale » de cette thèse.

Matériels et méthodes

I. Annotation des banques métagénomiques de termites

A. Assemblages puis sélection des contigs

Le but de cette étape est de reconstruire les fragments d'ADN clonés dans *E. coli* lors de la préparation de la banque métagénomique. Les séquences obtenues par pyroséquençage 454 (Medini et al. 2008) ont été analysées à l'aide d'une suite automatisée de logiciels bioinformatiques développée précédemment par Sandrine Laguerre au laboratoire.

Cette suite consiste à reconstruire les séquences (avec les programmes Cap3 et StatAssemblage), nettoyer la séquence du vecteur (programme CreateContigFileVM_VN), conduisant à des contigs de longueur comprise entre 1000 et 50000pb et qui serviront à la détection des cadres ouverts de lecture (ORF).

B. Recherche des cadres ouverts de lecture

Les cadres ouverts de lecture (ORFs) ont été recherchés dans les séquences nucléiques et protéiques par l'intermédiaire du programme MetaGene (Noguchi, Park, et Takagi 2006). Ensuite les contigs et les ORFs obtenus ont été analysés par alignements de séquences contre les banques nr (non-redundant) et Swissprot (Altschul et al. 1990).

C. Annotation fonctionnelle et taxonomique

L'annotation fonctionnelle des séquences protéiques correspondant aux ORFs identifiées a été réalisée par B. Henrissat et E. Drula de l'équipe CAZy du laboratoire Architecture et Fonction des Macromolécules Biologiques (AFMB) CNRS – Université Aix Marseille UMR7257.

Des BLASTs contre les COGs (Clusters of Orthologous Groups of proteins) ont été effectués afin d'obtenir des informations quant à la fonction générale des protéines obtenues (Tatusov, Koonin, et Lipman 1997 ; Meereis et Kaufmann 2008).

L'annotation taxonomique a été faite à l'aide du programme MEGAN (Min support = 1, Min score = 35,0, Top percent = 10,0 et Min complexity = 0,3) (Mitra, Klar, et Huson 2009).

II. Etude structurale et fonctionnelle de GH*

A. Sous clonage

GH* a été identifiée dans une banque métagénomique de ver de terre. Les banques métagénomiques ont été créées dans des cellules *E. coli* EPI 100 (tableau 4) et les fragments d'ADN, du fait de leur grande taille, sont dans des fosmides pCC1FOS.

	Nom	Génotype	Caractéristiques	Résistance
Vecteurs	pCC1FOS	Origine du facteur F d' <i>E. coli</i> , oriV, sites cos pour l'emballage par le phage lambda, Cm ^R	Fosmide, réplication monocopie, inducible multicopie	Chloramphénicol (Cm)
	pET28a	Promoteur T7, Tag-histidine, site de restriction thrombine, Multi-site de clonage, <i>lacI</i> , Km ^R , origine pBR322 et <i>fi</i>	Plasmide d'expression	Kanamycine (Km)
Souches	EPI100	[<i>F-mrcA Δ(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 recA 1 endA 1 araD139 Δ(ara, leu) 7697 galU galK λ-rpsL nupG tonA</i>]	Résistance phage T1	
	DH5α	<i>F-80dlacZ M15 (lacZYA-argF) U169 recA 1 endA 1 hsdR17 (rk- mk+) phoAsupE44 -thi-1 gyrA96 relA 1</i>	Souche optimisée pour la réplication de plasmide, permet la stabilité des inserts, utilisée en criblage blanc/bleur	
	BL21/DE3	<i>F-ompT hsdSb (r_B-m_B-) gal dcm λ(DE3)</i>	Dépourvue des protéases lon et OmpT, porte le prophage DE3	
	C41/DE3	<i>F-ompT hsdSb (r_B-m_B-) gal dcm λ(DE3)</i>	Souche dérivée de BL21/DE3, contient une mutation qui prévient la mort cellulaire associée à l'expression de protéines recombinantes toxiques	
	Tuner/DE3	<i>F-ompT hsdSb (r_B-m_B-) gal dcm lacY1 λ(DE3)</i>	Souche dérivée de BL21/DE3, porte en plus une mutation de la Lac perméase conférant un meilleur contrôle de l'expression	

Tableau 4 : Caractéristiques des vecteurs et souches utilisés au cours de ce travail.

La séquence codant GH* sans le peptide signal a été extraite à partir du fosmide par PCR. Pour ce faire, les cellules *E. coli* EPI 100, contenant le fosmide pCC1FOS sur lequel la séquence codant GH* est présente, ont été mises en préculture dans du milieu LB supplémenté en chloramphénicol à 12,5 µg/ml, une nuit à 37 °C sous agitation (160 rpm). Le fosmide pCC1FOS a ensuite été extrait et purifié en utilisant un kit de midiprep (nucleobond PC100 de Macherey Nagel).

Préparation des amorces pour amplification par PCR :

La présence d'un peptide signal a été prédite à l'aide du serveur SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>). Les amorces ont été dessinées en insérant les sites de restriction (soulignés) NdeI et BamHI sur les amorces sens (Fo) et anti-sens (Re) respectivement, ainsi qu'une séquence correspondant au vecteur de destination (située en amont du site de restriction). La séquence codant GH* a ensuite été insérée dans un plasmide pET28a afin d'obtenir la protéine étiquetée hexahistidines en N-terminal :

GHStar_NdeI-Fo_CGCGCGGCAGCCATATGGCTGAGCCGCCGCTGACG

GHstar_BamHI-Re_GCTCGAATTCGGATCCGTTTGGGCTGCTGATGACGCTG.

Séquence nucléique de GH * (sans le peptide signal) :

```
GCTGAGCCGCCGCTGACGGCTGCCATTATGCGCAGCAGCTTGGCGTCGGGATGGATGTGGACTGGGCGCGA
ACCGAGCGCGGCATACGCGAATTCGACCCGCTGGTGGTCCGCGATTTTCGTGCAAAAGGCATTACTCACGTGC
GCATCCGCGTGGCCGATGAGCCGACGGAAGCGCGGCTGATTACCTGCGCAAGCTGGTGAAGCCTGCGAGC
AGTATGGCGTGATCCCGATAATCGCTTATCAGGCTGATGAATATAAAAACGACCCAAAAGCCGATAACGAAAT
AGAGACGATCAACTGGTGGATTGCCGTTGCGCACTATTTGGTCAGAGCTACCCGCTGCTGGGTTTTGATCTCA
TTTATGAGCCAGCCGACAAGCTTAACCACAACCTCGCGTCGCTAAACCGCGTGATGAAAAACCATTAAGC
GATCCACGACATGGACCCGACGCGCATGATCTTTATTGCGCCGCGCCTGCGTGCCGCGCCGGAGATTATCC
AGCCTGAAGTGCTGCGCACAGCCAGAACTACCTGCTGGCGGAGTGGCATATTTTCCCGTGGGGGCGCTGA
AAAATAGCGGTAAATACCCGTGGACGTCCGGCACGGCGCGGAGAAAGCGGCTATCCATAACCGCATCAACG
CCGCGCTGCACTGGCAGCAAAAACCGGGCACGTTAGCTGGGTGGGGGCTGGGGCGTGGGAGAATCAAGC
CGCTTTACGCAACCGCTTCGAGCTGGCCTTCGCCACCTTTATGGCCTGCGAGCTGCAAAAGGCGAAAATCCC
GTATGCGCTTAACGCGGATTTCCAGTTTTACGACGGAGAAGAGGGCGCCTGGCGGCCAGCGCCAGAGCCGTT
ATTGCAGGCGATGATCGCCCCTGTCTGTGAAAAGCCCGGCGATAAGCCGGGCCATCATGCGGTTAAACCGGCT
GCTCGTGATGCGGGACACGCGACGCCAGCGGCAGCCAGCACAGTAAAATCAGCAGCCCCATCAGCGTCATCA
GCAGGCCCAAACCTAG
```

Les conditions expérimentales de la PCR ayant fait l'objet d'optimisation, elles sont détaillées dans le chapitre II.

Afin de vérifier si la séquence d'ADN amplifiée est la bonne, le produit PCR a été déposé sur gel d'agarose à 0,8 %. La migration sur gel a été réalisée pendant 20 min à 200 V. Après révélation au BET, la lecture de l'intensité de fluorescence à 300 nm a été réalisée avec le lecteur de gel Geldoc (Biorad). La séquence d'ADN migre à la masse attendue de 1 kb.

Le sous-clonage a été réalisé par recombinaison, en utilisant le kit « In-Fusion » de Clontech, dans un pET28a possédant une cassette de résistance à la kanamycine (kana) et une séquence codant une étiquette 6-histidines en N-terminal de la séquence insérée.

Pour valider le sous-clonage, une digestion par les enzymes de restriction NdeI et BamHI a été réalisée, et analysée après migration sur un gel d'agarose 0,8 %. La taille totale, du vecteur contenant l'insert, attendue est de 6369 pb.

Des cellules DH5 α ont été transformées avec le pET28a contenant la séquence codant la GH*, étalées sur boîtes LB agar contenant de la kanamycine à 50 μ g/ml et ont été mise à pousser à 37 °C sur la nuit. Deux colonies ont été mises en pré-culture (LB + kana à 50 μ g/ml) à 37 °C sur la nuit sous agitation (160 rpm) puis des miniprep ont été effectuées pour extraire et purifier le plasmide. Ce dernier a été séquençé (GATC, Constance, Allemagne) pour vérifier l'insertion et la séquence.

La masse moléculaire, le coefficient d'extinction molaire et le point isoélectrique de la protéine ont été estimés à l'aide du serveur ProtParam (<http://web.expasy.org/protparam/>). La masse moléculaire (MM) est estimée à 40234,5 g/mol, le point isoélectrique (pI) à 6,88 et le coefficient d'extinction molaire (ϵ) à 73005 l.mol⁻¹.cm⁻¹.

Séquence protéique de GH* (avec l'étiquette hexahistidine His) :

```

      10      20      30      40      50      60
MGSSHHHHHH SSGLVPRGSH MAEPPLTAAH YAQQLGVGMD VDWARTEGI REFDPVVVRD

      70      80      90     100     110     120
FRAKGITHVR IRVADEPTEA RLIHLRKLVE ACEQYGVIP IAYQADEYKN DPKADNEIET

     130     140     150     160     170     180
INWWIAVAHY FGQSYPLLG FDLIYEPADKL NHNLASLNRV YEKTIKAIHD MDPQRMIFIA

     190     200     210     220     230     240
PRLRAAPEDL SSLKLPASQ NYLLAEWHIF PWGPLKNSGK YPWTSGTAAE KAAIHNRIINA

     250     260     270     280     290     300
ALHWQKQTGH VSWVGGWVG ESSRFTPTAS QLAFATFMAC ELQKAKIPYA LNADFQFYDG

     310     320     330     340     350     360
EEGAWRPAP EPLLQAMIAPV CEKPGDKPGH HAVKPAARDA GHATPAAAST VKSAAPSASS

```

AGPN*

Préparation d'amorces pour obtenir des mutants de GH* :

Afin de vérifier si les résidus identifiés comme pouvant être les résidus catalytiques jouent réellement un rôle dans l'activité de GH*, des mutants de l'enzyme ont été réalisés par mutagenèse dirigée en utilisant le kit « QuikChange II Site-Directed Mutagenesis Kit » (Agilent technologies) avec les amorces ci-dessous. Le mutant du glutamate 117 de GH* en alanine a été obtenu avec les amorces GH* E117A-Fo (sens) et GH* E117A-Re (anti-sens), le

mutant de la tyrosine 116 en alanine a été réalisé en utilisant les amorces GH* Y116A-Fo (sens) et GH* Y116A-Re (anti-sens) et enfin le mutant glutamate 117 en glutamine a été obtenu en utilisant les amorces GH* E117Q-Fo (sens) et GH* E117Q-Re (anti-sens). Le codon correspondant à l'alanine est surligné en jaune pour le mutant E117A et en vert pour le mutant Y116A et le codon correspondant à la glutamine est surligné en bleu pour le mutant E117Q.

Tous ces mutants ont été obtenus par Franck Moncassin pendant son stage de 2^{ème} année de master Biologie Structurale et Fonctionnelle (M2R BSF, Toulouse).

GH* E117A-Fo : GT TTT GAT CTC ATT TAT **GCC** CCA GCC GAC AAG CTT AAC C

GH* E117Q-Re : GTT AAG CTT GTC GGC TGG GGC ATA AAT GAG ATC AAA ACC

GH* Y116A-Fo: GT TTT GAT CTC ATT **GCG** GAG CCA GCC GAC AAG CTT AAC C

GH* Y116A-Re : GTT AAG CTT GTC GGC TGG CTC CGC AAT GAG ATC AAA ACC

GH* E117Q-Fo : CTG GGT TTT GAT CTC ATT TAT **CAA** CCA GCC GAC AAG CTT AAC C

GH* E117Q-Re : GTT AAG CTT GTC GGC TGG TTG ATA AAT GAG ATC AAA ACC CAG C

B. Recherche des meilleures conditions pour l'expression de GH*

Afin de trouver les meilleures conditions de surexpression de la protéine, des tests de production ont été effectués dans différentes souches bactériennes d'*E. coli* : BL21/DE3, C41/DE3 et Tuner/DE3 (tableau 4).

Les cultures ont étéensemencées au 1/100^{ième} de préculture et placées à 37 °C jusqu'à atteindre une densité optique à 600 nm comprise entre 0,4 et 0,6 (début de la phase de croissance exponentielle), à la suite de quoi diverses conditions ont été testées selon les souches. Différentes conditions d'induction ont été testées : 4 h à 37 °C ou 16 h à 16 °C avec ou sans induction à 1, 0,2 ou 0,1 mM IPTG pour BL21/DE3, Tuner/DE3 ou C41/DE3 respectivement.

Finalement, le meilleur taux de production de la protéine recombinante a été obtenu dans une souche BL21/DE3 avec une préculture de 10 ml à 37 °C sur la nuit sous agitation. 1 l de culture estensemencé au 1/100^{ième} suivi d'une induction à 16 °C sur la nuit avec 1 mM final d'IPTG. Préculture, culture et induction se font sous agitation (160 rpm), l'antibiotique

kanamycine est utilisée à 50 µg/ml. La culture est arrêtée par centrifugation 20 min à 7700 *g* à 4 °C et les culots sont repris dans 20 ml de tampon 20 mM HEPES, 100 mM NaCl, pH 7,0. Enfin les culots resuspendus dans ce tampon sont placés à -20 °C jusqu'à leur utilisation.

C. Purification

1. Chromatographie d'affinité sur ions métalliques immobilisés (IMAC)

Un culot bactérien est décongelé à température ambiante. Il est ensuite transféré dans un bécher en verre et placé dans la glace afin de mieux maîtriser la température de l'échantillon durant la lyse qui est effectuée par sonication. Les cellules sont lysées avec une amplitude de 25 %, un temps de pulse de 1 s active et 1 s inactive, pendant 2 min. L'opération est répétée 3 fois avec un temps d'attente d'au moins 2 min entre chaque cycle. Après la lyse, l'échantillon est centrifugé 10 min à 12800 *g* à 4 °C pour sédimenter les débris cellulaires et récupérer dans le surnageant le contenu cytoplasmique des bactéries (extrait cellulaire).

La 1^{ère} étape de purification consiste en une chromatographie d'affinité en « batch » à température ambiante. Les éluions se font par gravité en faisant passer l'échantillon à travers une résine déposée dans une petite colonne en plastique de 20 ml. Pour cela, 5 ml de résine cobalt Talon (Clonetech) sont déposés dans une colonne. L'éthanol est retiré par gravité et ainsi un volume d'environ 2 – 2,5 ml de résine est obtenu. La résine est lavée avec 8 à 10 volumes (20 ml) d'eau ultrapure puis équilibrée par 10 volumes de tampon Tris (20 mM Tris, 300 mM NaCl, pH 8,0). Le surnageant après lyse des cellules est chargé sur la résine et la fraction non retenue est collectée et conservée à 4 °C. La protéine recombinante est retenue sur la résine par le biais de son étiquette à 6-histidines. La résine est ensuite lavée avec 20 ml de tampon pour retirer tout ce qui n'est pas fixé sur la résine. La fraction de lavage est collectée et conservée à 4 °C. Un premier lavage avec 5ml de tampon contenant 5 mM d'imidazole est effectué puis un second lavage avec 5ml de tampon contenant 10 mM d'imidazole est réalisé deux fois afin d'éliminer tout ce qui s'est fixé de façon non spécifique sur la résine. Toutes les fractions de lavage avec de l'imidazole sont collectées puis conservées à 4 °C. L'éluion est réalisée par deux lavages successifs avec 5 ml de tampon contenant 100 mM d'imidazole. Ces deux fractions sont collectées et placées à 4 °C. Enfin la dernière étape est un lavage avec du tampon contenant 200 mM d'imidazole. Cette fraction

est aussi collectée puis gardée à 4 °C. La résine est lavée avec 20 ml d'eau ultrapure puis placée dans une solution d'éthanol à 20 % afin d'être régénérée.

Après optimisation la composition du tampon est modifiée pour devenir : 20 mM HEPES, 100 mM NaCl, pH 7,0 et la première étape de lavage avec de l'imidazole à 5 mM a été remplacée par une étape de lavage à 10 mM d'imidazole en conservant les mêmes volumes.

En fin de purification 5 µL de chaque fraction sont déposés sur gel de polyacrylamide en conditions dénaturantes afin de déterminer dans quelles fractions la protéine recombinante a été éluée et quel est son degré de pureté.

Après avoir ajouté 5 µl de tampon de charge 2 X (lithium dodécyl sulfate (LDS) 4 X + DTT + eau milliQ) aux 5 µl d'échantillons, les échantillons sont mis à chauffer 5 min à 95 °C. Les échantillons sont ensuite déposés sur gel précoulé « stain free » SDS-PAGE à 12 % d'acrylamide (Biorad). La migration se fait pendant 25 min à 200 V. Après migration, le gel est révélé avec un lecteur Gel doc (Biorad) en mesurant la fluorescence intrinsèque des protéines due à la présence de tryptophanes et de tyrosines dans la séquence protéique et à des composés trihalogénés présents dans les gels. Les gels sont exposés 5 min aux UV et l'intensité de fluorescence est ensuite mesurée.

2. Chromatographie d'exclusion-diffusion

Une 2^{ème} étape de purification est nécessaire pour obtenir une protéine pure afin de mener les essais fonctionnels et de cristallisation. Cette étape est une filtration sur gel sur colonne S200 préparative (Hiload 16/60 de GE Healthcare) dans un tampon 20 mM citrate de sodium, 100 mM NaCl, pH 5,0. Les fractions d'IMAC contenant la protéine recombinante sont concentrées sur Centricon (Amicon ultra) ayant un seuil de coupure de 10 kDa jusqu'à obtenir un volume d'environ 2 ml. Ensuite l'échantillon est injecté sur la colonne de chromatographie d'exclusion/diffusion. L'élution se fait sur système FPLC (AKTA) à un débit de 1 ml/min et les fractions de 1 ml sont automatiquement récupérées dès que la densité optique à 280 nm est supérieure à 50 mAU.

5 µl des différentes fractions d'élution sont gardés pour être déposés sur gel dénaturant afin de déterminer dans quelles fractions la protéine est éluée ainsi que sa pureté.

Après avoir ajouté 5µl de tampon de charge 2 X (LDS 4 X + agent réducteur + eau milliQ) à 5 µl d'échantillons, les échantillons subissent une dénaturation thermique à 95 °C pendant 5 min. Les échantillons sont ensuite déposés sur gel de polyacrylamide en conditions

dénaturantes. La migration se fait pendant 25 min à 200 V. Après migration la présence de protéine est révélée par mesure de la fluorescence intrinsèque des tryptophanes et tyrosines avec un lecteur Gel doc (Biorad).

D. Fluorimétrie différentielle à balayage

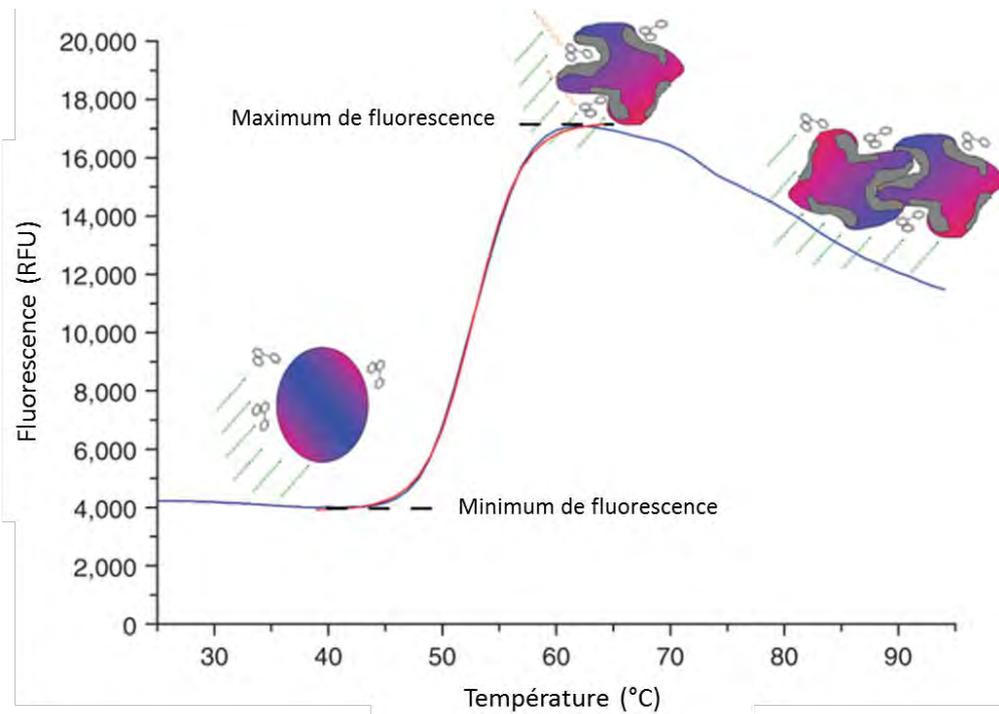


Figure 29 : Schéma expliquant les différentes étapes de la fluorimétrie différentielle à balayage (DSF) (Niesen, Berglund, et Vedadi 2007). Courbe de dénaturation d'une protéine : variation de fluorescence en fonction de la température.

Cette méthode, permettant de déterminer la température de fusion (T_m) d'une protéine, repose sur l'utilisation d'une sonde fluorescente, le Sypro-Orange (Invitrogene), dont la longueur d'onde d'absorption maximale est à 470 nm et la longueur d'onde d'émission maximale est à 570 nm. De plus, le Sypro-Orange possède la particularité d'avoir un rendement quantique de fluorescence faible en milieu hydrophile et élevé en milieu hydrophobe.

Au cours de la dénaturation, le cœur hydrophobe des protéines est exposé au solvant aqueux. Le Sypro-Orange entre alors en contact avec ces régions entraînant une augmentation de l'intensité de la fluorescence émise par cette sonde jusqu'à atteindre une intensité maximale où la protéine est dénaturée. Les zones hydrophobes interagissent alors

entre elles conduisant à un phénomène d'agrégation qui a pour conséquence de bloquer l'accès du Sypro-Orange à ces zones. L'intensité de fluorescence va alors décroître.

Le point d'inflexion de la courbe de fluorescence correspond au T_m de la molécule. La température de fusion (T_m), caractéristique d'une protéine dans un milieu donné est défini comme étant la température au point d'inflexion de la transition observée sur la courbe de dénaturation (Figure 29). Cette valeur caractéristique correspond à la température de demi-dénaturation. En pratique, la valeur du T_m est obtenue en calculant la dérivée de la courbe de fluorescence $-d(\text{RFU})/(dT)=f(T)$. La valeur de T_m correspond à la température au minimum de cette fonction dérivée.

Les expériences sont réalisées à l'aide d'un appareil de qPCR CFX96 (Biorad) en utilisant le mode FRET (longueurs d'ondes d'excitation de 450-490 nm et de détection de 560-580 nm).

Afin de savoir quel tampon est le plus adapté pour mener des essais de cristallisation de GH*, des expériences de DSF ont été réalisées en microplaques 96 puits. La température est progressivement augmentée par pas de 0,3 °C entre 25 °C et 100 °C avec une mesure de fluorescence faite à chaque incrément de température. Une première expérience a été menée pour savoir quelles concentrations de protéine et de Sypro-Orange utiliser. Les concentrations testées étaient 1, 2 et 5 µM d'enzyme et 2, 5 et 10 X de Sypro-Orange. Le meilleur signal a été obtenu pour 5 µM de GH* et 5 X Sypro.

A partir de là, les deux cribles pH/[sel] conçus dans l'équipe ont été utilisés. La gamme de pH va de 4,5 à 9,0 et la concentration en NaCl est de 0, 25, 150 ou 300 mM (tableau 5).

Des essais de DSF ont aussi été menés sur GH* sauvage et les mutants Y116, E117A et E117Q en présence ou non de Xylohexaose (X6) à 30 mM. L'enzyme est à 5 µM et les tests ont été effectués en triplicatas.

Crible 0-25	Crible 150-300
Sodium Acetate pH 4.5@100mM	Sodium Acetate pH 4.5@100mM, NaCl 150mM
Sodium Citrate pH 4.7@100mM	Sodium Citrate pH 4.7@100mM, NaCl 150mM
Sodium Acetate pH 5.0@100mM	Sodium Acetate pH 5.0@100mM, NaCl 150mM
Potassium Phosphate pH 5.0@100mM	Potassium Phosphate pH 5.0@100mM, NaCl 150mM
Sodium Phosphate pH 5.5@100mM	Sodium Phosphate pH 5.5@100mM, NaCl 150mM
Sodium Citrate pH 5.5@100mM	Sodium Citrate pH 5.5@100mM, NaCl 150mM
Sodium Acetate pH 4.5@100mM, NaCl 25mM	Sodium Acetate pH 4.5@100mM, NaCl 300mM
Sodium Citrate pH 4.7@100mM, NaCl 25mM	Sodium Citrate pH 4.7@100mM, NaCl 300mM
Sodium Acetate pH 5.0@100mM, NaCl 25mM	Sodium Acetate pH 5.0@100mM, NaCl 300mM
Potassium Phosphate pH 5.0@100mM, NaCl 25mM	Potassium Phosphate pH 5.0@100mM, NaCl 300mM
Sodium Phosphate pH 5.5@100mM, NaCl 25mM	Sodium Phosphate pH 5.5@100mM, NaCl 300mM
Sodium Citrate pH 5.5@100mM, NaCl 25mM	Sodium Citrate pH 5.5@100mM, NaCl 300mM
MES pH 5.8@100mM	MES pH 5.8@100mM, NaCl 150mM
Potassium Phosphate pH 6.0@100mM	Potassium Phosphate pH 6.0@100mM, NaCl 150mM
MES pH 6.2@100mM	MES pH 6.2@100mM, NaCl 150mM
Sodium Phosphate pH 6.5@100mM	Sodium Phosphate pH 6.5@100mM, NaCl 150mM
PIPES pH 6.5@100mM	PIPES pH 6.5@100mM, NaCl 150mM
MES pH 6.5@100mM	MES pH 6.5@100mM, NaCl 150mM
MES pH 5.8@100mM, NaCl 25mM	MES pH 5.8@100mM, NaCl 300mM
Potassium Phosphate pH 6.0@100mM, NaCl 25mM	Potassium Phosphate pH 6.0@100mM, NaCl 300mM
MES pH 6.2@100mM, NaCl 25mM	MES pH 6.2@100mM, NaCl 300mM
Sodium Phosphate pH 6.5@100mM, NaCl 25mM	Sodium Phosphate pH 6.5@100mM, NaCl NaCl 300mM
PIPES pH 6.5@100mM, NaCl 25mM	PIPES pH 6.5@100mM, NaCl NaCl 300mM
MES pH 6.5@100mM, NaCl 25mM	MES pH 6.5@100mM, NaCl NaCl 300mM
Potassium Phosphate pH 7.0@100mM	Potassium Phosphate pH 7.0@100mM, NaCl 150mM
HEPES pH 7.0@100mM	HEPES pH 7.0@100mM, NaCl 150mM
PIPES pH 7.3@100mM	PIPES pH 7.3@100mM, NaCl 150mM
Sodium Phosphate pH 7.5@100mM	Sodium Phosphate pH 7.5@100mM, NaCl 150mM
Tris pH 7.5@100mM	Tris pH 7.5@100mM, NaCl 150mM
Imidazole pH 8.0@100mM	Imidazole pH 8.0@100mM, NaCl 150mM
Potassium Phosphate pH 7.0@100mM, NaCl 25mM	Potassium Phosphate pH 7.0@100mM, NaCl 300mM
HEPES pH 7.0@100mM NaCl 25mM	HEPES pH 7.0@100mM, NaCl 300mM
PIPES pH 7.3@100mM, NaCl 25mM	PIPES pH 7.3@100mM, NaCl 300mM
Sodium Phosphate pH 7.5@100mM, NaCl 25mM	Sodium Phosphate pH 7.5@100mM, NaCl 300mM
Tris pH 7.5@100mM, NaCl 25mM	Tris pH 7.5@100mM, NaCl 300mM
Imidazole pH 8.0@100mM, NaCl 25mM	Imidazole pH 8.0@100mM, NaCl 300mM
HEPES pH 8.0@100mM	HEPES pH 8.0@100mM, NaCl 150mM
Tris-test pH 8.0@100mM	Tris-test pH 8.0@100mM, NaCl 150mM
Bicine pH 8.0@100mM	Bicine pH 8.0@100mM, NaCl 150mM
Tris pH 8.5@100mM	Tris pH 8.5@100mM, NaCl 150mM
Bicine pH 9.0@100mM	Bicine pH 9.0@100mM, NaCl 150mM
CHES pH 9@500mM	CHES pH 9@500mM, NaCl 150mM
HEPES pH 8.0@100mM, NaCl 25mM	HEPES pH 8.0@100mM, NaCl 300mM
Tris-test pH 8.0@100mM, NaCl 25mM	Tris-test pH 8.0@100mM, NaCl 300mM
Bicine pH 8.0@100mM, NaCl 25mM	Bicine pH 8.0@100mM, NaCl 300mM
Tris pH 8.5@100mM, NaCl 25mM	Tris pH 8.5@100mM, NaCl 300mM
Bicine pH 9.0@100mM, NaCl 25mM	Bicine pH 9.0@100mM, NaCl 300mM
CHES pH 9@500mM, NaCl 25mM	CHES pH 9@500mM, NaCl 300mM

Tableau 5 : compositions des cribles de DSF, l'un étant à 0 ou 25 mM et l'autre à 150 ou 300 mM de NaCl.

E. Recherche de l'activité de GH*

Une large gamme de substrats (oligo- et polysaccharides ainsi que des pNP-substrats) a été utilisée pour déterminer si GH* avait une activité glycoside hydrolase et sur quel substrat.

1. Activité sur pNP-substrats

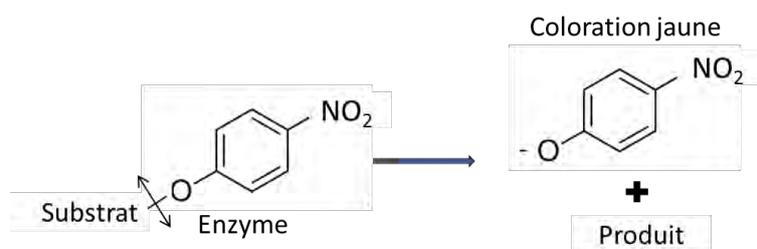


Figure 30 : Principe du dosage de l'activité sur pNP-glycosides. Le groupement paranitrophényl présent sur le pNP-Glycoside est libéré lors de l'hydrolyse enzymatique libérant du paranitrophénate ayant une coloration jaune à pH alcalin

Quand les pNP-glycosides sont hydrolysés (clivage de la liaison pseudo-glycosidique), le groupement paranitrophénate est libéré dans le milieu (figure 30). Ce groupement présente une coloration jaune en milieu alcalin et l'intensité de cette coloration est dépendante de la quantité de groupement paranitrophénol libéré dans le milieu. 19 pNP-glycosides différents ont ainsi été criblés dans le but d'identifier une potentielle activité exoglycosidase de GH*. Les tests d'activité ont été menés à 40 °C en tampon HEPES 200 mM pH 8,0 pour avoir une détection rapide de l'activité car la coloration jaune du paranitrophénate apparaît en milieu alcalin, en présence de 1 mM de substrat et 5 μM d'enzyme. Les premiers tests sur chaque substrat ont été faits en point final sur 24 h puis des cinétiques ont été réalisées sur 24 h avec des prélèvements à 0 min, 30 min, 1h, 2 h, 4 h, 8 h et 24 h sur les 6 meilleurs substrats (pNP-α-L-arabinofuranoside, pNP-α-L-arabinopyranoside, pNP-β-D-galactopyranoside, pNP-β-D-xylopyranoside, pNP-β-D-glucopyranoside et pNP-β-D-lactopyranoside). Chaque prélèvement est stocké à 4 °C. Les résultats ont été lus après le dernier prélèvement, en microplaques, à 401 nm avec le lecteur EON (Biotek). Chaque test a été fait en triplicatas.

2. Activité de phosphorolyse sur oligosaccharides

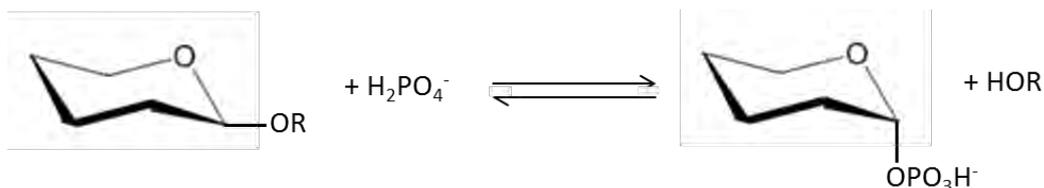


Figure 31 : Principe de la phosphorolyse. Le reste de la chaîne du substrat est représenté par le R.

11 oligosaccharides (di et tri-saccharides) ont été criblés afin de détecter une activité de phosphorolyse sur di- et trisaccharides (figure 31). Lors d'une phosphorolyse du phosphate inorganique est utilisé pour jouer le rôle de nucléophile qui est manquant dans la séquence de l'enzyme. Les tests ont été effectués à 40 °C, en tampon HEPES 200 mM à pH 8,0 en présence de 4,5 mg/ml de substrats et de 5 μM d'enzyme et en présence ou en absence de phosphate inorganique à 10 mM final. Des cinétiques ont été réalisées sur 24 h avec des prélèvements après 0, 1, 2, 4, 8 et 24 h. Chaque prélèvement est congelé immédiatement. Une fois tous les prélèvements effectués, les échantillons sont dilués pour être à une concentration finale en sucre de 2 mg/ml. L'activité a été détectée par chromatographie sur couche mince (CCM) en utilisant différents solvants : butanol/ acide acétique/eau (2/1/1) (pour chitobiose, chitopentaose, lactose, cellobiose, manobiose, xylobiose, galactobiose, xylosylcellobiose et galactosylmannobiose), propanol/eau (8,2/1,5) (pour arabinobiose et arabinosylxylobiose) et acétate d'éthyle/acide acétique/eau (2/1/1) (pour acide alduronique et héparosane). 2 μl à 2 mg/ml de chaque réaction ainsi que les standards sont déposés sur plaque de silice à 1 cm du bord inférieur de la plaque. Les plaques sont déposées dans des cuves dont l'atmosphère a été préalablement saturée par le solvant de migration. Quand le front de migration arrive à environ 1 cm du bord supérieur de la plaque, la migration est arrêtée et les plaques sont séchées. Pour révéler les tâches, les plaques sont plongées dans de l'orcinol sulfurique (10 % poids/volume d'orcinol dans une solution d'acide sulfurique/éthanol/eau dans un rapport 3/72,5/22,5) puis révélées par chauffage (méthode de Tillmans et Philippi, 1929). L'acide sulfurique provoque la déshydratation des sucres qui vont donner des dérivés de furfural. Ces dérivés vont se condenser avec l'orcinol pour donner une tâche brune. La comparaison de la migration des tâches issues des réactions

avec celles des standards (calcul du facteur de rétention) permet de savoir si le substrat a été hydrolysé par l'enzyme.

3. Activité sur polysaccharides

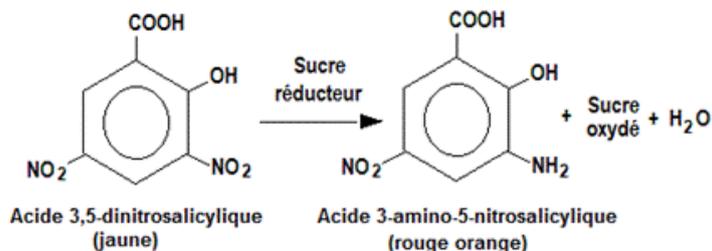


Figure 32 : principe du dosage au DNS.

26 polysaccharides ont été criblés afin de détecter une activité endoglycosidique. L'activité a été mesurée par dosage au DNS. L'acide dinitrosalicylique (DNS) a une coloration jaune en absence de sucre réducteur et une coloration orange voire rouge brique en présence de sucres réducteurs (figure 32), il s'agit d'une réaction de réduction. Plus la couleur est foncée plus la quantité de sucres réducteurs libérés est grande. Les tests ont été effectués à 40 °C en tampon citrate à 200 mM à pH 5,0 puis en tampon HEPES à 200 mM à pH 8,0. Les cinétiques ont été réalisées sur 24 h avec des prélèvements à 0 min, 30 min, 1 h, 2 h, 4 h, 8 h et 24 h. La réaction est arrêtée avec le DNS puis les échantillons sont chauffés 5 min à 95 °C. La lecture des résultats se fait en microplaque à 540 nm avec un lecteur TECAN SUNRISE. La quantité d'extrémités réductrices libérée est quantifiée grâce à une gamme étalon. Celle-ci est faite avec le monosaccharide correspondant à la sous-unité réductrice du sucre libéré lors de la réaction d'hydrolyse du polysaccharide, par exemple du xylose dans le cas de l'hydrolyse d'un xylane. Les concentrations de la gamme étalon sont comprises entre 0 et 1 mg/ml avec un pas de 0,1 mg/ml. Les tests ont été réalisés en triplicatas. Pour la caractérisation sur l'arabinoxylane de seigle (RAX), la gamme étalon a été réalisée avec du xylobiose (X2) car nous avons pu voir que GH* ne libère pas de xylose et il a été démontré qu'avec du xylobiose, les concentrations de la gamme étalon correspondent mieux aux quantités de sucre réducteurs libérés lors de l'hydrolyse du substrat (McCleary et McGeough 2015).

Des tests en présence de cofacteurs (NADH, calcium ou phosphate inorganique) ont été effectués dans les mêmes conditions que précédemment pour voir un éventuel effet sur l'activité.

F. Chromatographie échangeuse d'anion à haute performance couplée à un système de détection ampérométrique pulsée (HPAEC-PAD)

Les produits d'hydrolyse des polysaccharides sur lesquels GH* a une activité ont été analysés par HPAEC-PAD. La séparation des molécules se fait en fonction de leur taille ainsi que de leur charge. Les sucres étudiés sont placés dans de la soude à 150 mM à un pH supérieur à 13 afin de les charger négativement. En effet à cette valeur de pH les fonctions alcool des sucres vont être chargées négativement car leur pKa est d'environ 12. Les sucres, une fois chargés négativement, vont se fixer sur la résine échangeuse d'anions de la colonne. L'élution se fait par compétition entre le sel de l'éluant (150 mM NaOH, 500 mM acétate de sodium) et les sucres fixés sur la colonne. Plus il y a de charge plus le sucre sera difficile à décrocher, ainsi un monosaccharide (xylose par exemple) sera élué avant un disaccharide de même nature (xylobiose par exemple). La séparation d'épimères est possible car leur pKa est différent. La détection des sucres se fait grâce à la cellule ampérométrique. Les sucres sont détectés en mesurant le courant électrique généré par leur oxydation sur l'électrode en or. La colonne utilisée est une PA-100 Dionex ayant un diamètre de 4 mm. Cette colonne permet d'analyser les oligosaccharides et aussi les polysaccharides.

Les standards utilisés sont l'arabinose et des xylooligosaccharides allant de X1 à X6. Pour déterminer la nature des produits d'hydrolyse, les temps de rétention des pics provenant de l'hydrolyse des polysaccharides ont été comparés aux temps de rétention des standards.

La quantification des produits d'hydrolyse a été rendue possible par l'injection de chacun des standards à différentes concentrations (de 5 à 30 mg/ml avec un pas de 5 mg/ml) et grâce au calcul du facteur de réponse pour chaque standard. Le facteur de réponse est calculé en divisant l'aire du pic ($nC \cdot \text{min}$) par la concentration molaire (μM) pour chaque concentration en standard puis en faisant la moyenne de toutes les valeurs obtenues. L'aire du pic de chaque produit d'hydrolyse est divisée par le facteur de réponse du standard correspondant et ainsi la concentration en produit est obtenue.

L'activité de GH* a été testée sur des oligosaccharides. Les réactions ont été réalisées à 40 °C, dans 50 mM de tampon phosphate à pH 6,0, en présence de GH* à 5 µM final et de chaque xylooligosaccharide (de X3 à X6) à 100 µM final. Les réactions ont été suivies sur 24 h avec un prélèvement fait à 0 min, 30 min, 1 h, 2 h, 4 h, 8 h et 24 h. Chaque prélèvement a été injecté sur HPAEC-PAD et les produits d'hydrolyse ont pu être analysés en comparant leur temps de rétention aux standards (X1 à X6 aux différentes concentrations précisées ci-dessus).

G. Détermination des paramètres optimaux

Les paramètres optimaux (pH optimal et température optimale) ont été déterminés sur arabinoxylane de seigle (RAX) qui est le meilleur substrat de l'enzyme identifié à ce jour.

Pour déterminer le pH optimal, une gamme de pH allant de pH 2,6 à 8,0 avec un pas de 0,5 unité de pH a été utilisée. Plusieurs tampons ont été nécessaires : phosphate-citrate (mélange de Na₂HPO₄ à 0,2 M avec de l'acide citrique à 0,1 M ajusté au pH-mètre pour obtenir le pH souhaité) pour les pH 2,6 à 7,0 et phosphate de sodium (mélange de Na₂HPO₄ à 0,2 M avec du NaH₂PO₄ à 0,2 M ajusté au pH-mètre pour obtenir le pH souhaité) pour les pH 7,0 et 8,0.

Les réactions ont été effectuées à 40 °C sur 3 h en présence de GH* à 5 µM et de RAX à 2 mg/ml. Les réactions sont arrêtées par ajout d'un même volume de DNS (acide dinitrosalicilique) que de milieu réactionnel puis chauffées pendant 10 min à 95 °C afin de dénaturer l'enzyme.

La température optimale a été déterminée à partir d'une gamme de température allant de 30 °C à 60 °C avec un pas de 5 °C. Les réactions ont été faites à pH 5,5 en tampon phosphate-citrate (43 mM citrate, 114 mM phosphate), en présence de 5 µM de GH* et de 2 mg/ml de RAX. Les réactions ont aussi été arrêtées après l'ajout d'un même volume de DNS que de milieu réactionnel en incubant le mélange réactionnel pendant 10 min à 95 °C.

Toutes les réactions ont été effectuées en triplicatas et les résultats ont été analysés après lecture de l'absorbance à 540 nm de chacun des puits des microplaques en utilisant le spectrophotomètre EON.

H. Cristallisation

L'obtention d'un cristal de protéine est une des étapes limitantes de la cristallographie. En effet, de nombreuses conditions doivent être réunies afin d'obtenir un cristal. Il n'existe pas à ce jour, de moyen de prédire une condition de cristallisation. Ces conditions sont déterminées de manière empirique sur une échelle de temps s'étalant de quelques heures à quelques mois. Théoriquement, pour cristalliser, les protéines pures doivent se rapprocher les unes des autres, tout en conservant leur structure tertiaire, afin d'établir un ensemble d'interactions faibles assurant la cohésion de l'édifice (liaisons hydrogène, liaisons ioniques, forces de van der Waals). Ces interactions sont appelées contacts cristallins. La cristallisation d'une protéine repose sur un processus composé de trois étapes observable sur un diagramme de phase (Figure 33A) :

- La sursaturation : un état où la protéine est en concentration supérieure à sa limite de solubilité. La concentration en protéine est de l'ordre de la dizaine de mg.mL^{-1} . L'état métastable de sursaturation va contraindre le rapprochement des protéines en solution et évoluer, soit vers la précipitation, soit vers la nucléation des protéines,
- La nucléation : dans certaines conditions, l'état de sursaturation va favoriser l'apparition de germes cristallins,
- La croissance cristalline : un retour à l'équilibre par l'incorporation des macromolécules protéiques à l'édifice cristallin. Une fois revenue à la limite de solubilité de la protéine, la croissance s'arrête.

Le succès de la cristallisation d'une protéine repose sur un ensemble de paramètres physico-chimiques variés, tels que la nature chimique, la concentration, le pH du tampon utilisé et la nature de l'agent précipitant (qui permet de réduire la solubilité de la protéine), mais aussi de la température, la nature et la concentration des sels.

Il existe différentes techniques permettant d'amener les protéines à cristalliser, la plus utilisée est la technique de diffusion de vapeur. Une goutte contenant la solution de protéine mélangée à la solution de cristallisation est équilibrée contre un réservoir contenant la solution de cristallisation. La diffusion de vapeur va se faire de la goutte vers le réservoir, augmentant la concentration en protéine au sein de la goutte lentement jusqu'à l'entrée en sursaturation. Il existe deux variantes à cette technique :

- La diffusion de vapeur en goutte assise (figure 33B) : généralement utilisée lors d'un criblage robotisé de conditions de cristallisation, minimisant le volume de la goutte (100 à 200 nL) et donc la quantité d'échantillon biologique nécessaire. Ceci permettant une recherche plus exhaustive de conditions permettant d'obtenir des objets cristallins,
- La diffusion de vapeur en goutte suspendue (figure 33C) : essentiellement utilisée pour optimiser manuellement une condition de cristallisation préalablement identifiée par un criblage robotisé. Dans ce cas, les gouttes sont de l'ordre du μL décuplant ainsi les quantités de protéine nécessaires.

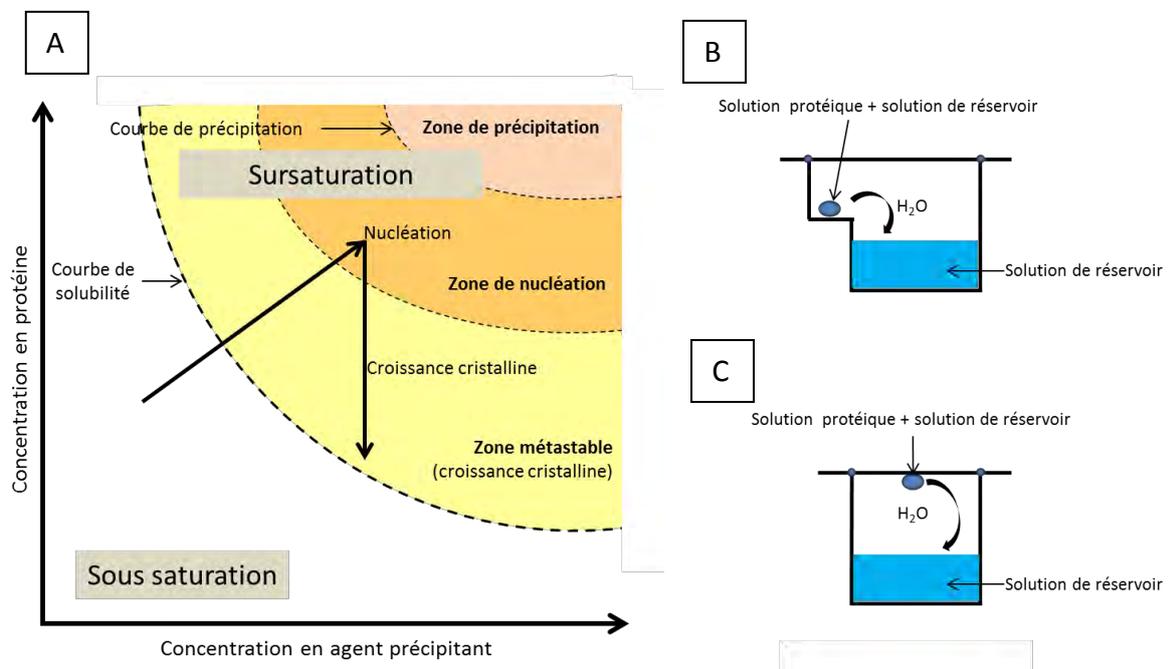


Figure 33 : A. Diagramme de phase. B. Diffusion de vapeur en goutte assise. C. Diffusion de vapeur en goutte suspendue.

Plusieurs cribles commerciaux ont été utilisés pour tenter de cristalliser GH*. Il s'agit des cribles JCSG I à IV, PEG I et II, pHclear I et II (Qiagen) ainsi que Morpheus et Nucleix (Molecular Dimensions). Différentes concentrations en protéine 12, 8, 4 et 3,5 mg/ml ont aussi été testées en présence ou non de L-arabinose ou de D-galactose à 100 mM. Les tests ont été effectués en microplaques (MRC 2 Well Crystallization Plate Hampton Research) en utilisant la méthode de la diffusion de vapeur en goutte assise. Les cribles commerciaux ont été distribués dans les puits à l'aide d'une pipette multicanaux (Liquidator, Rainin).

Les gouttes contenant 0,2 µl de solution protéique et 0,2 µl de solution de réservoir ont été préparées à l'aide d'un robot de cristallisation de type Nanodrop ExtY (Innovadyne technologies) ou Mosquito (TTPLabtech). Enfin, les plaques sont stockées à 12 °C et visualisées à intervalle régulier en lumière visible et en UV à l'aide du robot de visualisation Rock Imager (Formulatrix).

I. Congélation des cristaux et collecte de données

Les cristaux ont été récupérés à l'aide de boucles en nylon de taille adaptée puis cryoprotégés. Les cristaux obtenus en présence de 100 mM de L-arabinose n'ont pas nécessité de cryoprotection car le L-arabinose, comme beaucoup de monosaccharides a un effet cryoprotectant. Pour les cristaux obtenus en l'absence de L-arabinose, la cryoprotection a été effectuée en rajoutant 15 % (v/v) de PEG 400, 100 mM de L-arabinose, 100 mM de D-galactose ou 100 mM de D-glucose en fonction des besoins. Les cristaux sont ensuite congelés dans un flux d'azote gazeux à 100 K et analysés sur site synchrotron. Plusieurs jeux de données de diffraction ont pu être enregistrés sur différents cristaux de GH*. Le meilleur a été obtenu pour un cristal de cette protéine cocrystallisé en présence de 100 mM de L-arabinose dans les conditions E2 du crible commercial JCSG II (0,2 M NH₄Cl et 20 % (p/v) PEG3350). Ce cristal a permis d'obtenir des données de diffractions des rayons X à une résolution maximale de 1,57 Å sur la ligne ID23-2 de l'ESRF (Grenoble, France).

J. Détermination de la structure de GH*

L'indexation, l'intégration et la mise à l'échelle des réflexions a été faite en utilisant XDS (<http://xds.mpimf-heidelberg.mpg.de/>) (Kabsch 1993). Le cristal appartient au groupe d'espace orthorhombique $P2_12_12_1$ avec une seule molécule par unité asymétrique.

La structure de GH* a été résolue par remplacement moléculaire en utilisant l'endo-β-1,4-glucanase FnCel5A de *Fervidobacterium nodosum* (code pdb : 3NCO) comme modèle (Zheng et al. 2012) et à l'aide du logiciel Phaser (McCoy et al. 2007). Le modèle a été construit manuellement dans des cartes de densité électronique de type 2mFo-DFc pondérées par les SigmaA en utilisant WinCOOT (<http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/>) (Emsley et Cowtan 2004) et plusieurs cycles d'affinement ont été effectués avec Refmac5 (suite CCP4) (<http://www.ccp4.ac.uk/html/refmac5.html>) (Murshudov et al. 2011). La structure a été

résolue à 1,6Å de résolution et les données statistiques de la résolution de la structure de GH* sont présentées dans le tableau 15 (cf. chapitre III, partie IV), section B), p 145).

Chapitre I : Comparaison du potentiel hémicellulolytique des microbiotes de deux termites

I. Introduction

Ces dernières années, face à la demande grandissante de production d'énergie, au réchauffement climatique et à la raréfaction des énergies fossiles, trouver de nouvelles énergies renouvelables et moins polluantes est devenu un enjeu économique et environnemental majeur. L'utilisation de la biomasse végétale semble être une bonne alternative, notamment car il s'agit d'une source de carbone renouvelable constituée de 65 à 85% de polysaccharides en poids sec (Pettersen 1984) . Toutefois, la valorisation de cette biomasse nécessite la mise au point de méthodes de bioraffinerie moins polluantes.

Afin de développer des procédés plus propres, les biocatalyseurs jouent un rôle essentiel et notamment les glycosides hydrolases produites par les microorganismes qui suscitent depuis de nombreuses années un grand intérêt (Shallom et Shoham 2003 ; Cherubini 2010). Dans le but de découvrir de nouvelles enzymes pour hydrolyser plus efficacement la paroi végétale, les chercheurs se sont récemment intéressés à l'un des organismes les plus efficaces pour dégrader la biomasse végétale, à savoir le termite (Brune 2014). Les termites sont même parfois considérés comme des bioréacteurs d'une grande efficacité. En effet, on retrouve des termites sur tous les continents excepté l'antarctique (Brune 1998 ; Brune 2014).

Aujourd'hui, on compte près de 3000 espèces de termites réparties en 2 grands types : les termites inférieurs et les termites supérieurs (Kambhampati et Eggleton 2000). Les termites sont aussi répartis en 4 grands groupes en fonction de leur alimentation. Les termites du premier groupe se nourrissent principalement de bois morts et d'herbes. Ce groupe n'est constitué que de termites inférieurs. L'alimentation des termites du deuxième groupe est la plus variée (bois morts, lichens, herbes, feuillages). Les termites du troisième groupe se nourrissent de la couche supérieure de la litière composée de matière organique. Enfin, les

termite du quatrième groupe se nourrissent de la couche inférieure de la litière qui est plus minéralisée que la couche supérieure. Les termites de ces 3 derniers groupes ne sont que des termites supérieurs (Donovan, Eggleton, et Bignell 2001). Parmi les termites supérieurs on retrouve des termites xylophages stricts qui se nourrissent uniquement de bois, des termites humivores qui se nourrissent de matière en décomposition (humus), des termites humivores et xylophages qui se nourrissent d'humus et de bois en partie décomposé et, enfin, les termites champignonnistes qui se nourrissent de bois et de champignons qu'ils élèvent. D'autre part, il a été démontré que les termites humivores possèdent moins d'activité glycoside hydrolase que les autres termites supérieurs. A l'inverse, les termites xylophages possèdent le plus grand arsenal d'enzymes capables de dégrader les différents polysaccharides des parois végétales (Donovan, Eggleton, et Bignell 2001 ; Bastien-Uluis 2012 ; Arnal 2014). Tous les termites, inférieurs comme supérieurs, possèdent des enzymes endogènes capables de dégrader principalement la cellulose et un microbiote dans leur tractus digestif composé de bactéries et d'archaea. Les termites inférieurs possèdent en plus de ces micro-organismes des flagellés cellulolytiques (Ohkuma 2003 ; Ni et Tokuda 2013). Le microbiote du tractus digestif des termites produit de grandes quantités d'enzymes, ce qui leur permet de dégrader totalement la biomasse végétale (Brennan et al. 2004). D'autre part, le microbiote semble spécifique de l'alimentation du termite car il a été montré une évolution conjointe des microorganismes et de leurs hôtes (Hongoh et al. 2005 ; Warnecke et al. 2007 ; He et al. 2013).

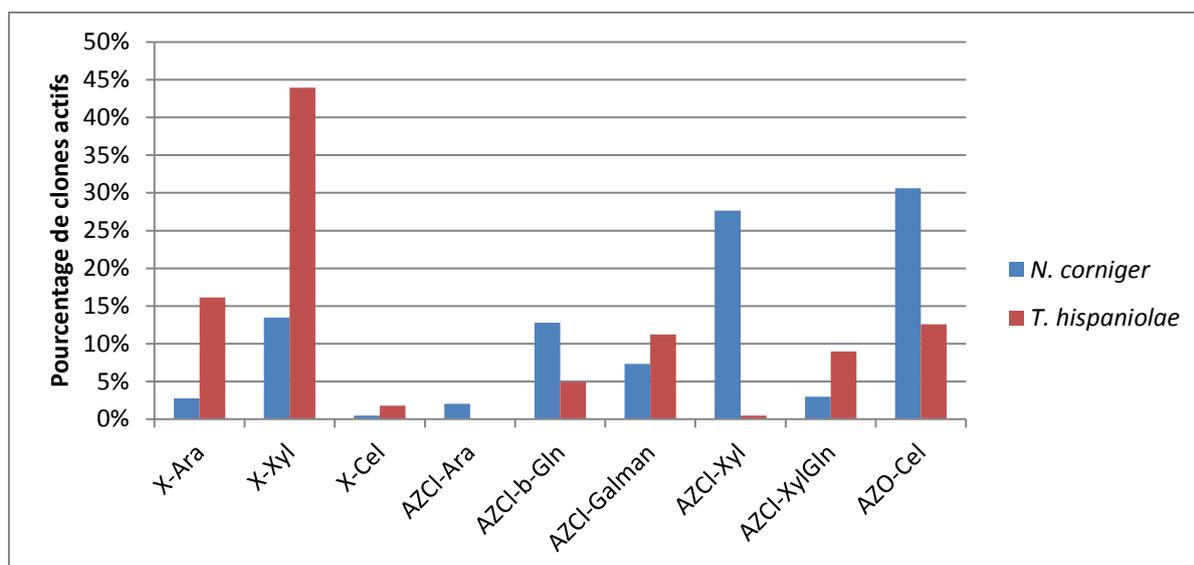
Dans ces travaux de thèse, qui font suite à ceux de Gregory Arnal (Arnal 2014), afin de trouver de nouvelles enzymes impliquées dans la dégradation de la biomasse, nous avons choisi d'explorer le microbiote intestinal de deux espèces de termites supérieurs appartenant à deux des différents groupes cités ci-dessus : *Nasutitermes corniger*, qui est un termite xylophage strict qui se nourrit de bois sain (sec ou humide) et de bois partiellement décomposé (He et al. 2013), et *Termes hispaniolae*, qui est un termite humivore et xylophage qui se nourrit de litière et de bois décomposé (Roisin et al. 2006). Comme ces deux espèces n'ont pas la même alimentation, étudier leurs microbiotes et plus particulièrement les enzymes impliquées dans la modification des glucides qu'ils sécrètent pourrait nous permettre de découvrir de nouvelles enzymes capables de dégrader la

biomasse végétale à différents stades de décomposition et ainsi de pouvoir dégrader plus efficacement la paroi végétale.

L'exploration du métagénome de ces termites a été réalisée par métagénomique fonctionnelle au cours des thèses de Géraldine Bastien et Gregory Arnal (Bastien-Uluis 2012 ; Arnal 2014). Les banques métagénomiques obtenues contiennent respectivement 12000 et 48000 clones pour *N. corniger* et *T. hispaniolae*. Un criblage fonctionnel, sur substrats chromogéniques, à haut débit a été effectué sur la plateforme ICEO du laboratoire et a permis d'isoler 348 clones pour *N. corniger* et 223 clones *T. hispaniolae*. Les fosmidés des clones les plus intéressants ont été séquencés et l'annotation réalisée. L'annotation des CAZymes a été possible grâce à la collaboration avec l'équipe de Bernard Henrissat qui gère la base de données CAZy (Carbohydrate Active enZymes). Les résultats des cribles fonctionnels ont mis en avant qu'une grande majorité des clones actifs de la banque de *N. corniger* (83,3 %) a une activité sur polysaccharides (AZO-cellulose et AZCI-xylane) ce qui correspond à une activité de type endoglycosidase. En revanche la majorité des clones actifs de la banque *T. hispaniolae* (61,9 %) a une activité sur X-substrat (X-xyl et X-ara) ce qui correspond à une activité de type exoglycosidase (tableau 6 et figure 34). Les résultats du criblage fonctionnel, semblent donc montrer que le microbiote de *N. corniger* produit majoritairement des enzymes à activité endoglycosidase alors que le microbiote de *T. hispaniolae* semble produire majoritairement des enzymes à activité exoglycosidase (tableau 6 et figure 34).

	<i>N. corniger</i>	<i>T. hispaniolae</i>
Taille de la banque (nombre de clones)	12000	48000
Nombre de clones criblés	11520	18432
Nombre de hits positifs	438 (3,8%)	223 (1,2%)
Nombre de clone à activité endoglycosidase	365 (83,3%)	85 (38,1%)
Nombre de clone à activité exoglycosidase	73 (16,7%)	138 (61,9%)
Nombre de clones multiactifs	74 (16,9%)	21 (9,4%)

Tableau 6 : Nombres de clones positifs confirmés pour chacune des banques métagénomiques de *N. corniger* et *T. hispaniolae*. Les chiffres entre parenthèses indiquent le pourcentage de clones actifs pour chaque type d'activité par rapport au nombre total de clones positifs. Le pourcentage de hits positifs est calculé par rapport à la totalité des clones criblés.



Figures 34 : Pourcentage de clones actifs sur les différents substrats testés lors du crible fonctionnel sur substrat chromogénique de chaque banque. X-arabinofuranoside (X-Ara), X-xylopyranoside (X-Xyl), X-cellobioside (X-Cel), AZCI-arabinogalactan (AZCI-Ara), AZCL- β -glucane (AZCI-bGln), AZCI-galactomannane (AZCI-Galman), AZCI-xylane (AZCI-Xyl), AZCI-xyloglucane (AZCI-XylGln) et AZO-cellulose (AZO-Cel).

D'autre part, les 2 banques contiennent des clones dits multi-actifs (tableau 7), à savoir qu'un même clone s'est révélé actif sur plusieurs substrats criblés. Les plus intéressants sont ceux qui présentent quatre activités car on peut penser qu'ils possèdent une grande variété d'enzymes différentes.

Crible fonctionnel sur substrat chromogénique	Activité criblée	<i>N. corniger</i>	<i>T. hispaniolae</i>
X-Ara/X-Xyl	Abf/Xyl	0	1
AZCI-Ara/AZCI-Xyl	Abn/XIn	2	0
AZCI-b-Gln/AZCI-xyl	b-Gln/XIn	5	0
AZCI-b-Gln/AZCI-Xylgln	b-Gln/X-Gln	1	0
AZCI-b-Gln/AZO-Cel	b-Gln/Cell	29	2
AZCI-Galman/AZCI-Xyl	Man/XIn	12	0
AZCI-Galman/AZO-Cel	Man/Cell	2	0
AZCI-Galman/AZCI-Xylglc	Man/X-Gln	0	1
AZCI-Galman/X-xyl	Man/Xyl	0	1
AZCI-Xyl/AZO-Cel	XIn/Cell	2	0
AZCI-XylGln/AZO-Cel	X-Gln/Cell	6	6
AZCI-Xyl/X-Xyl	XIn/Xyl	1	0
Total bi-fonctionnel		60	11
AZCI-Ara/AZCI-Galman/AZCI-Xyl	Abn//Man/XIn	4	0

AZCI-b-Gln/AZCI-Galman/AZCI-XylGln	b-Gln/Man/X-Gln	0	1
AZCI-b-Gln/AZCI-Galman/AZO-Cel	b-Gln/Man/Cell	0	1
AZCI-b-Gln/AZCI-Xyl/AZO-Cel	b-Gln/Xln/Cell	5	0
AZCI-b-Gln/AZCI-XylGln/AZO-Cel	b-Gln/X-Gln/Cell	3	5
AZCI-XylGln/AZO-Cel/X-Cel	X-Gln/Cell/Cbh	0	1
Total tri-fonctionnel		12	8
AZCI-b-Gln/AZCI-Galman/AZCI-XylGln/AZO-Cel	b-Gln/Man/X-Gln/Cell	1	1
AZCI-b-Gln/AZCI-XylGln/AZO-Cel/X-Ara	b-Gln/X-Gln/Cell/Abf	0	1
AZCI-b-Gln/AZCI-XylGln/AZO-Cel/X-Xyl	b-Gln/X-Gln/Cell/Xyl	1	0
Total tétra-fonctionnel		2	2

Tableau 7 : Tableau présentant les clones multiactifs pour chacune des 2 banques.

Arabinofuranosidase (Abf), xylosidase (Xyl), arabinanase (Abn), xylanase (Xln), glucanase (b-Gln), xyloglucanase (X-gln), cellulase (Cell), mannanase (Man) et cellobiohydrolase (Cbh).

La banque de *N. corniger* a été, pour la majeure partie, annotée par Gregory Arnal durant sa thèse. Mon travail ici a consisté à finir l'annotation de 24 clones de *N. corniger* et faire l'annotation des 66 clones de *T. hispaniolae* pour ensuite comparer le potentiel hémicellulolytique des microbiotes des deux espèces de termites et de répondre à la question suivante : est ce que les différences observées au niveau du criblage sont confirmées par l'annotation ?

II. Résultats de l'annotation fonctionnelle

Les fragments métagénomiques ont été séquencés puis annotés au niveau fonctionnel et taxonomique. Nous décrivons ici uniquement l'annotation fonctionnelle relative aux glucides. Pour la partie fonctionnelle, la majorité des modules protéiques trouvés sont des glycosides hydrolases (GH) dans les 2 espèces de termites (figure 35A et B). Des modules appartenant à des familles de carbohydate estérases (CE), de carbohydate binding modules (CBM), de polysaccharide lyases (PL) et de glycosyl transférases (GT) ont aussi été retrouvés dans les 2 banques mais dans différentes proportions. 2 modules appartenant à la famille 1 des AA ont été identifiés dans la banque de *T. hispaniolae*. A noter que lorsque Gregory Arnal a traité les données issues de la banque métagénomique de *N. corniger*, la classe des enzymes à activité auxiliaire n'avait pas encore été implémentée dans la base de données CAZy (Levasseur et al. 2013b). Il est donc possible que l'absence d'AA ne soit due

qu'au fait que ces enzymes n'étaient pas encore systématiquement recherchées lorsque l'annotation a été faite, cette annotation devra être mise à jour.

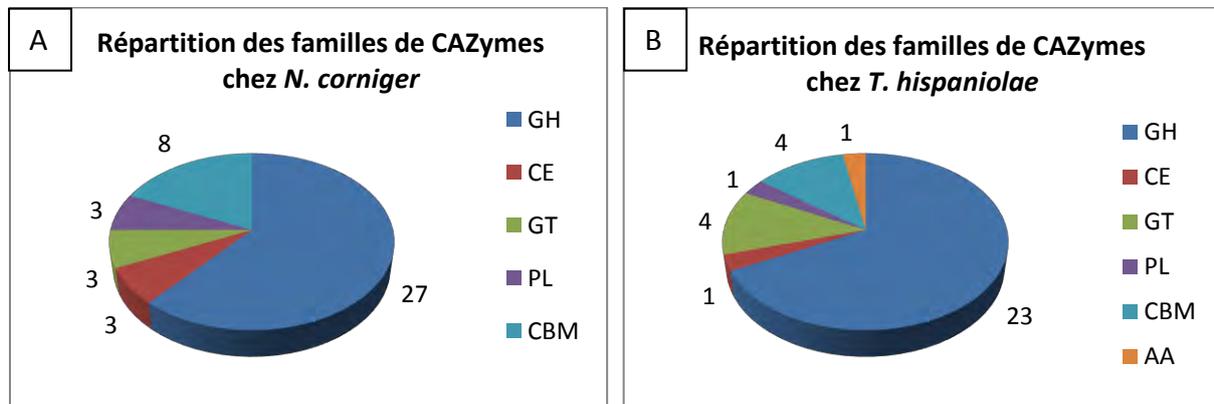


Figure 35 : CAZymes identifiées dans la banque de *N. corniger* (A) et dans la banque de *T. hispaniolae* (B). Glycoside hydrolase (GH), carbohydrate estérase (CE), glycosyl transférases (GT), polysaccharide lyase (PL), carbohydrate binding module (CBM) et Auxiliary activity (AA).

III. Comparaison des familles de GH retrouvées dans les deux banques

A. Familles retrouvées dans les deux banques

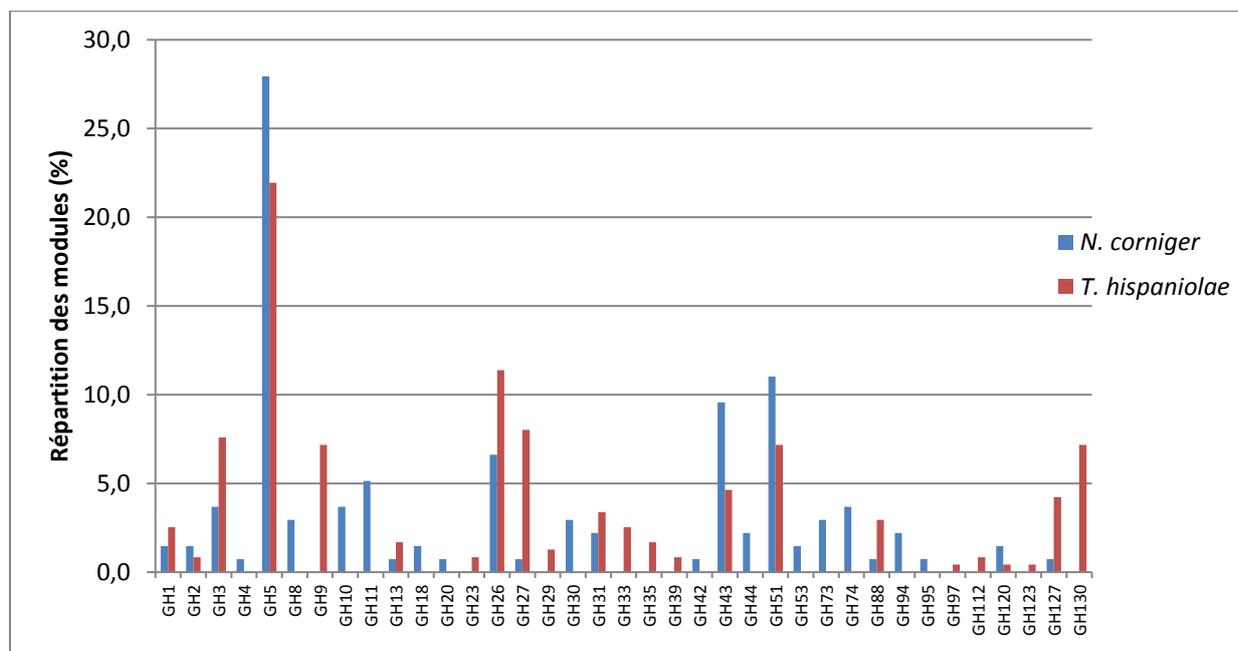


Figure 36 : Histogramme présentant la répartition en pourcentage des familles de GH dans les 2 banques. Les pourcentages sont calculés en divisant le nombre de modules protéiques trouvés pour

chaque famille d'enzymes par la totalité des modules protéiques annotés dans chaque banque métagénomique (tableau 8).

Plusieurs familles de GH sont communes aux 2 banques et 4 d'entre elles font partie des familles les plus représentées au sein de chaque banque (figure 36 et tableau 8). Il s'agit de 2 familles ayant principalement des activités endoglycosidases, les familles GH5 (endo- β -1,4-glucanases, endo- β -1,4-mannanases ou xylanases) et GH26 (β -xylanases ou β -mannanases) et 2 familles présentant majoritairement des activités exoglycosidases, les familles GH43 (α -L-arabinofuranosidases ou β -D-xylosidases) et GH51 (α -L-arabinofuranosidases). Ceci n'est pas surprenant puisque la cellulose est le polysaccharide majeur de la paroi végétale, les xylanes sont les hémicelluloses les plus représentées chez les céréales et les galactomannanes sont aussi très présents dans les parois végétales des bois tendres (Scheller et Ulvskov 2010 ; Lagaert et al. 2014) et les activités citées ci-dessus font donc partie des principales activités nécessaires à la dégradation de cette biomasse végétale.

La famille 5 des glycosides hydrolases est la famille la plus représentée dans les deux banques (figure 36 et tableau 8). Il s'agit principalement d'endo- β -1,4-glucanases mais aussi d'endo- β -1,4-mannanases ou d'endo- β -1,4-xylanases. Cette famille est une des plus grandes de la classification CAZy avec plus de 8000 séquences, dont 540 caractérisées, répertoriées à ce jour. Elle a été divisée en 53 sous-familles en fonction de l'activité et de la séquence (Aspeborg et al. 2012). La cellulose, qui est un polymère de D-glucopyranose lié en β -1,4 et étant le polysaccharide principal des parois végétales (Somerville et al. 2004), il n'est pas surprenant que la famille des GH5 soit très représentée dans le microbiote des termites. Par ailleurs pour dégrader totalement la cellulose, d'autres activités sont nécessaires comme des β -glucosidases (retrouvées notamment dans les familles de GH3, 4, 5, 9, et 30) ou des cellobiohydrolases (retrouvées notamment dans les familles de GH9 et 74) (Bornscheuer, Buchholz, et Seibel 2014). Des familles présentant ces activités sont retrouvées dans les deux banques.

La famille 26 (β -mannanases) est aussi retrouvée dans les deux banques (figure 36 et tableau 8) mais est plus représentée chez *T. hispaniolae*, ce qui est surprenant au vu des résultats des cribles fonctionnels car il y a plus d'activités sur AZCl-xylane et AZCl-galactomannane dans la banque métagénomique de *N. corniger* que dans celle de *T. hispaniolae* (figure 34).

Cependant, on peut noter qu'il y a une bonne corrélation entre les enzymes retrouvées et les cribles fonctionnels. Ces enzymes vont permettre la dégradation des mannanes en oligosaccharides qui seront ensuite pris en charge par d'autres enzymes telles que les β -D-galactosidases (Scheller et Ulvskov 2010 ; Lagaert et al. 2014). Les GH26 sont souvent liées à un CBM35 dans les 2 banques. Il a été démontré que les CBM35 se lient au xylane et/ou au mannane et permettent donc une meilleure accessibilité des GH26 à leur substrat (Montanier et al. 2009).

Les familles 43 (α -L-arabinofuranosidases ou β -xylosidases) et 51 (α -L-arabinofuranosidases) sont aussi retrouvées dans les 2 banques (figure 36 et tableau 8). La présence de ces 2 familles de GH dans le métagénome des termites est en accord avec la dégradation des arabinoxylanes puisque les L-arabinofuranoses et les D-xyloses sont les constituants de ce polysaccharide.

Les GH43 sont parfois liées à des CBM6 dans la banque de *N. corniger* alors que dans la banque de *T. hispaniolae*, les GH43 sont parfois liées à des CBM32. D'après Mewis *et al.* (Mewis et al. 2016) les CBMs de la famille 6 sont les modules les plus souvent retrouvés associés à des GH43 chez les bactéries et notamment aux sous-familles 15 et 16 avec 100% et 64% d'occurrence respectivement. Les CBM6 se lient notamment aux xylanes et aux glucanes, ce qui peut permettre une meilleure accessibilité aux ramifications de ces polysaccharides pour les GH43. Les CBM32 sont aussi retrouvés associés à des GH43 mais de façon beaucoup moins fréquente (autour de 15-20% d'occurrence).

Les GH51 ne sont pas retrouvées liées à des CBMs dans les deux banques. Cette famille, comme celles des GH43, est un peu plus retrouvée dans la banque de *T. hispaniolae* que dans l'autre ce qui est en accord avec les résultats des cribles fonctionnels car 16,1 % des clones actifs de la banque de *T. hispaniolae* ont été identifiés sur X-arabinofuranoside contre 2,7 % pour la banque de *N. corniger*.

GH	<i>N. Corniger</i>		<i>T. Hispaniolae</i>	
	Nombre	%	Nombre	%
GH1	2	1,5	6	2,5
GH2	2	1,5	2	0,8
GH3	5	3,7	18	7,6
GH4	1	0,7	0	0,0

GH5	38	27,9	52	21,9
GH8	4	2,9	0	0,0
GH9	0	0,0	17	7,2
GH10	5	3,7	0	0,0
GH11	7	5,1	0	0,0
GH13	1	0,7	4	1,7
GH18	2	1,5	0	0,0
GH20	1	0,7	0	0,0
GH23	0	0,0	2	0,8
GH26	9	6,6	27	11,4
GH27	1	0,7	19	8,0
GH29	0	0,0	3	1,3
GH30	4	2,9	0	0,0
GH31	3	2,2	8	3,4
GH33	0	0,0	6	2,5
GH35	0	0,0	4	1,7
GH39	0	0,0	2	0,8
GH42	1	0,7	0	0,0
GH43	13	9,6	11	4,6
GH44	3	2,2	0	0,0
GH51	15	11,0	17	7,2
GH53	2	1,5	0	0,0
GH73	4	2,9	0	0,0
GH74	5	3,7	0	0,0
GH88	1	0,7	7	3,0
GH94	3	2,2	0	0,0
GH95	1	0,7	0	0,0
GH97	0	0,0	1	0,4
GH112	0	0,0	2	0,8
GH120	2	1,5	1	0,4
GH123	0	0,0	1	0,4
GH127	1	0,7	10	4,2
GH130	0	0,0	17	7,2
Total	136		237	

Tableau 8 : tableau présentant le nombre et le pourcentage de modules retrouvés pour chaque famille dans les 2 banques.

B. Familles présentes dans une seule banque

1. Familles identifiées majoritairement chez *N. corniger* (termite xylophage) (tableau 8)

Dans la banque de *N. corniger* on retrouve des GH10 et 11 qui sont quasi exclusivement des endo- β -1,4-xylanases. Ces résultats vont dans le sens des cribles fonctionnels car de

nombreux clones ont été trouvés positifs sur AZCl-xylane. D'autre part, aucune GH10 ou GH11 n'a été identifiée dans la banque de *T. hispaniolae*. Ce résultat est très surprenant car les xylanes sont les hémicelluloses majoritaires de la paroi végétale de nombreux types de plantes (Scheller et Ulvskov 2010), donc on pourrait s'attendre à retrouver des xylanases dans le métagénome d'un termite quelle que soit l'espèce.

La famille GH53 qui correspond à des β -galactanases, n'est retrouvée que dans la banque *N. corniger*. Le β -galactane constitue la chaîne principale des arabinogalactanes qui appartiennent à la famille des pectines (Mohnen 2008). Cela impliquerait que ce termite, par l'intermédiaire de son microbiote, soit capable de dégrader, au moins en partie, les pectines. La famille des GH27 est également retrouvée. Il s'agit d' α -galactosidases qui peuvent être impliquées dans la dégradation des galactomannanes en hydrolysant les ramifications α -galactoside.

Enfin, plusieurs familles de chitinases ou chitosanases sont retrouvées chez *N. corniger* (GH8 et 18) ainsi que des N-acétylglucosaminidases (GH18, 20 et 73). La chitine et le chitosan, retrouvés entre autres dans les lichens et dans les champignons qui poussent généralement sur les arbres, sont des polysaccharides essentiellement composés de N-acétylglucosamine. La présence de ces quatre familles d'enzymes semble indiquer une dégradation totale de ces polysaccharides par le microbiote intestinal du termite. Sachant que *N. corniger* se nourrit de tout type de bois qu'il soit sec, humide ou en décomposition, il est possible qu'il se nourrisse aussi du lichen poussant sur ces mêmes bois.

2. Familles identifiées majoritairement chez *T. hispaniolae* (termite humivore et xylophage) (tableau 8)

Les GH9 (endo-glucanases) ne sont retrouvées que chez *T. hispaniolae*, tout comme les GH29 (α -L-fucosidases). La famille des GH9 peut être impliquée dans l'hydrolyse de la cellulose, du xyloglucane et du β -glucane.

La famille GH29 permet l'hydrolyse d' α -L-fucosides retrouvés sur des glycoprotéines et certaines ramifications du rhamnogalacturonane et de certains xyloglucanes (Scheller et Ulvskov 2010 ; Paper et al. 2012). Les bactéries du microbiote se fixent sur les résidus α -L-fucosides des glycanes présents à la surface des cellules du tractus digestif. Une fois fixées les bactéries sécrètent des fucosidases pour hydrolyser le L-fucose et s'en servir pour leur métabolisme (Stahl et al. 2011 ; Megson et al. 2015).

La famille 23 (muramidases) n'est retrouvée que chez *T. hispaniolae*. L'acide N-acétylmuramique est l'un des constituants du peptidoglycane de la paroi des bactéries gram positives. *T. hispaniolae* est un termite humivore, et compte tenu de la présence des familles de GH citées ci-dessus, ces résultats suggèrent qu'il est possible que ce termite puisse se nourrir également des bactéries du sol et qu'il utilise des GH23 pour se défendre contre des pathogènes, comme c'est le cas chez *Amitermes wheeleri* qui est un termite se nourrissant d'excréments bovins (He et al. 2013). En effet, les muramidases dégradent le peptidoglycane de la paroi des bactéries, ce qui les rend moins virulentes et plus sensibles face aux défenses de l'organisme infecté.

La famille 33 (sialidases) n'est retrouvée que chez *T. hispaniolae*. L'acide sialique est retrouvé principalement chez les animaux (glycoprotéines) et dans une moindre mesure chez les plantes, les champignons ou les bactéries. L'acide sialique est retrouvé à la surface du tractus digestif de nombreux animaux. Les GH33 peuvent donc avoir le même rôle que les GH29 en hydrolysant les acides sialiques à la surface des cellules.

Les familles à activité exoglycosidase GH35 (β -galactosidases) et 39 (β -xylosidases) ne sont retrouvées que chez *T. hispaniolae*. Ces familles d'enzymes permettent la dégradation totale des ramifications présentes sur les xyloglucanes. Les xylosidases peuvent aussi hydrolyser des xylooligosaccharides issus d'une première dégradation de xylanes.

La famille 97 (α -glucosidases et α -galactosidases) est, elle aussi, retrouvée uniquement chez *T. hispaniolae*. Le galactose lié en α est retrouvé dans les pectines. Une activité α -galactosidase permet de dégrader des galactooligosaccharides issus de l'hydrolyse d'un galactane par exemple. Le glucose lié en α est l'unité de base de l'amidon qui est un produit de réserve notamment retrouvé dans les racines mais aussi les graines. *T. hispaniolae* est humivore et donc il semble logique de trouver des produits d'une première hydrolyse ou des polysaccharides de réserve présents dans les racines ou les graines dans l'alimentation de ce termite.

La famille GH130 (mannooligosaccharides phosphorylases et β -mannosidases) n'est retrouvée que chez *T. hispaniolae*. Cette famille est impliquée dans le métabolisme du N-glycane notamment chez *B. thetaiotaomicron* (Ladevèze et al. 2013), espèce présente dans le microbiote intestinal du termite. D'autre part le β -mannose est retrouvé dans les glycanes de la paroi secondaire des plantes ce qui impliquerait que les GH130 participent à la dégradation de la paroi végétale (Cuskin et al. 2015).

La famille GH127 (β -L-arabinofuranosidase) est présente chez *T. hispaniolae*, cette famille contient des β -L-arabinofuranosidases. Les β -L-arabinofuranosides sont souvent rencontrés dans les glycoprotéines riches en hydroxyprolines qui sont très présentes dans la paroi végétale et essentielles à la croissance des racines. Ces hydroxyprolines arabinosylées joueraient aussi un rôle essentiel dans la croissance reproductive et végétative des plantes (Ito et al. 2014 ; Velasquez et al. 2011). De la même manière que dans le cas des GH97, retrouver des constituants des racines dans l'alimentation du termite (humivore) semble logique et donc retrouver ce type d'activité semble cohérent.

C. Liens avec le mode de vie des termites

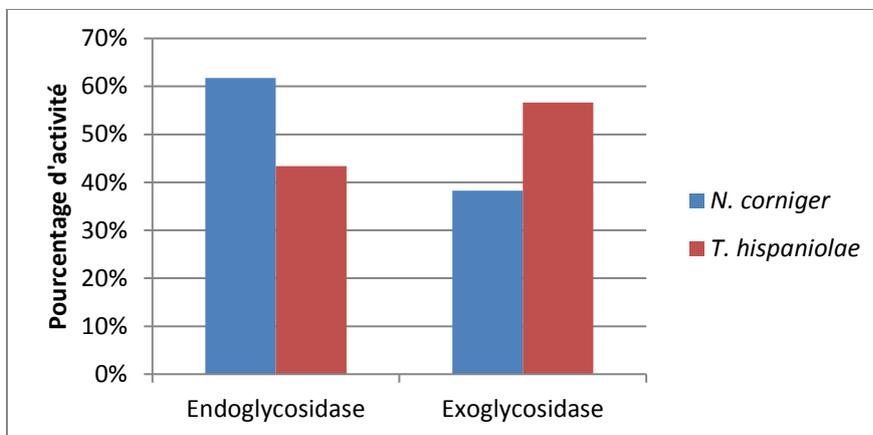


Figure 37 : Pourcentage de GH à activité endoglycosidase et exoglycosidase pour chaque banque.

Le crible fonctionnel sur substrat chromogénique indique une majorité d'activité endoglycosidase chez *N. corniger* (83,3 % contre 16,7 % d'activité exoglycosidase, tableau 6). D'après l'annotation fonctionnelle, on constate que l'écart entre activité exoglycosidase et endoglycosidase est moins important avec 62 % de familles de GH à activité endoglycosidase contre 38 % de GH à activité exoglycosidase (figure 37). De plus, les activités endoglycosidase sont réparties dans 12 familles différentes qui apportent une grande diversité d'activités correspondant à la diversité de polysaccharides retrouvés dans les cribles ainsi que dans les parois végétales (notamment dans les hémicelluloses et les pectines). Les activités exoglycosidase retrouvées sont cohérentes avec les sucres composants les ramifications des polysaccharides de la paroi végétale. Sachant que *N. corniger* est un termite xylophage strict et qu'il y a une diversité des polysaccharides qui est

fonction du type d'arbre, trouver une grande diversité de GH semble logique car cette espèce de termite se nourrit de tout type de bois.

Le crible fonctionnel sur substrat chromogénique de *T. hispaniolae* indique une majorité d'activités exoglycosidase (61,9 % contre 38,1 % d'activité endoglycosidase, tableau 6). Au niveau de l'annotation fonctionnelle, la tendance est toujours vraie mais tend, là encore, à s'équilibrer avec 57 % de familles de GH à activité exoglycosidase, contre 43 % d'activité endoglycosidase (figure 36). Parmi les familles de GH à activité exoglycosidase on trouve une grande diversité. Les activités identifiées vont dans le sens de l'hydrolyse de substrats complexes et aussi de l'hydrolyse de substrats déjà en partie dégradés (oligosaccharides) ainsi que des activités spécifiques de résidus osidiques trouvés dans les racines (amidon ou encore glycoprotéines riches en hydroxyprolines arabinosylées). Ces activités semblent donc en bonne corrélation avec une alimentation contenant du bois et de la matière en décomposition et sont donc en adéquation avec un termite humivore et xylophage.

Les différences des rapports d'activité endo/exoglycosidase observées entre les cribles fonctionnels sur substrats chromogéniques et les annotations fonctionnelles peuvent s'expliquer par la stratégie de grands inserts métagénomiques clonés lors de la préparation des banques métagénomiques. Nous avons donc trouvé, après séquençage des clones sélectionnés pour leur capacité à dégrader certains substrats chromogéniques type AZCL-xylane, des enzymes possédant les activités correspondant aux substrats utilisés (xylanases et xylosidases). Toutes les autres enzymes présentes sur les contigs peuvent ou non contribuer à la dégradation du substrat naturel correspondant, faisant ainsi augmenter ou diminuer la proportion des activités endoglycosidase ou exoglycosidase découvertes.

D. Comparaison avec d'autres espèces

Une étude menée par Warnecke et collaborateurs (Warnecke et al. 2007) présente une analyse du microbiote de termites supérieurs appartenant au genre *Nasutitermes* par métagénomique et une analyse des protéines retrouvées dans l'intestin par spectrométrie de masse. On retrouve dans ces termites quasiment toutes les familles de GH trouvées chez *N. corniger*. Ces termites ayant un microbiote proche (essentiellement composé de bactéries des genres *Treponema* et *Fibrobacter*) et le même biotope, il était attendu de retrouver les mêmes familles de GH.

Une autre étude comparative (He et al. 2013), par métagénomique 16S et métatranscriptomique, chez *N. corniger* a montré que les GH5 sont aussi les GH les plus abondantes chez ce termite. Les autres familles les plus retrouvées par He et coll. sont les GH3, 10, 2, 9, 43, et 42 qui sont toutes retrouvées dans notre étude de *N. corniger* sauf la famille 9 qui contient principalement des endo-glucanases (que nous ne retrouvons que chez *T. hispaniolae*). Toutefois de nombreuses autres familles d'endoglucanases sont retrouvées chez *N. corniger* (GH5, 8, 30, 44), donc le termite semble capable d'hydrolyser assez largement la cellulose et les glucanes. La différence entre l'étude de He et coll. et celle ayant l'objet de ces travaux de thèse quant aux les enzymes identifiées vient des techniques utilisées.

L'étude de He et coll. a aussi porté sur d'autres espèces du genre *Nasutitermes* venant du Costa Rica et sur *A. wheeleri* (He et al. 2013). Dans les autres espèces du genre *Nasutitermes*, la famille des GH5 est la deuxième plus représentée derrière les GH3. Toutes les autres familles les plus retrouvées chez les termites du genre *Nasutitermes* dans l'étude de He (2013) sont les mêmes que celles trouvées pour *N. corniger* (citées ci-dessus), par les auteurs de ce papier, mais dans des proportions différentes. Par exemple, il y a plus de 17% de GH5 chez *N. corniger* et 14% chez les autres espèces du genre *Nasutitermes*. Pour les GH2 c'est 7% contre 13,6% respectivement. Les espèces appartenant au genre *Nasutitermes* sont toutes des xylophages, donc retrouver les mêmes familles était attendu. La différence de proportion peut venir de la technique utilisée pour explorer les génomes ou du type de bois sur lequel se nourrissent les différentes espèces. En effet dans l'étude de He et coll. (He et al. 2013), l'exploration du métagénome a été faite par métagénomique 16S. La métagénomique 16S permet d'identifier tous les ARNm du métagénome alors qu'avec la technique que nous avons utilisée, la métagénomique fonctionnelle, nous n'avons séquencé que les clones actifs sur les substrats testés. Il est donc logique d'identifier plus d'enzymes à partir des données de métagénomique 16S en se basant sur la similarité de séquence.

IV. Comparaison des autres enzymes participant à la dégradation des polysaccharides

A. Les estérases

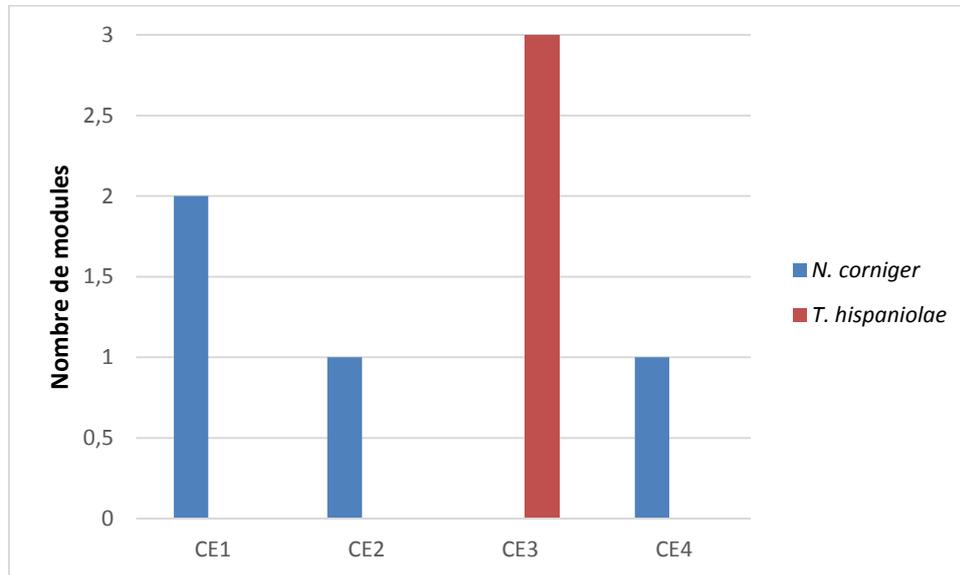


Figure 38 : Histogramme de la répartition des estérases dans les deux banques.

Les familles CE1 (acétylxylane estérases, féruloylestérases), CE2 (acétylxylane estérases) et CE4 (chitine déacétylases, acétylxylane estérases) sont retrouvées exclusivement chez *N. corniger* (figure 38). Les acétylxylane estérases et les féruloylestérases participent à la dégradation des xylanes. En effet, les estérifications des unités osidiques de la chaîne principale ou des ramifications gênent l'activité de certaines GH (Wong 2005 ; Chávez, Bull, et Eyzaguirre 2006) en empêchant l'accès au substrat. La présence de ces estérases est en accord avec l'idée que le microbiote de *N. corniger* lui permettrait de dégrader en grande partie la biomasse végétale.

Chez *T. hispaniolae* seule la famille des CE3 (acétylxylane estérases) a été identifiée (figure 38). La présence de très peu de familles d'estérases chez *T. hispaniolae* et notamment d'aucune famille de féruloylestérase semble indiquer que le termite se nourrit de substrats déjà partiellement hydrolysés sur lesquels il n'y aurait que très peu d'estérifications. Cette famille n'est pas retrouvée chez *N. corniger*, toutefois la même activité est retrouvée dans les familles de CE1, 2 et 4.

B. Les polysaccharides lyases

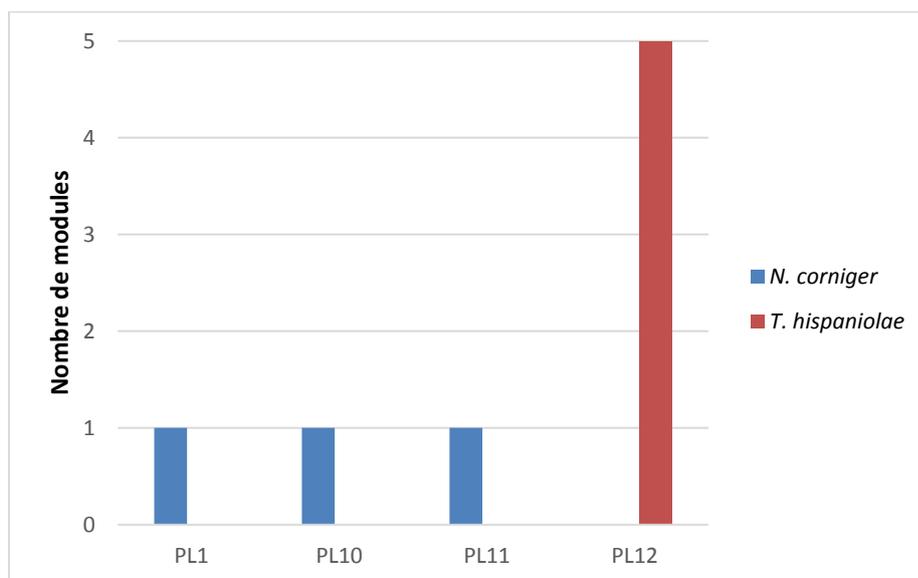


Figure 39 : Histogramme de la répartition des polysaccharide-lyases dans les deux banques.

Un module de chacune des familles PL1, 10 et 11 a été identifié chez *N. corniger* (figure 39). Il s'agit de pectate/pectine lyases, de pectate lyases et de rhamnogalacturonane lyases respectivement. Ces enzymes interviennent dans la dépolymérisation des pectines. *N. corniger* est un termite xylophage qui semble pouvoir dégrader tous les polysaccharides de la paroi végétale (cellulose, hémicelluloses, pectines).

Chez *T. hispaniolae*, plusieurs modules appartenant à la famille 12 des polysaccharides lyases (héparine-sulfate lyases) ont été retrouvés (figure 39). Les héparines sulfates sont des glycosaminoglycanes retrouvés notamment dans les parois bactériennes, ce qui implique que cette famille d'enzymes est peut-être impliquée dans la dégradation de la paroi bactérienne et donc dans les mécanismes de défense du termite contre ces bactéries comme c'est le cas chez *A. wheeleri* (He et al. 2013).

C. Domaines de liaison aux glucides (CBM)

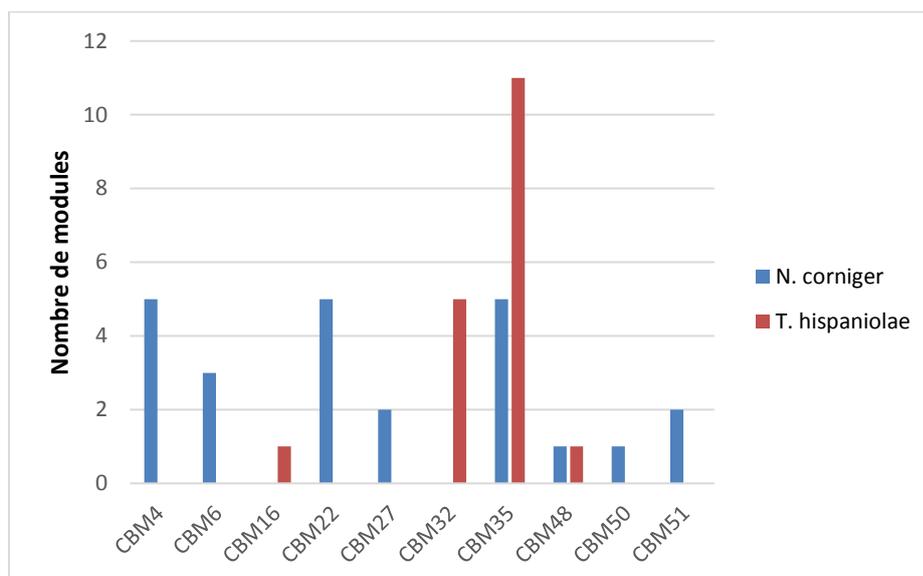


Figure 40 : Histogramme de la répartition des modules de liaison aux glucides dans les 2 banques.

Différentes familles de CBM ont été observées chez *N. corniger* et chez *T. hispaniolae*. Chez *N. corniger*, les CBM majoritairement rencontrés sont les CBM4 (liaison au xylane/glucane/cellulose), les CBM22 (liaison au xylane/glucane) et les CBM35 (liaison au xylane/mannane/galactane). Les CBM4 (liaison au xylane/glucane/cellulose) sont retrouvés associés à des GH10 ou des GH11 dans les deux banques, ils permettent donc un ancrage de l'enzyme à un xylane qui est son substrat, ou bien un ancrage dans l'environnement proche du substrat de l'enzyme. Les CBM22 (liaison au xylane/glucane) sont liés soit à des GH10, soit à des GH5. Les CBM35 (liaison au xylane/mannane/galactane) sont tous liés à des GH26. D'autres familles de CBM sont rencontrées : les CBM6 (liaison au xylane/glucane), les CBM27 (liaison au mannane), les CBM51 (liaison au galactose), puis les CBM48 (liaison au glycogène) et les CBM50 (liaison au chitine/peptidoglycane). On trouve les CBM6 (liaison au xylane/glucane) liés à des GH10 ce qui va aussi permettre la liaison de l'enzyme à un xylane. Les CBM27 (liaison au mannane) retrouvés ne font pas partie d'organisation multimodulaires, mais on peut noter la présence de GH26 sur le contig. Comme les GH26 sont des mannanases, la présence de CBM27 corrèle bien l'activité de l'enzyme et celle du clone puisqu'il a été identifié sur AZCl-galactomannane. Des CBM51 (liaison au galactose) sont retrouvés. Ils ne sont jamais liés à des GH mais leurs gènes sont associés en cluster avec plusieurs gènes codant des GHs. Le seul module appartenant à la famille des CBM48 (liaison

au glycogène) n'est pas lié à un autre module protéique. Sur le clone, on trouve une GH44. Enfin, deux CBM50 (liaison au chitine/peptidoglycane) sont associés sur la même séquence protéique. Sur le contig on trouve une GH5 et une CE4. Toutes ces activités vont dans le sens d'une hydrolyse de la chitine chez *N. corniger*, ce qui est en accord avec les activités trouvées après annotation (figure 40, tableau 9 A et B). Tous ces CBM semblent corrélés avec les activités des GHs qui ont été trouvées chez ce termite et tous sont impliqués dans la liaison aux polysaccharides de la paroi végétale.

Chez *T. hispaniolae*, les CBM majoritairement rencontrés sont les CBM35 (liaison au xylane/mannane/galactane). Ils sont retrouvés tous associés à des GH26 de la même façon que chez *N. corniger*. Ensuite, la famille CBM32 (liaison au lactose/galactose) est la plus rencontrée. Ils sont toujours associés à des GH43. Certaines fois, il s'agit de protéines contenant 4 modules (2 GH43 et 2 CBM32) mais pas toujours dans le même ordre, association déjà rapportée dans la littérature (Mewis et al. 2016). Enfin, les CBM48 (liaison au glycogène) et les CBM16 (liaison au cellulose/glucomannane) sont présents. Un seul module de la famille des CBM48 (liaison au glycogène) est retrouvé lié à une GH13 qui est une pullulanase. Le CBM permet donc à la GH de se fixer sur son substrat. Pour finir, un seul CBM16 (liaison au cellulose/glucomannane) est retrouvé. Il est présent sur une protéine trimodulaire CE3-GH5-CBM16 (figure 40, tableau 10).

Le CBM16 de cette protéine peut se fixer sur la cellulose ou sur le glucomannane. La partie GH5 de la protéine peut hydrolyser, notamment, de la cellulose ou du mannane. Il peut donc y avoir fixation sur le substrat qui sera dégradé par la suite. Dans un contexte d'une paroi végétale plus ou moins intacte, la protéine peut se fixer sur un polysaccharide et dégrader les substrats l'entourant. L'activité du module CE3 est connue pour hydrolyser les groupements acétyles des xylanes. D'autre part, la partie GH5 peut exercer une activité glucanase ou endo-xylanase. On peut donc imaginer que cette protéine peut se fixer sur la cellulose, par exemple, par son CBM16 pour ensuite retirer les acétylations présentes sur un arabinoxylane environnant grâce à la CE3 avant d'hydrolyser la chaîne principale de ce dernier avec le module GH5 pour libérer des xylooligosaccharides ramifiés.

A

clone	substrat testé	annotation taxonomique	annotation CAZy
cor161	AZO-Cel	<i>Treponema sp.</i>	GH5_4, CBM50-CBM50, CE4
cor187	AZO-Cel	<i>Treponema sp.</i>	GH5_4
cor27	AZO-Cel	<i>Fibrobacter sp.</i>	GH11-CBM4, CBM35-GH26
cor63	AZO-Cel/bglc	<i>Treponema sp.</i>	GH5_4
cor65	AZO-Cel/bglc	<i>Treponema sp.</i>	GH5_4
cor67	AZO-Cel/bglc	<i>Treponema sp.</i>	GH5_4
cor70	AZO-Cel/bglc	<i>Treponema sp.</i>	GH5_4, GH13_31
cor71	AZO-Cel/bglc	<i>Fibrobacter sp.</i>	GH8 (x2), GT2-GT2, GT2
cor72	AZO-Cel/bglc	<i>Treponema sp.</i>	GH5_4
cor74	AZO-Cel/bglc	<i>Treponema sp.</i>	GH5_4
cor157	AZO-cel/AZCL-Xyloglc	<i>Treponema sp.</i>	GH5_4
cor173	AZO-cel/AZCL-Xyloglc	<i>Treponema sp.</i>	GH5_4
cor183	AZO-cel/AZCL-Xyloglc	<i>Treponema sp.</i>	GH5_10
cor107	AZO-Cel/AZCL-Xylan	<i>Treponema sp.</i>	GH5_4
cor222	AZO-Cel/AZCL-Xylan	<i>Fibrobacter sp.</i>	GH5_4, GH11, GT2
cor22	AZCL-gal/AZCL-Xylan	<i>Fibrobacter sp.</i>	GH10, GH11-CBM4, CBM35-GH26, CE2
cor78	AZCL-bglc/AZCL-Xylan	<i>Treponema sp.</i>	GH5_4-CBM22, GH73
cor60	AZCL-b-Gln/AZCI-Xyl/AZO-Cell	<i>Treponema sp.</i>	GH5_4-CBM22, GH73
cor79	AZCL-b-Gln/AZCI-Xyl/AZO-Cell	<i>Treponema sp.</i>	GH5_4-CBM22, GH73
cor83	AZCL-b-Gln/AZCI-Xyl/AZO-Cell	<i>Fibrobacter sp.</i>	GH5_4 (x3), GH11, GH74, CBM51, PL10_1
cor93	AZCL-b-Gln/AZCI-Xylgln/AZO-Cell	<i>Treponema sp.</i>	GH5_4 (x2)
cor95	AZCL-b-Gln/AZCI-Xyl/AZO-Cell	<i>Treponema sp.</i>	GH5_4-CBM22, GH73
cor98	AZCL-b-Gln/AZCI-Xylgln/AZO-Cell	<i>Treponema sp.</i>	GH44
cor33	AZCI-b-Gln/AZCI-Galman/AZCI-Xylgln/AZO-cell	<i>Pilobacter Propionicus</i>	GH5_4

B

clone	substrat testé	annotation taxonomique	annotation CAZy
cor 2	AZCL-Ara	<i>Fibrobacter sp.</i>	GH30, GH43, GH5 (x3), CBM51
cor 15	AZCI-Galman/AZCI-Xylan	<i>Fibrobacter sp.</i>	GH30, GH43, GH5 (x3), CBM51
cor 3	AZCI-Ara/AZCI-Galman/AZCI-Xylan	<i>Fibrobacter sp.</i>	GH30, GH43, GH5 (x3)
cor 4	AZCI-Ara/AZCI-Galman/AZCI-Xylan	<i>Fibrobacter sp.</i>	GH30, GH43, GH5 (x3), CBM51
cor 6	AZCI-Ara/AZCI-Galman/AZCI-Xylan	<i>Fibrobacter sp.</i>	GH30, GH43, GH5 (x3), CBM51
cor 10	AZCI-Ara/AZCI-Galman/AZCI-Xylan	<i>Fibrobacter sp.</i>	GH30, GH43, GH5 (x3), CBM51
cor 425	X-Ara	<i>Fibrobacter sp.</i>	GH30, GH43, GH5 (x3), CBM51
cor 5	AZCL-Ara	<i>Fibrobacter sp.</i>	GH11, GH5 (x3), GH74 (x2), GH43, GH43-CBM6, PL1
cor 1	AZCI-Galman/AZCI-Xylan	<i>Fibrobacter sp.</i>	GH5 (x2), GH30, GH74 (x2), GH43
cor 9	AZCL-Ara	<i>Fibrobacter sp.</i>	GH53, PL11, GT4
cor 369	X-Xyl	<i>Treponema sp.</i>	GH3
cor 370	X-Xyl	<i>Treponema sp.</i>	GH3
cor 372	X-Xyl	<i>Treponema sp.</i>	GH3 (x2)
cor 368	X-Xyl	<i>Treponema sp.</i>	GH1, GH94
cor 367	X-Xyl	<i>Treponema sp.</i>	GH43, GH51, GH120, CE1 (x2)
cor 427	X-Ara	<i>Treponema sp.</i>	GH51
cor 431	X-Ara	<i>Treponema sp.</i>	GH51
cor 432	X-Ara	<i>Treponema sp.</i>	GH51
cor 433	X-Ara	<i>Treponema sp.</i>	GH51
cor 434	X-Ara	<i>Treponema sp.</i>	GH51
cor 436	X-Ara	<i>Treponema sp.</i>	GH2, GH51
cor 430	X-Ara	<i>Treponema sp.</i>	GH43, GH51
cor 426	X-Ara	<i>Treponema sp.</i>	GH51, GH127
cor 429	X-Ara	<i>Treponema sp.</i>	GH4, GH51, GT2
cor 435	X-Ara	<i>Treponema sp.</i>	GH27, GH31, GH51
cor 428	X-Ara	<i>Treponema sp.</i>	GH2, GH51, GH53, GH88, GT5
cor 437	X-Cel	ND.	GH1, GH94

cor 438	X-Cel	ND	GH1, GH94
cor 43	AZCL-b-Gln	<i>Fibrobacter sp.</i>	GH8 (x2), GT2, GT4
cor 45	AZCL-b-Gln/AZCL-Xylan/AZO-Cell	<i>Treponema sp.</i>	GH5
cor 356	AZCL-Xyloglucan/AZO-Cell	<i>Treponema sp.</i>	GH5
cor 355	AZCL-Xyloglucan/AZO-Cell	<i>Treponema sp.</i>	GH3, GH5
cor 44	AZCL-b-Gln/AZCL-Xyloglucan	<i>Treponema sp.</i>	GH5 (x2), GH43, GT2
cor 69	AZCL-b-Gln/AZCL-Xyloglucan/AZO-Cell/X-Xyl	<i>Treponema sp.</i>	GH43, GH44, GH51
cor 357	AZCL-Xyloglucan/AZO-Cell	<i>Treponema sp.</i>	GH5, GH18 (x2), GH20, GH94
cor 358	AZCL-b-Gln/AZCL-Xyloglucan/AZO-Cel	<i>Treponema sp.</i>	GH44, CBM48
cor 14	AZCL-Galman	<i>Treponema sp.</i>	GH26
cor 24	AZCL-Galman	<i>Treponema sp.</i>	GH26, CBM27
cor 28	AZCL-Galman	<i>Treponema sp.</i>	GH5, GH26, GH31, GH42, GH74, GH95, CBM27, GH43-CBM6
cor 249	AZCL-Xylan	<i>Fibrobacter sp.</i>	GH11-CBM4, CBM35-GH26, GT2 (x2)
cor 247	AZCL-Xylan/X-Xyl	<i>Treponema sp.</i>	CBM22-GH10, GH30 (x2), GH31, GH43 (x3), GH43-CBM6, GH51 (x2), GH120
cor 29	AZCL-Galman/AZCL-Xylan	<i>Fibrobacter sp.</i>	GH11, GH26, GT2
cor 13	AZCL-Galman/AZCL-Xylan	<i>Fibrobacter sp.</i>	GH10-CBM4, CBM35-GH26
cor 38	AZCL-Galman/AZCL-Xylan	<i>Fibrobacter sp.</i>	GH10, GH10-CBM4, CBM35-GH26
cor 336	AZCL-Galman/AZCL-Xylan	<i>Fibrobacter sp.</i>	GH10, GH10-CBM4, CBM35-GH26

Tableau 9 A et B : Annotation fonctionnelle et taxonomique de la banque *N. corniger*. A : Annotation faite par mes soins. B : Annotation faite par Grégory Arnal durant sa thèse. Les clones redondants sont indiqués en gris.

clone	substrat testé	annotation taxonomique	annotation CAZy
his1	AZCL-Galman	<i>Firmicutes</i>	GH5 (x2)
his2	AZCL-Galman	<i>Firmicutes</i>	CBM35-GH26
his3	AZCL-Galman	<i>Firmicutes</i>	GH5_4, CBM35-GH26
his4	AZCL-Galman	<i>Proteobacteria</i>	GH26
his5	AZCL-Galman	<i>Firmicutes</i>	CBM35-GH26
his8	AZCL-Galman	<i>Firmicutes</i>	CBM35-GH26
his9	AZCL-Galman	<i>Firmicutes</i>	GH26, CBM35-GH26
his10	AZCL-Galman	<i>Ascomycetes</i>	AA1 (x2)
his12	AZCL-Galman	<i>Firmicutes</i>	CH5_4, GH130, CBM35-GH26
his14	AZCL-Galman	<i>Firmicutes</i>	GH13, CBM35-GH26
his17	AZCL-Galman	<i>Firmicutes</i>	GH5_4, GH130, CBM35-GH26
his19	AZCL-Galman	<i>Firmicutes</i>	GH26, GH26-CE3, CE3
his20	AZCL-Galman	<i>Firmicutes</i>	CBM35-GH26
his23	AZCL-Galman	<i>Firmicutes</i>	CBM35-GH26
his25	AZCL-Galman	<i>Firmicutes</i>	CE3-GH5-CBM16
his68	AZCL-Xyloglc	<i>Firmicutes</i>	GH1, GH5_4, GH43-CBM32-GH43-CBM32, GH130
his71	AZCL-Xyloglc	<i>Firmicutes</i>	GH3, GH5_4, GH29, GH31
his72	AZCL-Xyloglc	<i>Firmicutes</i>	GH5_4 (x2), GH9, GH27
his74	AZCL-Xyloglc	<i>Firmicutes</i>	GH3, GH5_4, GH29
his77	AZCL-Xyloglc	<i>Bacteroidetes</i>	GH3, GH5_4 (x2), GH9, GH26 (x2), GH27, GH130
his83	AZCL-Xyloglc	<i>Bacteroidetes</i>	GH3, GH5_4 (x2), GH9, GH26 (x2), GH27, GH130
his65	AZCL-Xylan	<i>Bacteroidetes</i>	GT3-GT35, GT4
his37	AZO-cel	<i>Bacteroidetes</i>	GH5_4 (x2), GH27
his40	AZO-cel	<i>Firmicutes</i>	GH9
his43	AZO-cel	<i>Bacteroidetes</i>	GH5_4 (x2), GH9, GH27
his44	AZO-cel	<i>Firmicutes</i>	GH5_4 (x2), GH27, GH97
his47	AZO-cel	<i>Fibrobacter</i>	GH5_2, GH23, GT2
his52	AZO-cel	<i>Bacteroidetes</i>	GH3, GH5_4 (x2), GH9, GH26 (x2), GH27, GH33 (x2), GH130

his61	AZO-cel	<i>Bacteroidetes</i>	GH3, GH5_4 (x2), GH9, GH26 (x2), GH27, GH130
his62	AZO-cel	<i>Fibrobacter</i>	GH5_2, GT2, GT2-GT2
his63	AZO-cel	<i>Bacteroidetes</i>	CH5_4, GH27
his184	X-ara	<i>Firmicutes</i>	GH31, GH43 (x2), GH51, GH88
his186	X-ara	<i>Firmicutes</i>	GH31, GH35, GH51 (x2), GH88, GH127 (x2), PL12_1
his187	X-ara	<i>Spirochaetes</i>	GH51
his188	X-ara	<i>Firmicutes</i>	GH31 (x2), GH43 (x2), GH51, GH88
his189	X-ara	<i>Actinobacteria</i>	GH39, GH51
his190	X-ara	<i>Firmicutes</i>	GH31, GH35, GH43-GH43-CBM32-CBM32, GH51 (x2), GH88, GH127 (x2), PL12_1
his191	X-ara	<i>Firmicutes</i>	GH31, GH35, GH43, GH51 (x2), GH88, GH127 (x2), PL12_1
his192	X-ara	<i>Actinobacteria</i>	GH39, GH51
his194	X-ara	<i>Firmicutes</i>	GH31, GH35, GH51 (x2), GH88, GH112, GH127 (x2), PL12_1
his199	X-ara	<i>Firmicutes</i>	GH51
his101	X-xyl	<i>Bacteroidetes</i>	GH23, GH123
his109	X-xyl	<i>Firmicutes</i>	GH3
his117	X-xyl	<i>Actinobacteria</i>	GH13_18, GH120
his120	X-xyl	<i>Sinergistetes</i>	GH2, GH3
his121	X-xyl	<i>Actinobacteria</i>	GH3, GH13_13, CBM48-GH13
his124	X-xyl	<i>Firmicutes</i>	GH29, GH51, GH112
his138	X-xyl	<i>Firmicutes</i>	GH3
his220	X-cel	<i>Firmicutes</i>	GH2, GH3
his221	X-cel	<i>Firmicutes</i>	GH1
his223	X-cel	<i>Firmicutes</i>	CBM35-GH26
his32	bGlc/Cel	<i>Fibrobacter</i>	GH5_4, GT4
his42	XylGlc/Cel	<i>Bacteroidetes</i>	GH5_4 (x2), GH9, GH27, GH130
his53	XylGlc/Cel	<i>Bacteroidetes</i>	GH3, GH5_4 (x2), GH9, GH26 (x2), GH27, GH33 (x2), GH130
his56	XylGlc/Cel	<i>Fibrobacter</i>	GT4
his64	XylGlc/Cel	<i>Bacteroidetes</i>	GH3, GH5_4 (x2), GH9, GH26 (x2), GH27, GH130

his171	X-Ara/X-Xyl	<i>Actinobacteria</i>	GH1 (x2)
his15	bGlc/Gal/Cel	<i>Bacteroidetes</i>	GH3, GH5_4 (x2), GH9, GH26 (x2), GH27, GH130
his24	bGlc/Gal/Xylglc	<i>Bacteroidetes</i>	GH5_4, GH27
his26	bGlc/Xylglc/Cel	<i>Firmicutes</i>	GH1, GH5_4, GH43-CBM32 (x2), GH130
his27	bGlc/Xylglc/Cel	<i>Bacteroidetes</i>	GH5_4 (x2), GH9, GH27
his30	bGlc/Xylglc/Cel	<i>Bacteroidetes</i>	GH3, GH5_4 (x2), GH9, GH26 (x2), GH27, GH130
his34	bGlc/Xylglc/Cel	<i>Bacteroidetes</i>	GH3, GH5_4 (x2), GH9, GH26 (x2), GH27, GH33 (x2), GH130
his49	Xylglc/Cel/X-Cel	<i>Bacteroidetes</i>	GH3, GH5_4 (x2),GH5_7, GH9, GH26 (x2), GH27, GH130
his21	bGlc/Xylglc/Cel/Gal	<i>Bacteroidetes</i>	GH3, GH5_4 (x2),GH5_7, GH9 (x2), GH26 (x2), GH27, GH130
his28	bGlc/Xylglc/Cel/X-Ara	<i>Firmicutes</i>	GH1,GH5_4, GH43, GH51 (x2), GH88, GH127 (x2), GH130, GH43-CBM32-GH43-CBM32,PL12_1

Tableau 10 : Annotation fonctionnelle et taxonomique de la banque *T. hispaniolae*.

V. Comparaison des microbiotes

Lors de l'annotation fonctionnelle, une annotation taxonomique est réalisée dans le but d'identifier l'organisme d'origine des gènes mis en évidence. Cependant il est rare de pouvoir pousser l'annotation taxonomique jusqu'à l'identification de l'espèce, les séquences étant relativement divergentes de celles présentes dans les bases de données car il existe encore peu de génomes de référence pour les termites (Huson et al. 2007). Chez *N. corniger* on retrouve très majoritairement deux genres bactériens : *Treponema* et *Fibrobacter* (tableau 9 A et B). Ces genres appartiennent respectivement aux embranchements des *Spirochaetes* et des *Fibrobacteres*. Ces résultats sont en accord avec la publication de He et coll. (He et al. 2013), dans laquelle les auteurs ont comparé le microbiote des termites *N. corniger* et *A. weheeleri*. Ils ont montré que les *Spirochaetes* et *Fibrobacteres*, font partie des embranchements spécifiques des termites supérieurs et qu'ils sont très majoritaires chez *N. corniger*. Les résultats que nous avons obtenus avec une méthode d'analyse différente sont en accord avec ceux de l'étude de He et collaborateurs. Par contre il est intéressant de constater que le microbiote de ce termite n'est pas très diversifié, ce qui peut être lié au fait que *N. corniger* est un termite xylophage strict. La diversité de l'alimentation étant moins importante, la collection d'enzymes de deux genres bactériens seulement peut être suffisante pour digérer ces substrats.

Chez *T. hispaniolae* les deux embranchements bactériens majoritairement retrouvés sont les *Firmicutes* et les *Bacteroidetes*. Contrairement au microbiote de *N. corniger*, d'autres embranchements sont également présents : les *Actinobacteria*, les *Fibrobacteres*, les *Proteobacteria*, les *Ascomycetes*, les *Spirochaetes* et les *Synergistetes* (tableau 10). Ces deux embranchements principaux sont retrouvés chez les termites supérieurs et inférieurs ainsi que dans le microbiote d'autres animaux (bovins, ovins, humains...) (He et al. 2013 ; Flint et al. 2008). Les *Bacteroidetes* sont aussi retrouvés dans le sol. Les *Actinobacteria* et les *Fibrobacteres* étant les embranchements spécifiques des termites supérieurs, il était attendu de les retrouver (He et al. 2013). Le fait que *T. hispaniolae* soit un termite humivore et xylophage explique probablement une plus grande diversité du microbiote liée à la grande diversité de substrats.

Dans des études antérieures comme celle sur le microbiote de *A. weheeleri* (He et al. 2013), qui est un termite se nourrissant sur des excréments de bovins, on retrouve beaucoup d'espèces différentes appartenant à l'embranchement des *Firmicutes*. On retrouve également des *Spirochaetes* et quelques bactéries appartenant à d'autres embranchements tels que les *Synergistetes* et plus rarement les *Bacteroidetes*. Donc, le microbiote est ici composé de bactéries appartenant aux embranchements spécifiques des termites supérieurs ainsi qu'à d'autres embranchements. On peut donc penser que, comme pour *T. hispaniolae*, le microbiote est plus diversifié que celui de *N. corniger* en raison d'une alimentation plus variée.

Dans une étude récente (Rossmassler et al. 2015), six espèces de termites supérieurs appartenant à quatre groupes différents ont été étudiées : *Nasutitermes corniger* et *Microcerotermes parvus* qui sont des termites xylophages, *Cornitermes sp.* se nourrissant de litière, *Termes hospes* et *Neocapritermes taracua* qui sont humivores, et enfin, *Cubitermes ugandensis* qui se nourrit sur le sol. Dans l'un des compartiments du tractus digestif des termites se nourrissant de bois et de litière, l'embranchement principalement retrouvé est celui des *Spirochaetes*. Cet embranchement est retrouvé en faible proportion chez les termites se nourrissant d'humus et de sol. Les embranchements *TG3* et *Fibrobacteres* ne sont retrouvés que chez les termites se nourrissant de bois et de litière. Dans un autre compartiment, on trouve très majoritairement des *Firmicutes* dans les six espèces étudiées. Dans certaines espèces, on y retrouve aussi des *Spirochaetes* et des *Actinobacteria*. Dans un troisième compartiment le microbiote est plus diversifié avec cependant une grande proportion de *Bacteroidetes*.

Dans notre étude, les embranchements *Firmicutes*, *Actinobacteria* et *Bacteroidetes* ne sont pas retrouvés chez *N. corniger*. Nos annotations taxonomiques ont uniquement été faites à partir des données de métagénomique fonctionnelle alors que dans l'étude de Rossmassler les données taxonomiques viennent de métagénomique 16S et de RNAseq, qui est une méthode plus adaptée à l'identification taxonomique que la métagénomique. Cela explique pourquoi nous avons trouvé une plus faible diversité dans le microbiote de *N. corniger* que dans l'étude de Rossmassler.

Dans l'étude de Rossmassler, les embranchements trouvés pour des termites humivores correspondent à ceux que nous avons identifiés pour *T. hispaniolae*.

Nos résultats, comparés à d'autres études sur des termites supérieurs, vont dans le sens d'un microbiote totalement adapté à l'alimentation du termite (Rossmassler et al. 2015). Ces résultats sont en accord avec une étude de Hongoh et collaborateurs (Hongoh et al. 2005) dans laquelle ils ont montré que le microbiote des termites a co-évolué avec son hôte.

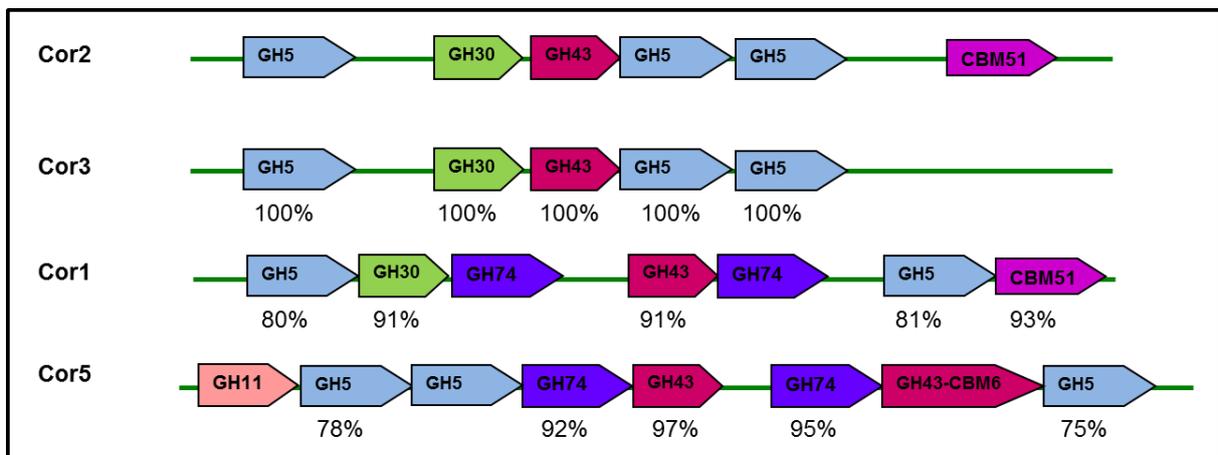
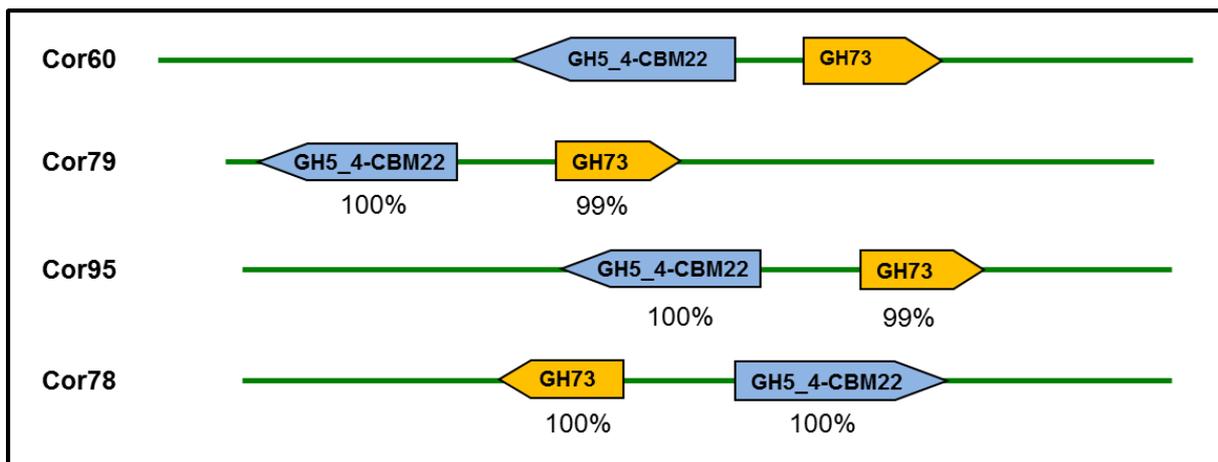
VI. « Polysaccharide utilization locus » (PUL)

Les PULs sont des regroupements de gènes codant des protéines dont les activités permettent de dégrader totalement et spécifiquement un type de polysaccharide. Le premier système décrit et qui est le plus étudié est le système Sus (« Starch utilization system ») qui permet la reconnaissance, la dégradation et l'internalisation de l'amidon et de ses produits de dégradation par la bactérie (cf introduction p. 42). Les systèmes fonctionnant sur le même principe et spécifiques à d'autres polysaccharides sont dits Sus-like. Les bactéries les plus connues pour posséder des PULs tels que le système SUS sont les *Bacteroidetes* (Martens et al. 2009 ; Martens et al. 2011 ; Bolam et Koropatkin 2012).

Lors de l'annotation fonctionnelle, des protéines appartenant à des systèmes Sus-like ont été identifiées dans la banque de *T. hispaniolae* sur une dizaine de clones. Les protéines similaires à SusD et/ou SusC ont été trouvées. SusD est une protéine ancrée à la membrane externe des bactéries et responsable de la reconnaissance du polysaccharide. SusC est une protéine transmembranaire située dans la membrane externe de la bactérie et permet l'internalisation des produits de première hydrolyse du polysaccharide (Martens et al. 2009). Comme l'embranchement des *Bacteroidetes* est l'un des plus présents dans le microbiote de *T. hispaniolae*, il semble logique de retrouver des protéines appartenant à des systèmes Sus. Par contre, aucune protéine de ce type n'a été identifiée dans le métagénome de *N. corniger*, qui est dépourvu de bactéries appartenant à l'embranchement des *Bacteroidetes*. L'hypothèse est que si de telles séquences existent dans le métagénome de *N. corniger*, elles sont peut-être trop divergentes de celles référencées dans les banques de données et donc elles n'ont pas été identifiées. Cependant plusieurs exemples de PUL ont pu être mis en

évidence chez d'autres termites comme par exemple le termite champignonniste *P. militaris*, donc ce résultat est plutôt surprenant (Bastien et al. 2013).

D'autre part, certaines associations de modules ont été observées sur différents clones de *T. hispaniolae* sans pour autant présenter des séquences analogues aux protéines Sus (GH3, GH5_4 (x2), GH9, GH26 (x2), GH27, GH130 ou encore GH31, GH35, GH51 (x2), GH88, GH112, GH127 (x2), PL12_1) et sont majoritairement en adéquation avec les substrats chromogéniques sur lesquels les clones ont été identifiés (tableau 10, figure 41). Par contre, ces associations ne sont pas référencées dans la PUL database (<http://www.cazy.org/PULDB/>). Par exemple, dans la PULDB, l'association GH5, GH26, GH27, GH130 n'a été observée que chez *Flavobacterium johnsoniae* qui appartient à l'embranchement des *Bacteroidetes*. Les *Bacteroidetes* étant connues pour posséder des PULs, il est possible que ceux identifiés dans notre étude ne soient pas encore référencés.



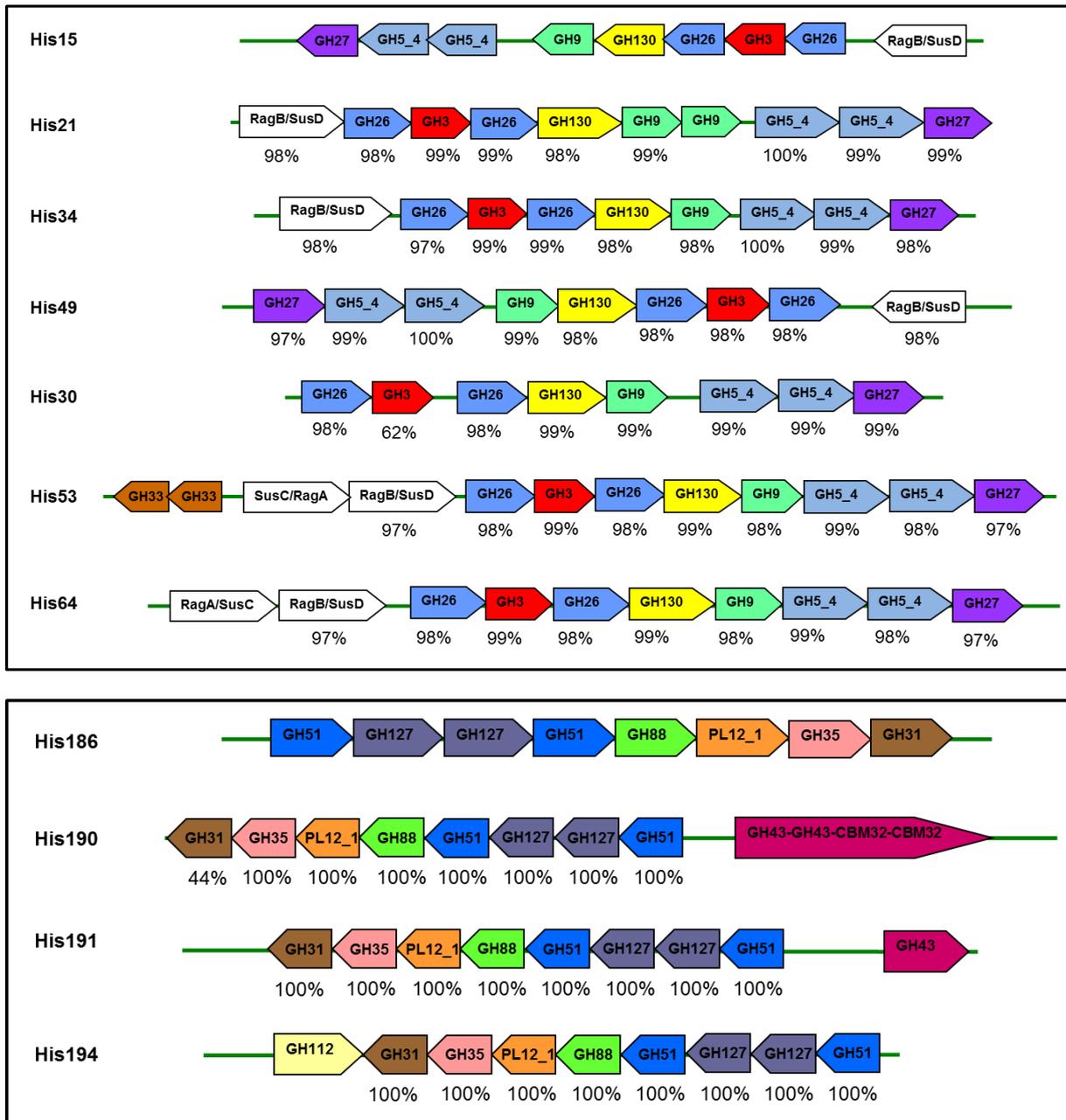


Figure 41 : Représentation schématique de regroupements de gènes chez *T. hispaniolae* (His) et *N. corniger* (Cor). Le sens de la flèche indique l'orientation du gène. Les pourcentages indiqués correspondent au pourcentage d'identité de séquence par rapport au groupe de gènes en haut de chaque encadré noir.

Chez *N. corniger*, des regroupements de CAZymes ont été observés sur peu de clones (GH5_4-CBM22, GH73 ou encore GH5, GH30, GH43) (tableau 9 A et B, figure 41). Seule l'association GH5, GH30, GH43 est retrouvée chez des espèces de *Bacteroidetes* dans la PUL

database. De la même manière que pour *T. hispaniolae*, il est possible que les associations observées ici n'aient pas encore été décrites.

VII. Conclusion – Discussion

Dans cette étude, nous avons pu montrer qu'il existe bien une corrélation entre les activités identifiées et l'alimentation des deux termites étudiés.

En effet *N. corniger*, qui est un termite xylophage strict, se nourrit exclusivement de bois et notamment de bois sain dont la paroi végétale est intacte et difficile à dégrader (Somerville et al. 2004). Le crible fonctionnel primaire sur substrat chromogénique semblait indiquer une majorité d'enzymes à activité endoglycosidase pour dégrader principalement les polysaccharides. Le microbiote intestinal de ce termite semble produire un arsenal d'enzymes capables de dégrader efficacement tous les polysaccharides qui composent la paroi végétale. Par rapport au crible fonctionnel sur substrat chromogénique, l'annotation fonctionnelle indique que la différence de proportion d'activités endoglycosidase et exoglycosidase n'est pas aussi importante, ce qui est logique puisque qu'il faut un mélange de ces activités pour dégrader totalement les polysaccharides de la paroi végétale. D'autre part les cribles sont faits avec certains substrats choisis (pour trouver certaines activités en particulier), qui apportent donc un biais dans les activités identifiées. Ensuite, les fosmides contiennent des inserts de grande taille (40kb) qui permettent de préserver l'organisation des métagénomés d'origine. Il est donc logique de trouver beaucoup de GH par l'annotation fonctionnelle mais par contre sans aucune garantie de leur réelle activité.

Les résultats de l'analyse fonctionnelle et taxonomique corrélient avec d'autres études faites sur la même espèce de termite et d'autres appartenant au même genre.

Chez *T. hispaniolae*, un termite humivore et xylophage, les cribles fonctionnels primaires laissaient penser que la majorité des enzymes produites par le microbiote intestinal présente des activités exoglycosidase. Ces résultats allaient dans le sens d'une alimentation de bois en partie dégradé et de matière en décomposition. L'annotation fonctionnelle des enzymes présentes dans le métagénome du termite montre qu'elles sont aussi composées en majeure partie d'enzymes à activité exoglycosidase mais la tendance tend à s'équilibrer.

Pour les mêmes raisons que pour la banque métagénomique de *N. corniger*, il est logique de trouver beaucoup plus de GH lors de l'annotation fonctionnelle que lors des cribles car tous les substrats existant dans la nature ne sont pas criblés. Par contre on trouve moins de GH actives sur les pectines, ce qui pourrait indiquer que les protéines nécessaires sont codées par d'autres génomes ou dans d'autres zones du génome puisque les activités pectinases n'ont pas été recherchées lors des criblages. A l'inverse, un nombre important de modules actifs contre les polysaccharides de réserve et des protéines glycosylées ont pu être identifiées. Ces substrats sont en grande partie retrouvés dans les racines. Tout cela corrèle donc bien avec un termite se nourrissant au niveau du sol d'humus et de bois partiellement décomposé.

L'étude taxonomique a montré que le microbiote semble spécifique du type d'alimentation du termite ce qui corrèle avec d'autres études comme celles de Hongoh, de Bastien ou de He (Hongoh et al. 2005 ; Bastien et al. 2013 ; He et al. 2013).

Les annotations fonctionnelles ont permis d'identifier d'autres catégories d'enzymes telles que des estérases, des lyases et des modules de liaison au glucides. Ces différentes protéines sont essentielles à la dégradation totale ou partielle de substrats complexes comme ceux retrouvés dans les parois végétales. Ces modules ainsi que des GH sont codés dans des regroupements de gènes, notamment chez *T. hispaniolae* et sont en accord avec les activités des clones dans lesquels ils ont été identifiés.

Enfin, on a pu trouver dans ces métagénomes des gènes codant des protéines ayant des organisations multimodulaires originales qui pourraient être intéressantes à caractériser d'un point de vue fonctionnel et structural, comme par exemple la protéine contenant les modules CE3-GH5-CBM16 identifiée dans la banque de *T. hispaniolae* et qui pourrait être active sur substrats complexes.

Chapitre II : Etude fonctionnelle de GH*, une nouvelle enzyme impliquée dans la dégradation de polysaccharides de la paroi végétale ?

I. Introduction :

La paroi végétale est une structure complexe composée de polysaccharides (cellulose, hémicelluloses et pectines), de lignine et de protéines (Somerville et al. 2004 ; Burton, Gidley, et Fincher 2010 ; Gorshkova, Kozlova, et Mikshina 2013). Il s'agit de la plus abondante source de carbone organique renouvelable de la biosphère.

Les polysaccharides de la paroi végétale sont organisés en un maillage dense et rigide. Leur organisation fait de la paroi végétale une structure difficile à hydrolyser car il faut casser les liaisons existantes entre les différents polysaccharides pour ensuite les dégrader.

Pour maximiser sa valorisation, il est nécessaire de développer de nouveaux biocatalyseurs capables de dégrader efficacement la cellulose et les hémicelluloses qu'elle contient. Et les glycosides hydrolases (GHs) semblent être des biocatalyseurs de choix pour assurer cette dégradation.

Pour ce faire, différentes banques métagénomiques ont été créées dont une à partir du ver de Madagascar *Ponthoscolex corethrurus*. Ce ver est géophage et vit dans des zones humides. Il est connu pour tolérer une grande variation de propriétés physicochimiques du sol et de températures. Il est très efficace pour assimiler de la matière organique de faible qualité et est capable de coloniser des territoires (Lavelle et al. 1987).

Le criblage fonctionnel, sur substrats chromogéniques, de la banque métagénomique de ce ver a permis d'identifier de nombreux clones actifs sur divers substrats représentatifs de la biomasse lignocellulosique allant du pNP-arabinose à l'AZCL-cellulose. Un des 7 clones actifs sur X-cellobioside a particulièrement attiré notre attention (figure 42).

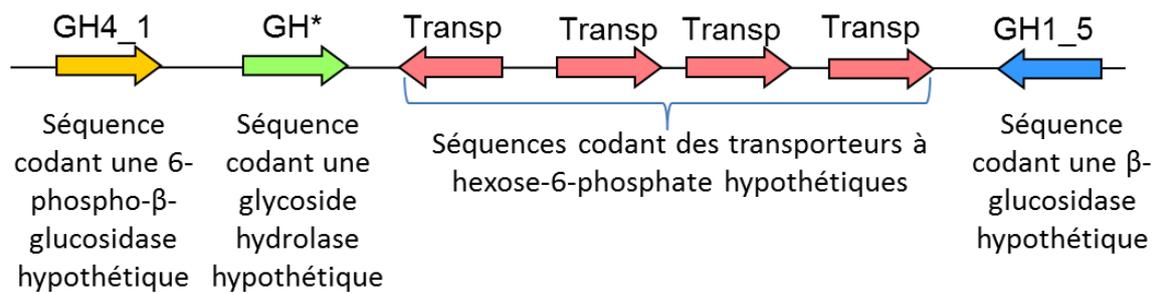


Figure 42 : Représentation schématique du fragment de génome sur lequel se trouve la séquence de GH*. Le sens des flèches indique l'orientation du gène.

L'annotation fonctionnelle a révélé 3 GH hypothétiques dont 2 ont pu être annotées comme appartenant aux familles GH1 et 4 en se basant sur la classification CAZy (<http://www.cazy.org/>) (Lombard et al. 2014). Ces familles présentent majoritairement des activités β-glucosidases et 6-phospho-β-glucosidases respectivement. De façon intéressante, la 3^{ème} GH a été annotée comme étant une GH hypothétique qui n'a pu être assignée à aucune des 135 familles de GH existantes actuellement dans la base de données CAZy. Cette enzyme notée GH* pourrait être une protéine appartenant à une nouvelle famille de GH si son activité glycoside hydrolase peut être démontrée. En tenant compte de l'activité du clone sur lequel l'enzyme a été identifiée, nous pouvons penser que GH* pourrait avoir une activité sur des oligo et/ou des polysaccharides constituant la paroi végétale. Néanmoins, l'activité cellobiohydrolase détectée lors du criblage fonctionnel peut être conférée par une autre GH du fosmide. Il est donc nécessaire de finement caractériser cette enzyme.

En réalisant un alignement de séquences, les meilleurs scores d'alignement, pour des enzymes caractérisées ou annotées, ont été obtenus avec des GH5 mais ayant une identité de séquence avec GH* assez faible (Tableau 11).

Protéine	Organisme	Score	Pourcentage de couverture	E-value	Identité
Protéine hypothétique	<i>Enterobacter absuriae</i>	742	100%	0,0	99%
Protéine périplasmique	<i>Enterobacter cloacae</i>	741	100%	0,0	99%
Protéine périplasmique	<i>Enterobacter</i>	739	100%	0,0	99%
Protéine hypothétique	<i>Enterobacter absuriae</i>	734	100%	0,0	98%
Protéine périplasmique	<i>Enterobacter absuriae</i>	722	100%	0,0	96%
Protéine périplasmique	<i>Enterobacter sp.</i>	722	100%	0,0	96%
Protéine hypothétique	<i>Enterobacter cloacae</i>	721	100%	0,0	96%
Protéine périplasmique	<i>Enterobacteriaceae</i>	721	100%	0,0	96%
Protéine hypothétique	<i>Enterobacter absuriae</i>	720	100%	0,0	96%
Protéine hypothétique	<i>Enterobacter cloacae</i>	720	100%	0,0	96%
Protéine hypothétique	<i>Enterobacter absuriae</i>	719	100%	0,0	96%
GH5	<i>Zymoseptoria tritici</i>	40	51%	4,2	23%
Protéine hypothétique	<i>Dactylellina haptotyla</i>	40	48%	4,2	22%
Protéine hypothétique	<i>Arthrotrichum oligospora</i>	39,7	54%	5,8	17%
GH	<i>Sistotremastrum suecicum</i>	39,7	49%	6,6	20%
Protéine hypothétique	<i>Acromobacter xylosoxidans</i>	36,2	12%	6,7	37%
GH5	<i>Bipolaris victoriae</i>	39,3	38%	7,6	23%
Régulateur de transcription LuxR	<i>Mycobacterium sp.</i>	39,7	46%	7,8	25%
Protéine hypothétique	<i>Dactylellina haptotyla</i>	39,3	54%	8,8	23%
Régulateur de transcription LuxR	<i>Mycobacterium sp.</i>	39,3	46%	9,2	25%

Tableau 11 : Résultats du BLAST de la séquence protéique de GH* contre la banque nr (non redondant). Les résultats sont classés de haut en bas en % d'identité décroissant. Les 363 résidus de GH* (peptide signal inclus) ont été alignés. Le pourcentage de couverture correspond à la taille de la région alignée par rapport à la taille totale de GH*. Le pourcentage d'identité est calculé sur la région alignée.

Cependant, d'après Bernard Henrissat, responsable de la base de données CAZy, cette enzyme ne peut pas être classée dans la famille des GH5 car il lui manque des signatures caractéristiques de GH5, notamment un des deux acides glutamiques catalytiques (celui qui joue le rôle de nucléophile). Sur la figure 43 on peut voir qu'une des signatures de la famille 5 n'est pas conservée chez GH*. En effet, au lieu de retrouver le motif NEP ou NEV, c'est le motif YEP qui est retrouvé dans la séquence de GH*. L'acide glutamique du motif NEP/V chez les GH5 est celui jouant le rôle du résidu catalytique acide/base, cela laisse penser que

l'acide glutamique 117 de GH* pourrait jouer un rôle identique. D'autre part, on peut voir au bas de la figure 43 qu'un autre motif conservé, EWG ou EFG n'est pas retrouvé chez GH*. A la place, on trouve GWG. Or, l'acide glutamique de ce motif est celui jouant le rôle de nucléophile dans les GH5. Cela confirme donc la non-conservation du glutamate jouant le rôle de nucléophile dans la séquence de GH*.

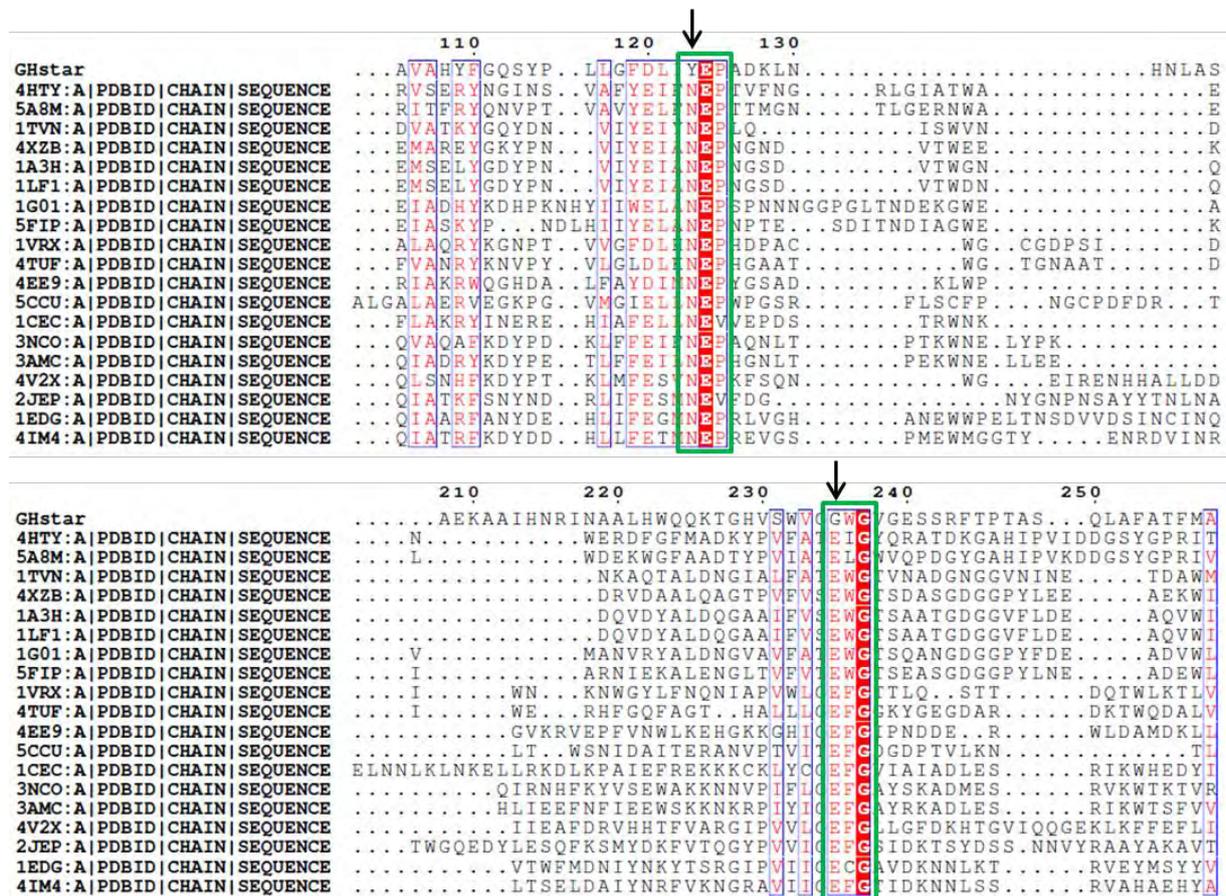


Figure 43 : Parties d'alignement de GH* avec des GH5 caractérisées. Ces parties contiennent les motifs de résidus conservés autour des résidus catalytiques.

D'autre part, cette enzyme a été annotée par l'équipe CAZy comme appartenant au clan des GH-A, auquel appartient aussi la famille 5. Toutes les enzymes de ce clan possèdent un mécanisme par rétention de la configuration absolue du carbone anomérique et les deux résidus catalytiques sont des acides glutamiques (Henrissat et Bairoch 1996), dont un pourrait être absent chez GH*. Comme on peut le voir en figure 44, ce résidu est essentiel car impliqué dans la formation d'un intermédiaire réactionnel glycosyl-enzyme. L'absence de ce résidu pourrait causer l'inactivité de GH*.

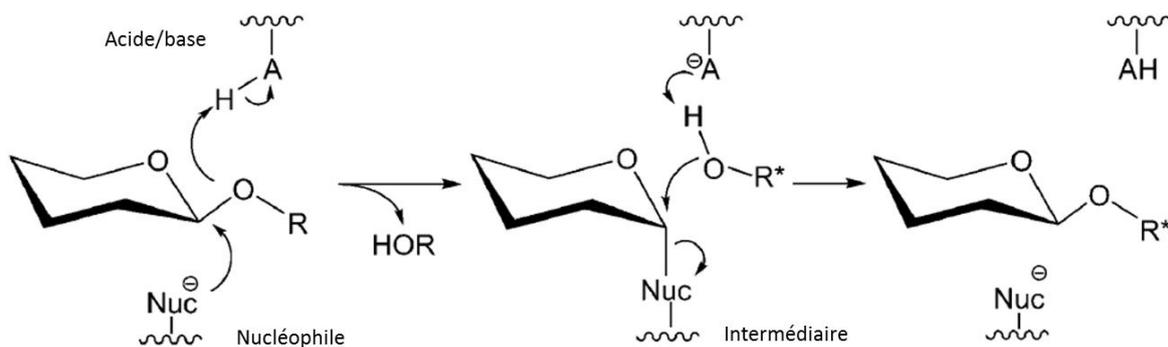


Figure 44 : Schéma présentant le mécanisme d'hydrolyse par rétention de la configuration absolue du carbone anomérique. Résidu acide/base : A ; résidu nucléophile : Nuc.

Une étude fonctionnelle a donc été menée afin de déterminer si GH* est active sur des polysaccharides ou oligosaccharides et si elle peut jouer un rôle dans la dégradation de la biomasse végétale.

II. Détermination des conditions optimales pour l'expression de GH*

A. Sous-clonage

Le fosmide N°405-E6 de la banque métagénomique contenant le gène d'intérêt a été utilisé pour amplifier la séquence de GH*. La taille attendue du gène est de 1 kb et on peut voir sur le gel de la figure 45A que la bande la plus intense des pistes 3 et 4 migre à 1 kb. Il semble donc que la séquence codant GH* ait bien été amplifiée. Le gène a été inséré dans un plasmide pET28a entre les sites de restrictions Nde1 et BamH1. La taille attendue pour le vecteur avec l'insert est de 6,4 kb. Sur le gel de la figure 45B, on voit que la bande de la piste 3 fait 5,5 kb et celle de la piste 4 fait 6,5 kb. La différence de migration peut s'expliquer par une différence d'enroulement du plasmide.

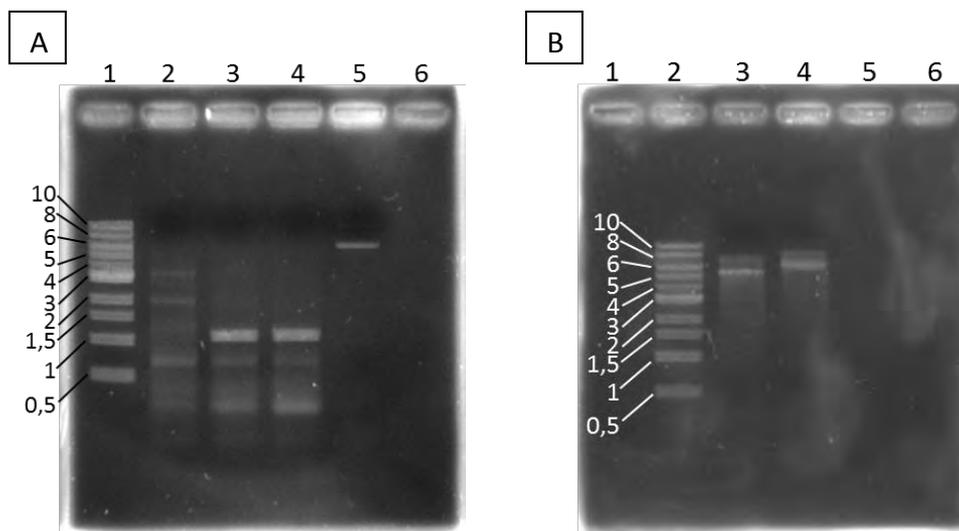


Figure 45 : A. Gel d'agarose après amplification de la séquence codant GH* par PCR. Piste 1 : marqueur de taille (kb) ; piste 2 : témoin négatif ; piste 3 : GH* (PCR avec 50 ng d'ADN matrice) ; piste 4 : GH* (PCR avec 100 ng d'ADN matrice) ; piste 5 : témoin de dépôt (100 ng d'ADN). **B. Gel d'agarose après extraction et purification du plasmide contenant la séquence de GH*.** Piste 2 : marqueur de taille (kb) ; piste 3 : pET28a-GH* issu du clone 1 (à 90 ng/ μ l) ; piste 4 : pET28a-GH* issu du clone 1 (à 117 ng/ μ l).

Les plasmides purifiés ont été envoyés à séquencer et les résultats ont validé le clonage de GH*. Une fois la séquence validée, les tests d'expression ont pu être réalisés.

B. Expression

Afin de trouver les meilleures conditions pour l'expression de GH*, différentes souches d'*E. coli* (C41/DE3, Tuner/DE3 ou BL21/DE3) ont été testées. Les cultures ont été réalisées à 37 °C et deux conditions d'induction à l'IPTG ont été testées : 4 h à 37 °C et 20 h à 16 °C. Les extraits cellulaires contenant les protéines solubles et les fractions insolubles ont été analysés par SDS-PAGE (figure 46). Les résultats montrent que l'induction est très faible dans les C41/DE3 quelles que soient les conditions d'induction. Dans les 2 autres souches l'induction à 37°C pendant 4h avec 0,2 ou 1 mM IPTG entraîne une importante production de protéine recombinante dans la fraction insoluble. Les 2 meilleures conditions d'expression semblent donc être à 16 °C sur la nuit dans les souches BL21/DE3 et Tuner/DE3 car la majorité de la protéine se trouve dans la fraction soluble.

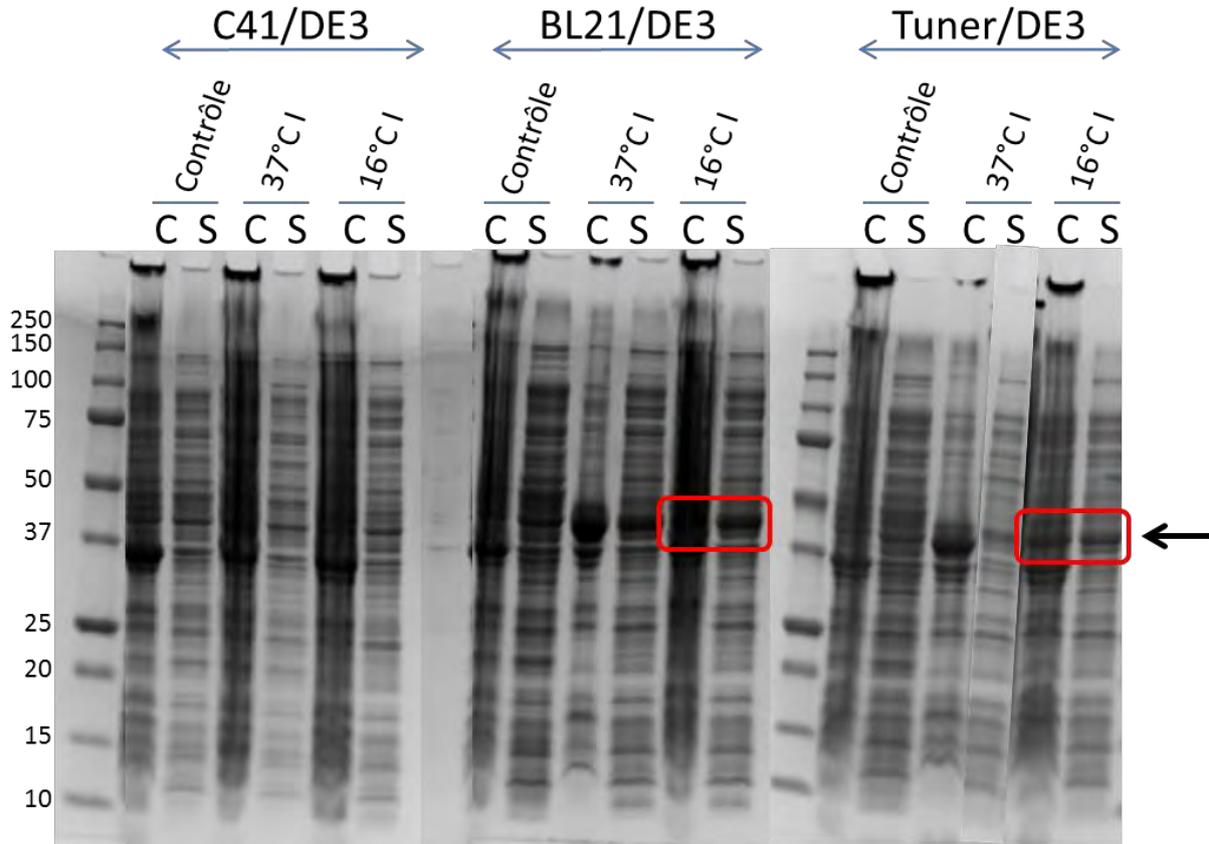


Figure 46 : Gels SDS-PAGE présentant les résultats des tests d'induction. Induit à l'IPTG (I), fraction insoluble (culot, C), fraction soluble (surnageant, S). La flèche montre la migration de GH*. Les encadrés rouges montrent les plus fortes expressions de GH*.

Pour valider le meilleur protocole d'expression de GH* et sélectionner la meilleure souche, un litre de culture des deux souches d'*E. coli* BL21/DE3 et Tuner/DE3 transformées avec le plasmide pET28-GH* a été réalisé. Les deux cultures ont été induites par l'ajout d'1 mM et de 0,2 mM d'IPTG respectivement à 16 °C sur la nuit. Les fractions solubles extraites ont été chargées sur résine de cobalt pour purifier GH* par chromatographie d'affinité. Les fractions d'élution de la chromatographie d'affinité ont été déposées sur gel SDS-PAGE pour quantifier la production de GH* et déterminer quelle souche est la meilleure pour l'expression de GH*. Les résultats des gels sont présentés en figure 47. GH* est éluée dans les fractions à 100 mM d'imidazole et la protéine migre à la masse attendue. La bande dans la 1^{ère} fraction d'élution à 100 mM d'imidazole est un peu plus intense sur le gel correspondant à l'expression dans la souche BL21/DE3. L'intensité de la bande de la 2^{ème} fraction à 100 mM est équivalente dans les deux gels. Les fractions à 100 mM d'imidazole rassemblées ont permis d'obtenir 10mg de protéine pour chaque production.

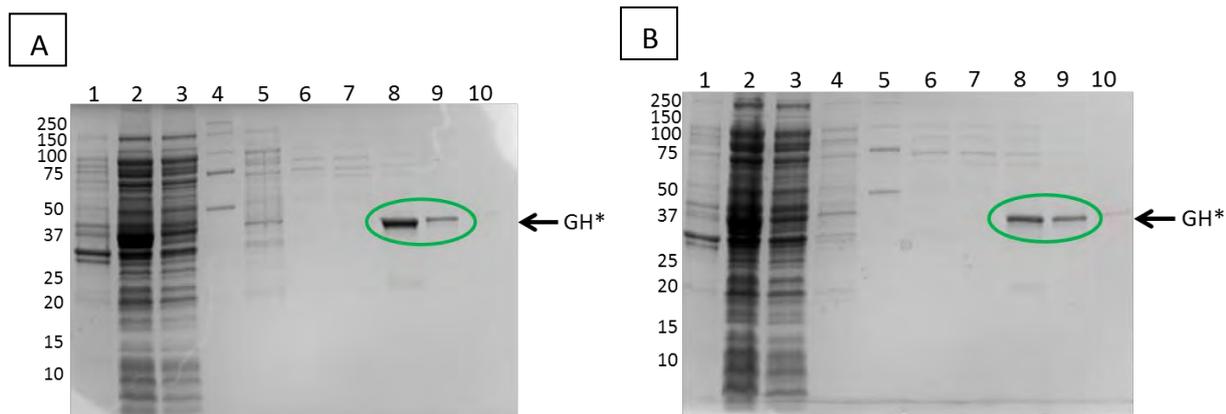


Figure 47 : Gels SDS-PAGE présentant la purification de GH* après expression dans la souche BL21/DE3 (A) et dans la souche Tuner/DE3 (B). Piste 1 : culot après sonication ; piste 2 : surnageant après sonication ; piste 3 : non retenu sur la colonne ; piste 4 : marqueurs de taille (gel A) et lavage avec le tampon sans imidazole (gel B) ; piste 5 : lavage avec le tampon sans imidazole (gel A) et marqueur de taille (gel B) ; piste 6 : lavage à 5 mM imidazole ; piste 7 : lavage à 10 mM imidazole ; pistes 8 et 9 : élution de GH* à 100 mM imidazole ; piste 10 : lavage à 200 mM imidazole. Les bandes correspondant à GH* sont entourées en vert.

Les conditions de culture bactérienne permettant de surexprimer la protéine recombinante GH* ont donc été fixées comme suit : dans la souche d'*E. coli* BL21/DE3, la culture est dans un premier temps réalisée à 37 °C jusqu'à ce que la DO à 600 nm atteigne une valeur de 0,6, la culture est ensuite incubée pendant 20 h à 16 °C en présence de 1 mM d'IPTG pour induire la production de la protéine.

III. Optimisation des conditions de purification

A. Chromatographie d'affinité sur ions métalliques immobilisés

La première purification par chromatographie d'affinité sur ions métalliques immobilisés (IMAC) a été réalisée en « batch » (cf matériels et méthodes) avec 2 ml de résine cobalt (TALON), dans un tampon 20 mM Tris, 300 mM NaCl, pH 8,0. Après lyse des cellules, la fraction soluble est déposée sur la résine. Un lavage avec 20 ml de tampon 20 mM Tris, 300 mM NaCl, pH 8,0 est alors effectué pour retirer tout ce qui ne s'est pas fixé sur la résine. Initialement, un lavage à 5 mM d'imidazole était effectué mais il a été supprimé car il n'apportait rien. Une partie des contaminants est élue lors des deux lavages avec 5 ml de

tampon 20 mM Tris, 300 mM NaCl, pH 8,0 contenant 10 mM d'imidazole (figure 48, pistes 6 et 7). GH* est ensuite éluée avec deux fois 5 ml de tampon 20 mM Tris, 300 mM NaCl, pH 8,0 contenant 100 mM d'imidazole (figure 48, pistes 8 et 9). Toutefois, des contaminants sont co-élués avec la protéine d'intérêt lors de cette étape (figure 48, pistes 8 et 9). Enfin un lavage avec 5 ml de tampon 20 mM Tris, 300 mM NaCl, pH 8,0 contenant 200 mM d'imidazole est effectué afin de retirer tout ce qui est resté fixé sur la résine. Avec ce protocole de purification il semble qu'un peu de GH* soit éluée lors des différents lavages. Cela peut être dû à un mauvais repliement de la protéine qui empêcherait l'étiquette 6-histidine de se fixer au cobalt correctement.

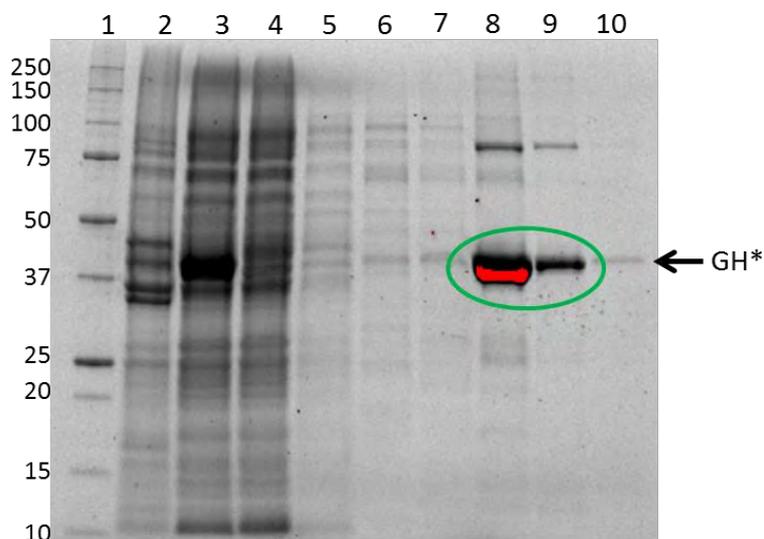


Figure 48 : Gel SDS-PAGE présentant la purification de GH* sur IMAC. Piste 1 : marqueurs de taille ; piste 2 : culot après sonication ; piste 3 : surnageant après sonication ; piste 4 : non retenu sur la colonne ; piste 5 : lavage avec le tampon sans imidazole ; pistes 6 et 7 : lavage à 10 mM ; pistes 8 et 9 : élution de GH* à 100 mM imidazole ; piste 10 : lavage à 200 mM imidazole. L'élution de GH* est entourée en vert.

Nous avons remarqué que lors du passage du tampon d'IMAC à pH 8,0 à celui de cristallisation à pH 5,0 (plus d'informations dans la partie DSF), des agrégats ou des précipités se formaient, probablement dû à l'écart de pH et à un passage trop rapide de l'un à l'autre. Il a donc été décidé de faire l'étape de purification par affinité dans un tampon à pH 7,0 afin de diminuer l'écart de pH entre les tampons. Le tampon choisi est 20 mM HEPES, 100 mM NaCl, pH 7,0.

Une expérience comparative entre les 2 tampons de purification a donc été réalisée et les résultats sont présentés en figure 49. Deux culots de culture ont été resuspendus, l'un dans du tampon Tris (20 mM Tris, 300 mM NaCl, pH 8,0), l'autre dans du tampon HEPES (50 mM HEPES, 100 mM NaCl, pH 7,0). Les mêmes volumes de lavage que précédemment ont été utilisés. Quand on compare les pistes 8 et 9 (figure 49) des 2 gels de la figure 49 à l'aide du logiciel Image Lab (Biorad), il apparaît que les bandes correspondant à GH* ont la même intensité. Il semble donc que la même quantité de protéine soit éluée dans les 2 tampons (environ 10mg). De plus, il y a moins de contaminants de haut poids moléculaire détectés sur le gel correspondant à l'élué à pH7,0 (figure 49B). Le changement de tampon semble avoir un effet positif au regard de la diminution de contaminants co-élués avec GH*. Enfin, la comparaison des pistes 10 des 2 gels (figure 49) semble indiquer qu'il y a moins de GH* éluée à 200 mM imidazole. Par conséquent, il semblerait que moins de protéine reste fixée sur la colonne après l'élué à 100 mM.

Le passage de pH 8,0 en tampon Tris (20 mM Tris, 300 mM NaCl, pH 8,0) à un tampon HEPES (50 mM HEPES, 100 mM NaCl, pH 7,0) permet donc d'avoir une protéine un peu plus pure en sortie d'IMAC.

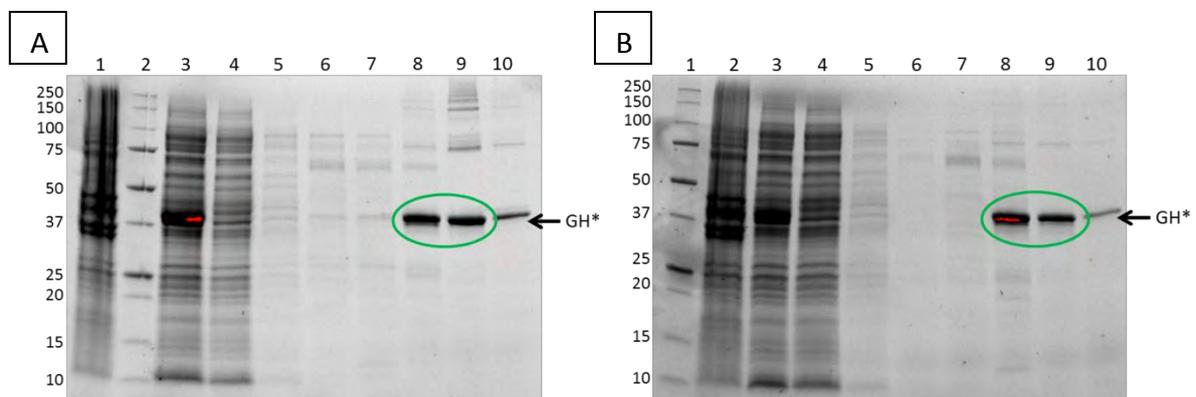


Figure 49 : A. Gel SDS-PAGE présentant les résultats de l'IMAC en tampon Tris à pH8,0. Piste 1 : culot après sonication ; piste 2 : marqueurs de taille ; piste 3 : surnageant après sonication ; piste 4 : non retenu sur la colonne ; piste 5 : lavage avec le tampon sans imidazole ; pistes 6 et 7 : lavage avec 10 mM imidazole ; pistes 8 et 9 : élué de GH* avec 100 mM imidazole ; piste 10 : lavage avec 200 mM imidazole. **B. Gel SDS-PAGE présentant les résultats de l'IMAC en tampon HEPES à pH7,0.** Piste 1 : marqueurs de taille ; piste 2 : culot après sonication ; piste 3 : surnageant après sonication ; piste 4 : non retenu sur la colonne ; piste 5 : lavage avec le tampon sans imidazole ; pistes 6 et 7 : lavage avec 10 mM imidazole ; pistes 8 et 9 : élué de GH* avec 100 mM imidazole ; piste 10 : lavage avec 200 mM imidazole. L'élué de GH* est entourée en vert.

Après les IMAC à pH 8,0 et 7,0 les fractions contenant la protéine d'intérêt sont dialysées dans un tampon 20 mM citrate de sodium, 100 mM NaCl, pH 5,0. La quantité de protéine précipitée semble être beaucoup moins importante dans le boudin de dialyse en passant du tampon HEPES à pH 7,0 au tampon citrate de sodium à pH 5,0 qu'en passant du tampon Tris à pH 8,0 au tampon citrate de sodium à pH 5,0. En fin d'IMAC environ 10 mg de GH* fortement enrichie sont obtenus. Ces résultats confirment l'utilisation du tampon 20 mM HEPES, 100 mM NaCl, pH 7,0 pour la chromatographie d'affinité.

B. Chromatographie d'exclusion-diffusion

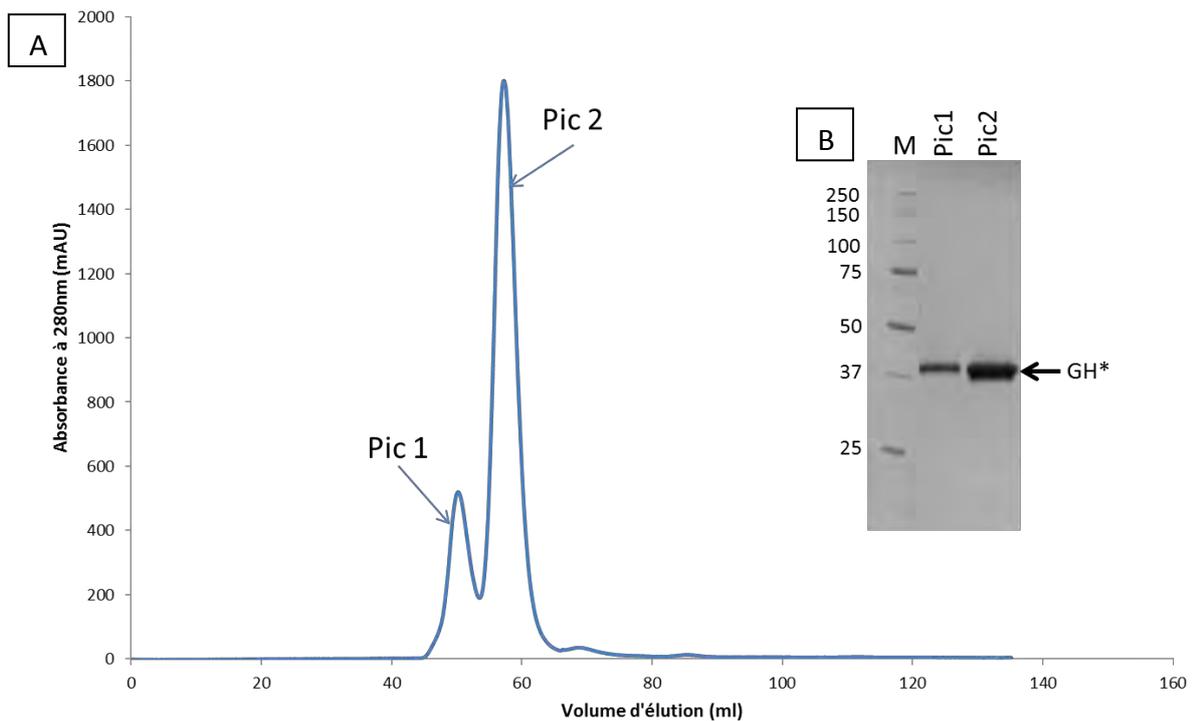
Suite à l'IMAC, les 2 fractions à 100 mM imidazole sont réunies et concentrées jusqu'à un volume final d'environ 2 ml pour être injectées sur colonne de filtration sur gel.

La première purification par chromatographie d'exclusion-diffusion a été réalisée sur une colonne Sephadex S75 préparative 16/60 HiLoad (GE Healthcare) car la protéine possède une masse moléculaire d'environ 40 kDa et que la gamme de séparation de cette colonne va de 3 à 70 kDa (avec une séparation optimale entre 8 et 50 kDa). Le tampon d'élution est composé de 50 mM de phosphate de sodium à pH 7,0. Le chromatogramme est présenté en figure 50A. La protéine éluée avec un volume d'élution de 58 ml correspond à une protéine d'environ 40 kDa en accord avec la masse moléculaire de GH*. Le gel SDS-PAGE (figure 50B) ne présente qu'une seule bande pour la piste correspondant au pic 2. Cela indique donc que la protéine est pure après l'étape de filtration sur gel. Toutefois, on peut voir sur le chromatogramme (figure 50A) qu'une protéine est éluée avec un volume d'élution à la limite du volume mort qui est de 46,5 ml. Sur gel SDS-PAGE, la fraction correspondant au pic 1 du chromatogramme présente une seule bande correspondant à une protéine de 40 kDa environ. Le pic 1 du chromatogramme indique donc probablement la présence de GH* sous forme oligomérique ou agrégée dans les fractions correspondantes.

Afin de caractériser la protéine éluée avec un volume d'élution de 46,5 ml sur S75, la filtration sur gel a été faite en sortir d'IMAC sur colonne S200 préparative 16/60 HiLoad (GE Healthcare) dont la gamme de séparation s'étend de 10 à 600 kDa avec une séparation optimale de 30 à 250 kDa. Les résultats de cette chromatographie sont présentés figure 50C. On peut voir qu'une protéine est éluée avec un volume d'élution de 65 ml. Grâce à la calibration de la colonne, nous avons pu estimer qu'il s'agit d'une protéine ayant une masse moléculaire d'environ 80 kDa. Le gel SDS-PAGE (figure 50D) montre une seule bande

correspondant à GH* (40 kDa). On peut donc en conclure que c'est très probablement un dimère de GH* qui est élué avec un volume d'éluion de 65 ml. La protéine éluée avec un volume d'éluion de 77 ml correspond à une protéine d'environ 40 kDa et donc à GH* sous forme monomérique.

Afin de confirmer que la protéine éluée avec un volume d'éluion de 65 ml est bien un dimère de GH*, un gel SDS-PAGE en condition réductrice et non réductrice a été réalisé (figure 50E). En condition réductrice (en présence de β -mercaptoéthanol), une seule bande correspondant à GH* est présente. En condition non-réductrice (en l'absence de β -mercaptoéthanol), deux bandes sont présentes, une migre à 80 kDa et l'autre à 40 kDa. Ces résultats ont permis de mettre en évidence que la bande à 80 kDa correspond bien à un dimère de GH*. Il doit donc se former un pont disulfure intermoléculaire pour former le dimère.



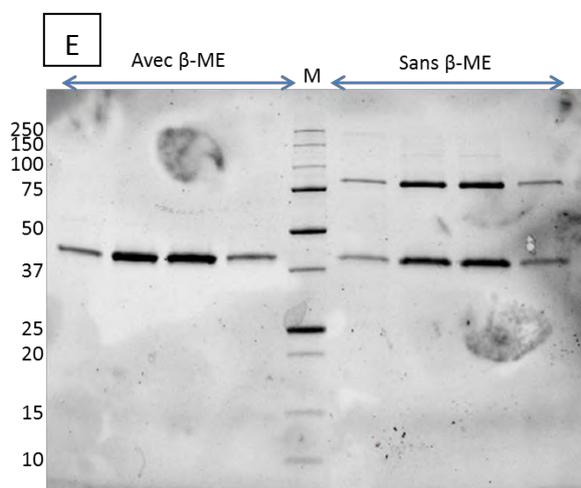
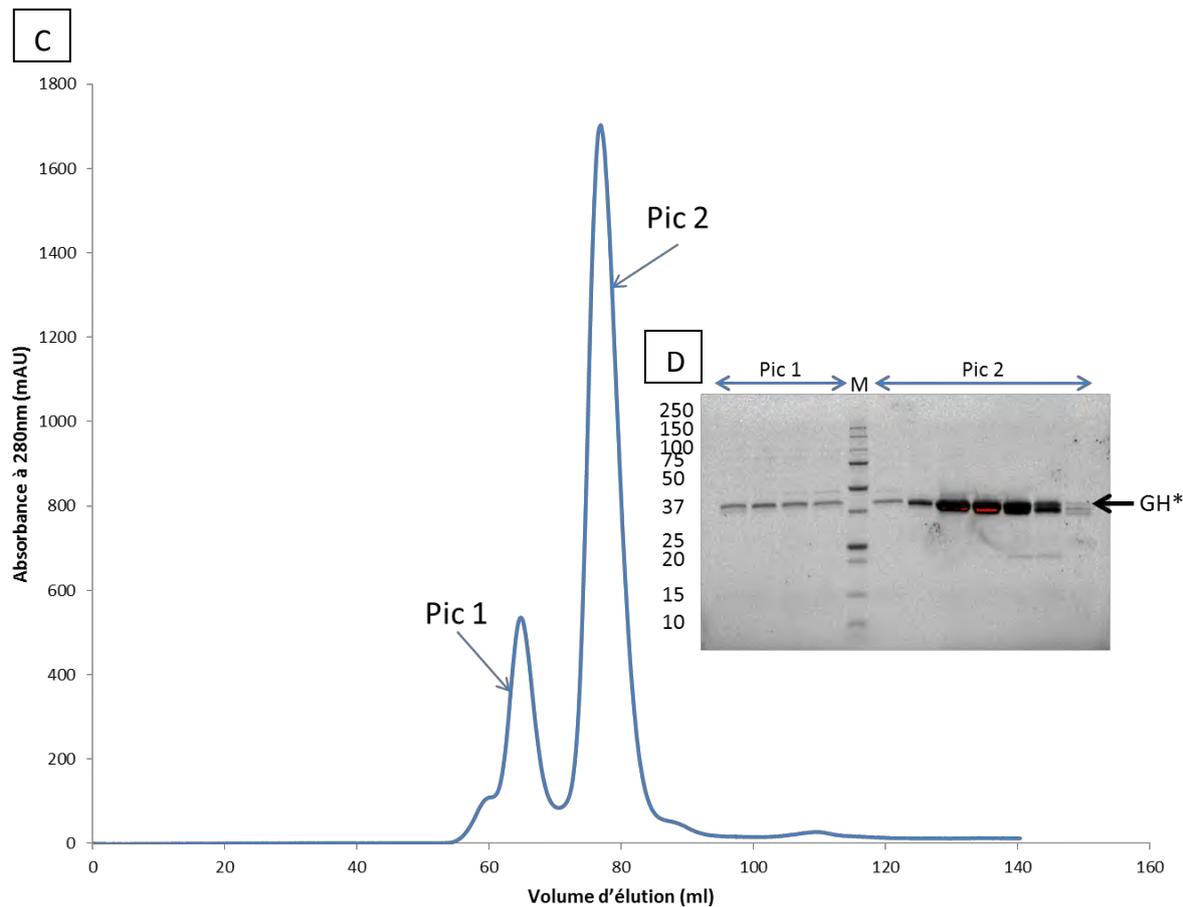


Figure 50 : Purification de GH* sur colonne de filtration sur gel. Chromatogramme de l'élution de GH* sur colonne de filtration sur gel HiLoad Superdex S75 16/60 (A) et gel SDS-PAGE pour vérifier la pureté après purification sur S75 (B). Les volumes d'élution des pics 1 et 2 sont respectivement de 50 ml et 58 ml. Chromatogramme de l'élution de GH* sur colonne de filtration sur gel HiLoad Superdex S200 16/60 (C) et gel SDS-PAGE pour vérifier la pureté après purification sur S200 (D). Les volumes d'élution des pics 1 et 2 sont respectivement 65 ml et 77 ml. Marqueurs de taille (M). Gel SDS-PAGE après S200 des fractions correspondant au pic 1 en présence et en absence de β-mercaptoéthanol (β-ME) (E).

Tous ces résultats montrent que GH* est pure, majoritairement sous forme monomérique, après une 2^{ème} étape de purification. La colonne permettant la meilleure séparation est la S200 préparative car la gamme de masse de séparation est plus adaptée.

En fin de purification il reste environ 9 mg de protéine soit un rendement de 90 % entre les 2 étapes de purification.

Des expériences de fluorimétrie différentielle à balayage (DSF) (présentées ci-après) ont montré que le meilleur tampon pour GH* est un tampon 20 mM citrate de sodium, 100 mM NaCl, pH 5,0. Pour les purifications sur S200 préparative, nous avons donc choisi d'effectuer l'élution à 1 ml/min et dans le tampon 20 mM citrate de sodium, 100 mM NaCl, pH 5,0.

IV. Stabilité de GH*

Des expériences de DSF ont été menées afin de déterminer quel tampon est le plus favorable pour la stabilité de GH* en mesurant sa température de fusion (T_m). Des tests préliminaires ont permis de déterminer les conditions optimales de concentration en protéine et en Sypro-Orange pour effectuer ces essais à 5 μ M de GH* en présence de Sypro-Orange à 5 X. Dans ces conditions (5 X Sypro, 5 μ M GH*, tampon 50 mM phosphate de sodium pH 7,0), le T_m de la protéine obtenu est de 45,90 °C. A partir de ces données, un crible développé dans l'équipe a été utilisé (cf Matériels et méthodes p. 84). Il permet de cribler une large gamme de tampon et de pH (de 4,5 à 9,0) à différentes concentrations en NaCl. Le crible a révélé que GH* est plus stable à des pH autour de 5,0 et 6,0 (T_m les plus hauts) avec un T_m autour de 58,5 °C alors que la protéine semble être déstabilisée dans des tampons à pH alcalin (T_m les plus bas autour de 35 °C) (figure 51).

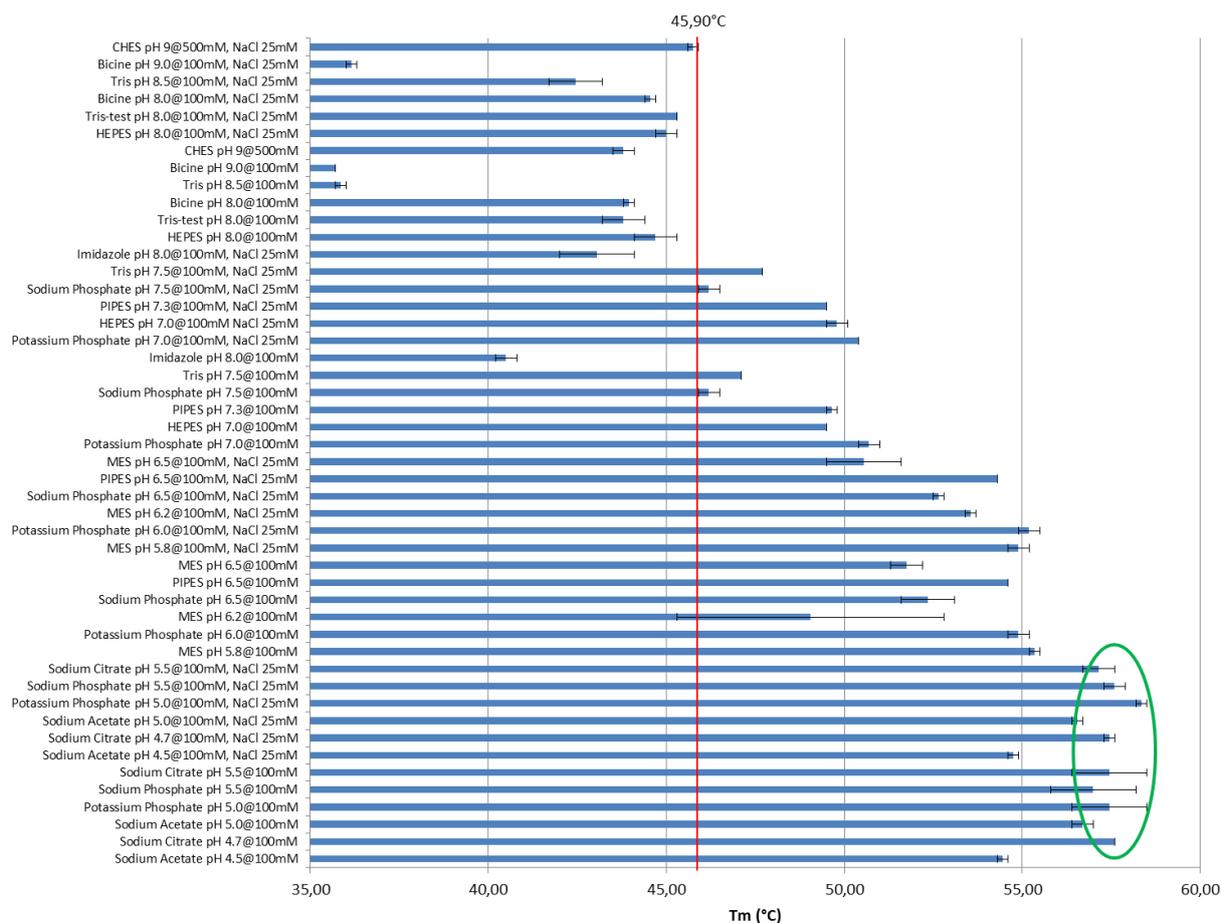


Figure 51 : Graphique présentant la température de fusion (T_m) de GH* en fonction du tampon utilisé. Le trait rouge correspond au T_m de GH* obtenu dans le premier test, la zone de plus haut T_m de l'enzyme est entourée en vert.

Le tampon choisi pour garder la protéine ainsi que pour lancer les essais de cristallisation est un tampon 20 mM citrate de sodium, 100 mM NaCl, pH 5,0 car ce tampon est un de ceux dans lesquels GH* a la meilleure stabilité et car pour la cristallisation nous avons préféré éviter la présence de phosphate dans le tampon. La salinité a été choisie à 100 mM car il faut du sel dans le tampon pour l'étape de chromatographie d'exclusion-diffusion et le T_m de GH* est diminué de 1 à 2 °C à des concentrations en sel plus importantes.

V. Détermination de l'activité de GH*

Afin de découvrir l'activité de GH*, l'activité sur X-cellobioside (qui correspond à l'activité du clone) a d'abord été évaluée mais n'a pas été validée. En conséquence, tous les substrats

commerciaux entrant dans la composition de la paroi végétale ainsi que tous les oligosaccharides modèles ont été criblés ainsi que des polysaccharides d'algues marines et des sucres phosphates.

A. Criblage des pNP-glycosides

Ces substrats imitent des di- ou trisaccharides. Comme le groupement paranitrophénate libéré par l'hydrolyse du substrat a une coloration jaune en milieu alcalin, les premiers cribles ont été effectués à 40 °C dans un tampon à 200 mM HEPES à pH 8,0 afin de rapidement cribler l'activité. Parmi les 19 pNP-glycosides testés, une activité a été détectée sur 8 d'entre eux : pNP- α -L-arabinofuranoside, pNP- β -D-galactopyranoside, pNP- α -L-arabinopyranoside, pNP- β -D-glucopyranoside, pNP- β -D-xylopyranoside, pNP- β -D-lactopyranoside, pNP-acide sialique et pNP-acide glucuronique (tableau 12). Ces activités témoigneraient d'une activité exoglycosidase de GH*. Il semblerait donc que la protéine possède une ou plusieurs de ces activités : α -L-arabinofuranosidase, β -D-galactosidase, α -L-arabinopyranosidase, β -D-glucosidase, β -D-xylosidase, lactopyranosidase, sialidase ou glucuronidase. Ces activités sont visibles après un temps relativement long (au moins 2 h, voire plus de 4 h pour certaines) ce qui implique une activité faible et/ou des conditions non optimales pour l'activité de GH*.

	Substrat	Activité
pNP-glycosides	pNP-β-D-Galp	+/-
	pNP-β-D-Galf	-
	pNP-α-D-Galp	-
	pNP-β-D-6-P-Galp	-
	pNP-N-ac-Galp	-
	pNP-α-L-Arap	+/-
	pNP-β-L-Arap	-
	pNP-α-L-Araf	+/-
	pNP-β-D-Glcp	+/-
	pNP-α-D-Glcp	-
	pNP-N-ac-Glcp	-
	pNP-α-L-Rhap	-
	pNP-β-D-Xylp	+/-
	pNP-β-D-Lacp	+/-
	pNP-α-L-Fucp	-
	pNP-cellobioside	-
	pNP-β-D-Manp	-
	pNP-sialic acid	+/-
pNP-glucuronide	+/-	
Oligosaccharides	Galactobiose	-
	Arabinobiose	-
	Mannobiose	-
	Galactosylmannobiose	-
	Xylobiose	-
	Arabinosylxylobiose	-
	Xylosylcellobiose	-
	Cellobiose	-
	Lactose	-
	Chitobiose diacétylé	-
	Chitopentaose pentaacétylé	-
Polysaccharides	Galactane pectique (pomme de terre)	-
	Galactane pectique (lupin)	-
	Rhamnogalacturonane I	-
	Rhamnogalacturonane II	-
	Arabinogalactane	-
	Galactane	-
	Arabinane	-
	Arabinane linéaire	-
	Acide polygalacturonique	-
	Galactomanane	-
	Xyloglucane	-
	Arabinoxylane de blé (WAX)	+
	WAX débranché enzymatiquement	+
	WAX débranché par traitement acide	-
	Arabinoxylane de seigle	+
	Xylane (hêtre)	+
	β-glucane	-
	Mannane	-
	1,4-β-mannane	-
	Chitine	-
	Amidon	-
Carboxyméthylcellulose	-	
Alginate	-	
Carraghénane	-	
Acide aldo-uronique	-	
Héparosane	-	

Tableau 12 : Résultats des cribles sur pNP-glycosides, oligosaccharides et polysaccharides.

Paranitrophényl (pNP), galactose (Gal), arabinose (Ara), glucose (Glu), rhamnose (Rha), xylose (Xyl), lactose (Lac), fucose (Fuc), mannose (Man), phosphate (P), acide (ac), pyranose (p) et furanose (f). Les + représentent des activités positives, les +/- une activité faible et les – l'absence d'activité.

Afin de confirmer ces activités, des cinétiques ont été réalisées sur 24 h avec des prélèvements à 0 min, 30 min, 1 h, 2 h, 4 h, 8 h et 24 h, à 40 °C, dans un tampon 20 mM citrate de sodium, 100 mM NaCl à pH 5,0. La plupart n'ont pas donné de résultats positifs. Pour quelques substrats, des activités ont pu être mesurées mais ces dernières sont très faibles. Par exemple, nous avons mesuré une activité en présence de pNP- α -L-arabinofuranoside d'environ 10^{-4} $\mu\text{mol}/\text{min}/\text{mg}$. Cette activité est quasi-nulle en comparaison des arabinofuranosidases de familles classiques comme les GH43 ou 51 dont les activités spécifiques sont de l'ordre de 10^2 $\mu\text{mol}/\text{min}/\text{mg}$ (Paës et al. 2008 ; Lee et Lee 2014).

B. Recherche d'activité de phosphorolyse sur oligosaccharides

Selon l'annotation des experts de CAZy, GH* semble ne posséder que le glutamate jouant le rôle de résidu catalytique acide/base (E117) et une activité de phosphorolyse a donc été envisagée. En effet les enzymes ayant ce type d'activité ne possèdent que l'acide/base catalytique et utilisent une molécule de phosphate inorganique pour remplacer le résidu nucléophile catalytique manquant (Cuskin et al. 2015 ; Ladevèze et al. 2013). Comme les enzymes faisant de la phosphorolyse ont une préférence pour les oligosaccharides, des tests sur di- et trisaccharides en présence de phosphate inorganique à 10 mM ont été faits sur 24 h et le résultat a été observé sur CCM mais aucune activité de phosphorolyse n'a été observée (tableau 12). Ces tests d'activités sur oligosaccharides avec révélation par CCM n'ont pas non plus permis de détecter une hydrolyse.

C. Criblage sur polysaccharides

De la même manière que pour les pNP-glycosides, des cribles sur polysaccharides ont été effectués. L'activité a été mesurée sur 24 h et révélée par dosage à l'acide dinitro-salicylique (DNS). Les substrats testés sont présentés dans le tableau 12.

Afin de déterminer sur quel substrat l'enzyme est la plus active, les hydrolyses des substrats permettant une détection de l'activité ont été comparées sur 23 h en faisant un prélèvement après 16 h et un autre après 23 h (figure 52). On voit sur le diagramme que l'activité spécifique la plus élevée est mesurée en présence d'arabinoxylane de seigle (RAX, barres violettes) puis d'arabinoxylane de blé (WAX, barres rouges) et de xylane de hêtre (barres bleues).

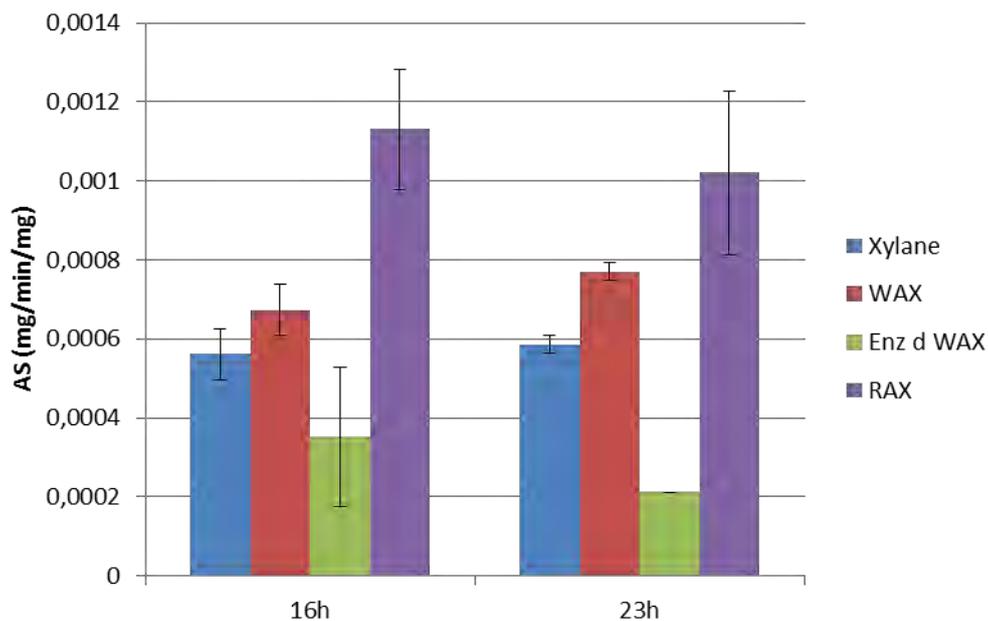


Figure 52 : Activité spécifique de GH* en mg/min/mg d'enzyme après 16 h et 23 h d'incubation. Arabinoxylane de blé (WAX), WAX débranché par traitement enzymatique (Enz d WAX) et arabinoxylane de seigle (RAX). Le xylane est un xylane de hêtre. Les expériences ont été faites en triplicatas (n=3).

Ce résultat permet de penser que GH* a une activité xylanase avec une capacité à accommoder les substitutions. En effet, les arabinoxylanes de blé et de seigle sont constitués d'une chaîne principale de D-xyloses liés en β -1,4, ramifiée par des résidus L-arabinofuranosides liés en α -1,2 ou en α -1,3. La répartition des ramifications n'est pas la même entre les deux polysaccharides. Le xylane de hêtre possède lui aussi une chaîne principale de D-xyloses liés en β -1,4, par contre les ramifications sont majoritairement des résidus d'acide glucuronique liés en α -1,2 et minoritairement des résidus L-arabinofuranosides liés en α -1,2 ou en α -1,3.

Les activités sur RAX et WAX semblent plus fortes que celle sur WAX débranché enzymatiquement, l'hypothèse que GH* puisse accommoder les résidus α -L-arabinofuranoside semble confirmée. En effet le RAX et le WAX possèdent 62 % de β -D-xylose (chaîne principale) et 38 % d' α -L-arabinose (ramifications) alors que le WAX débranché enzymatiquement possède 70 % de β -D-xylose et 30 % d' α -L-arabinose. Ce substrat a été traité par l' α -L-arabinofuranosidase de *Bacteroides ovatus* qui hydrolyse spécifiquement les liaisons α -1,3 sur les xyloses di-substitués. L'activité plus faible obtenue sur ce substrat amène deux autres hypothèses : la 1^{ère} est que la faible activité vient du fait qu'il y a moins de ramifications et la 2^{ème} serait que GH* a une préférence (non exclusive) pour les liaisons α -1,3 et donc quand il y a moins de ramifications en α -1,3, l'enzyme est moins efficace.

Le WAX débranché par traitement acide a été testé. Il ne possède quant à lui que 22 % d' α -L-arabinose (contre 78 % de β -D-xylose). Le fait qu'il y ait peu de ramifications doit expliquer l'absence d'activité de GH* sur ce substrat.

La détection d'une activité sur tous ces substrats va dans le sens d'une activité endo-xylanase ou éventuellement β -xylosidase. La différence d'activité entre ces substrats peut s'expliquer par la différence de distribution et du nombre des ramifications tout au long du polysaccharide.

Une cinétique sur 24 h a été réalisée sur le substrat qui semble être le meilleur, le RAX, afin de déterminer l'activité spécifique de GH* (figure 53).

Cette activité est de $0,023 \pm 0,002$ $\mu\text{mol}/\text{min}/\text{mg}$. Ce qui est très faible au regard des xylanases connues dans la littérature. Par exemple la xylanase (GH11) de *T. xylanilyticus* a une activité spécifique de 1750 ± 90 U/mg (Paës et al. 2007) ou encore la xylanase Xyn10N18 venant d'une banque métagénomique de rumen présente une activité spécifique de 822,6 U/mg (Gong et al. 2013).

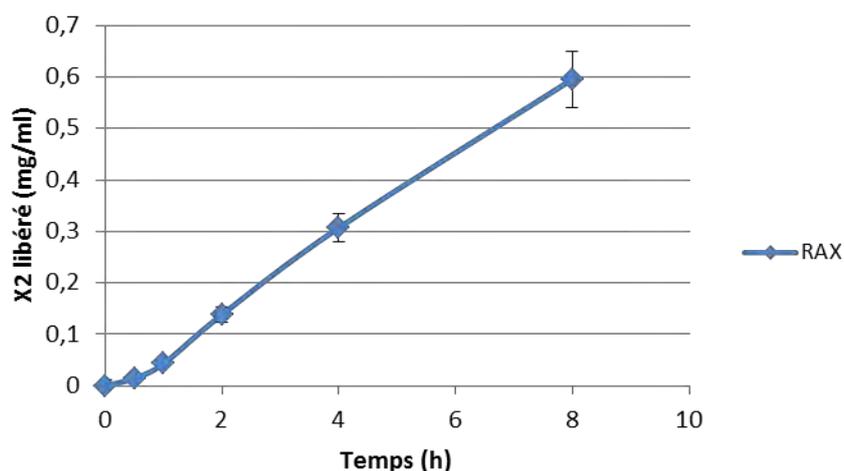


Figure 53 : Courbe présentant la quantité de xylobiose (X2) libéré en fonction du temps. L'hydrolyse du RAX a été faite sur 8 h en présence de GH*. La régression linéaire donne la fonction suivante : $y = 0,077x - 0,0156$ avec un $R^2 = 0,9967$.

En conséquence, comme l'activité de GH* est très faible et sachant que certaines GH ont besoin d'un cofacteur, des tests sur polysaccharides en présence soit de NADH, soit de calcium, soit de phosphate inorganique ont été effectués. Les activités obtenues sur polysaccharides restent inchangées en présence ou en absence de ces cofacteurs.

VI. Conditions retenues pour l'étude fonctionnelle de GH*

A. Détermination du pH optimal

Ces mesures ont été effectuées par Franck Moncassin pendant son stage de M2R sur le meilleur substrat de GH*, le RAX. Les résultats obtenus sont présentés en figure 54. La gamme de pH va de 2,6 à 8,0 et les mesures ont été réalisées à 40 °C.

On peut voir sur la figure 54 que GH* présente un maximum d'activité à pH 5,5 dans un tampon citrate-phosphate. D'autre part, nous avons constaté que GH* présente une activité supérieure à 90 % entre des valeurs de pH allant de 4,5 à 6,0. Or, par DSF nous avons déterminé que GH* est plus stable à des valeurs de pH situées entre 4,5 et 5,5. Il apparaît donc que l'enzyme présente la meilleure activité quand elle est dans un tampon où elle est la plus stable. Sur la figure 54, on voit aussi que l'activité diminue pour des valeurs de pH supérieures 6,0 ce qui corrèle aussi avec les résultats de DSF qui avaient mis en avant que

GH* est moins stable à des pH alcalins (entre pH 8,0 et 9,0). On note aussi que GH* perd quasiment 100 % de son activité à partir de valeurs de pH inférieures à 4,0.

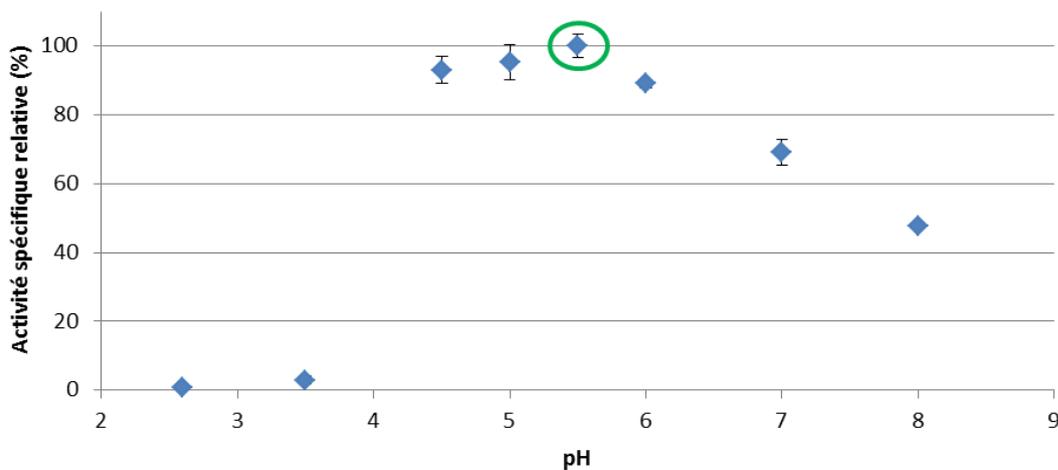


Figure 54 : Courbe présentant l'activité spécifique relative de GH* en fonction du pH. La valeur de pH optimal est entourée en vert.

B. Détermination de la température optimale

Ces tests ont aussi été effectués par Franck Moncassin pendant son stage de M2R sur le meilleur substrat de GH*, le RAX. Les résultats obtenus sont présentés en figure 55 pour une gamme de température allant de 30 à 60 °C dans un tampon citrate-phosphate (43 mM citrate, 114 mM phosphate) à pH 5,5.

Les résultats indiquent que GH* présente un maximum d'activité entre 45 et 50 °C. Pour des températures inférieures ou égales à 40 °C, l'enzyme perd plus de 60 % de son activité. A 55 °C GH* perd 40 % de son activité et l'activité de l'enzyme est totalement perdue dès 60 °C. Le T_m de GH* déterminé par DSF est d'environ 58 °C à pH 5,5, les résultats de thermoactivité sont donc en adéquation avec les expériences de DSF qui indiquent qu'au-delà de 58 °C, la protéine est dénaturée à plus de 50 %.

La température optimale choisie pour l'activité de GH* est 45°C pour être la plus éloignée du T_m tout en conservant une activité maximale.

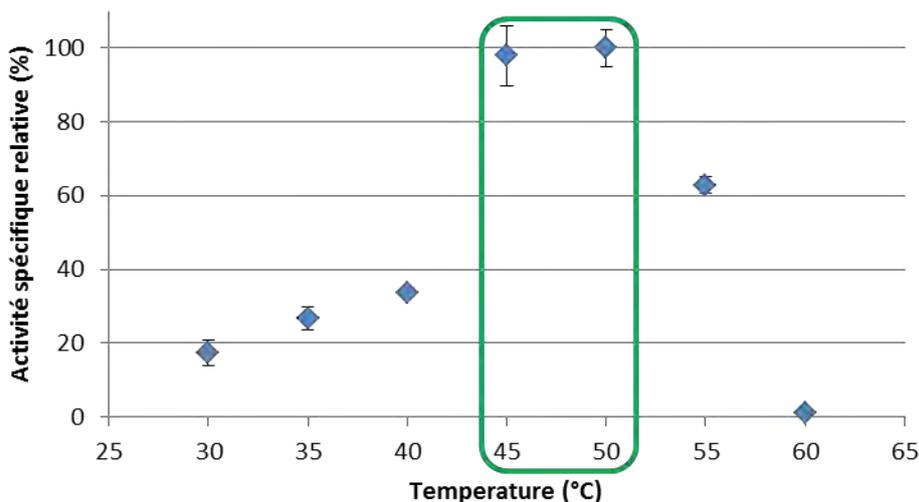


Figure 55 : Courbe présentant l'activité spécifique relative de GH* en fonction de la température. Les températures, pour lesquelles l'activité de GH* est la plus forte, sont encadrées en vert.

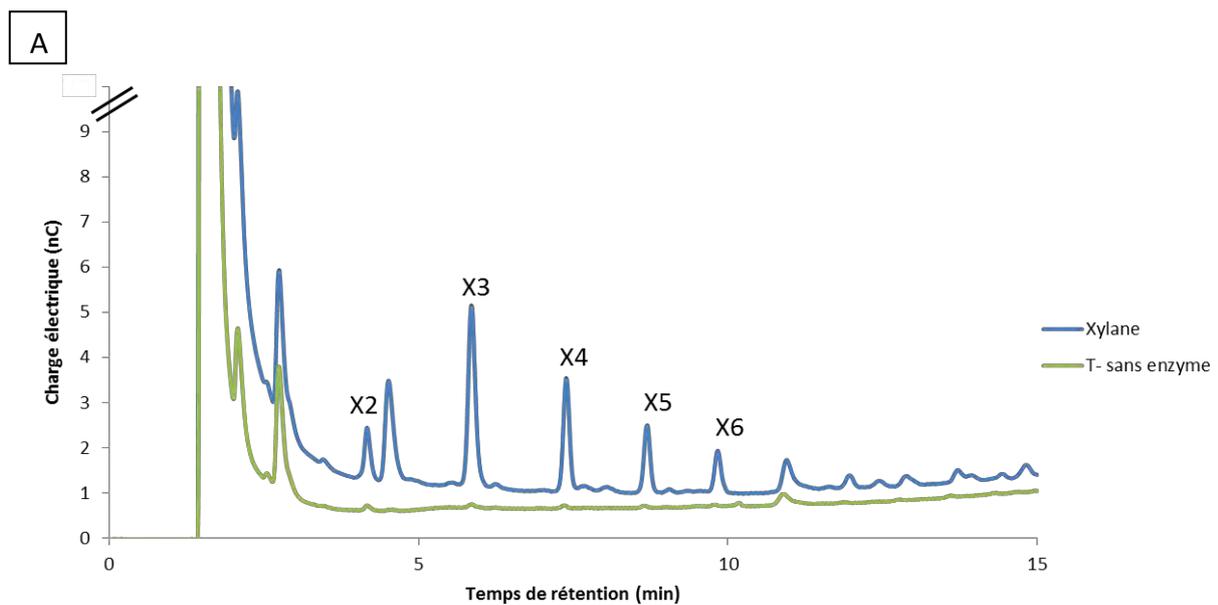
Lors du traitement des données, nous avons pu constater que GH* conserve 98 % de son activité sur 3 h à 45 °C. Les réactions d'hydrolyse ont donc été réalisées par la suite dans un tampon phosphate-citrate (43 mM citrate, 114 mM phosphate) à pH 5,5 et à 45 °C.

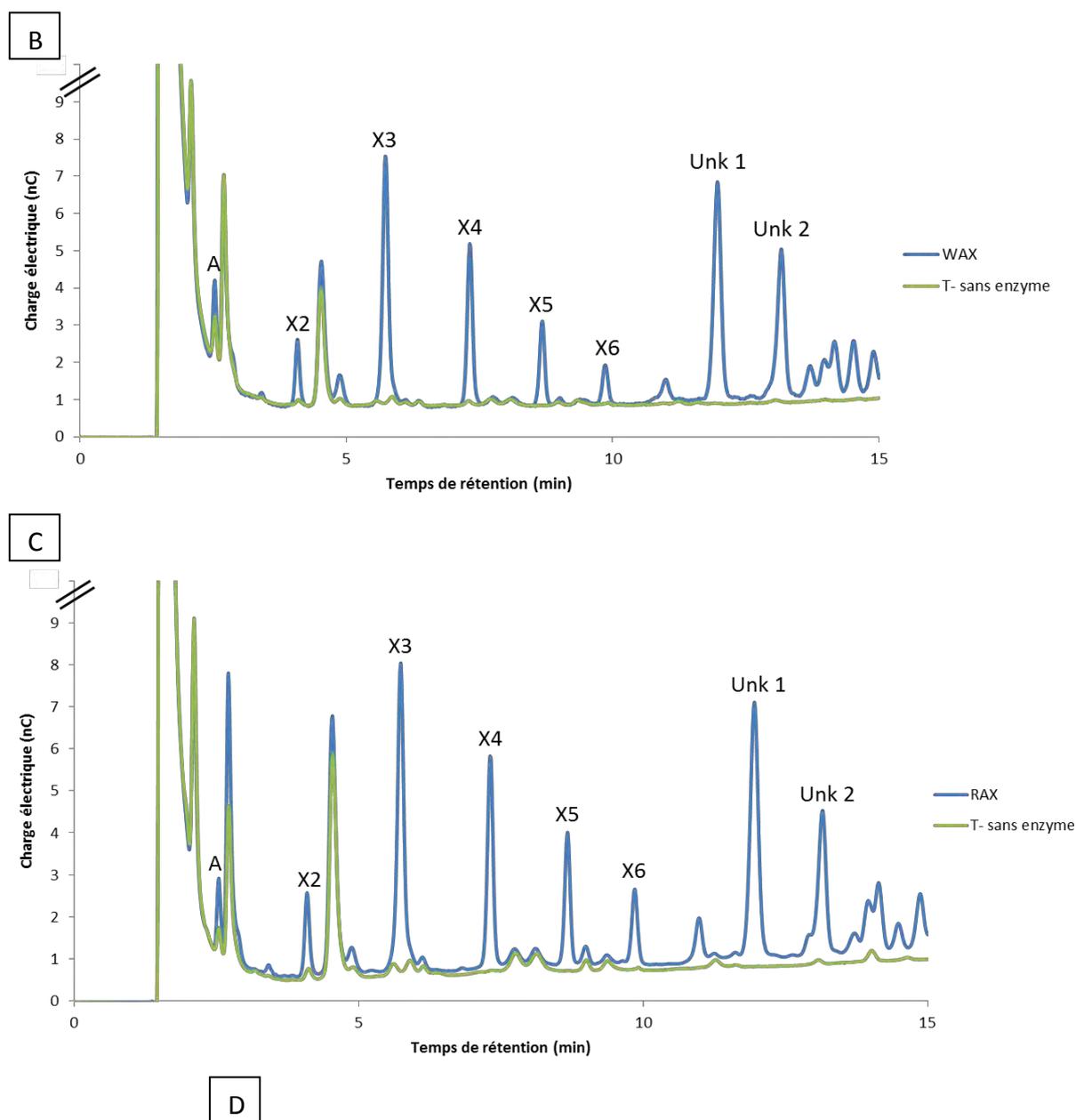
VII. Analyse des produits d'hydrolyse du xylane de hêtre et des arabinoxylyanes de blé (WAX) et de seigle (RAX)

Pour déterminer ce qui est libéré par GH* lors de l'hydrolyse du xylane de hêtre, du WAX et du RAX, des analyses par HPAEC-PAD ont été menées. Les chromatogrammes montrent qu'il n'y a pas d' α -L-arabinofuranose produit à partir du xylane (figure 56A) et seulement de très faibles quantités sont produites par GH* lors de l'hydrolyse du WAX et du RAX (figure 56A, B et C). Ces résultats semblent logiques pour le xylane car celui-ci ne contient quasiment pas d' α -L-arabinofuranose. Par contre, les résultats sur les arabinoxylyanes indiquent que GH* posséderait une activité α -L-arabinofuranosidase négligeable étant donné leurs fortes proportions en résidus α -L-arabinofuranoside.

De plus, il est intéressant de voir que des xylooligosaccharides allant du xylobiose (X2) au xylohexaose (X6) sont plus largement libérés par GH* pour chacun des polysaccharides. Le xylotriose (X3) est l'oligosaccharide produit majoritairement par GH* devant le xylo tétraose

(X4) et le xylopentaose (X5). Les quantités de X6 libérées sont un peu plus faibles que celles de X5. Le X2 est l'oligosaccharide le moins produit par GH* lors de l'hydrolyse du xylane, du WAX et du RAX (figure 56B et C). Ces résultats nous permettent de confirmer que GH* possède une activité endo-xylanase. Il semblerait que GH* ne libère pas de xylose (X1) ce qui indique que GH* n'aurait pas d'activité β -xylosidase. De plus on peut voir que pour le RAX et le WAX, 2 produits d'hydrolyse supplémentaires sont libérés et sortent avec des temps de rétention de 11,9 et 13,2 min (pics Unk 1 et 2 respectivement). Ces produits n'ont pas pu être caractérisés car ils ne correspondent à aucun des témoins injectés. Etant donné les temps de rétention de ces 2 composés et la nature des polysaccharides de départ, l'hypothèse la plus vraisemblable est que ces 2 produits correspondent à des xylooligosaccharides de plus haut degré de polymérisation (DP) ou à des xylooligosaccharides de DP 5 ou 6 ramifiés.





Oligosaccharides libérés	WAX	RAX	Xylane
Xylobiose (X2)	0,33 ± 0,02 (μM)	0,36 ± 0,05 (μM)	0,67 ± 0,40 (μM)
Xylotriose (X3)	1,74 ± 0,07 (μM)	1,86 ± 0,18 (μM)	3,49 ± 2,36 (μM)
Xylotétraose (X4)	0,79 ± 0,07 (μM)	0,86 ± 0,11 (μM)	1,83 ± 1,36 (μM)
Xylopentaose (X5)	0,38 ± 0,03 (μM)	0,49 ± 0,06 (μM)	1,28 ± 1,00 (μM)
Xylohexaose (X6)	0,19 ± 0,02 (μM)	0,29 ± 0,05 (μM)	0,92 ± 0,73 (μM)

Figure 56 : Résultats de l'analyse des produits d'hydrolyse par HPAEC-PAD. Chromatogrammes de l'hydrolyse du xylane de hêtre (A), du WAX (B) et du RAX (C). La courbe bleue correspond à la réaction d'hydrolyse du substrat par GH* et la courbe verte correspond au témoin négatif (polysaccharide sans enzyme). Xylobiose (X2), xylotriose (X3), xylotétraose (X4), xylopentaose (X5),

xylohexaose (X6), arabinose (Ara), pic non attribué (Unk). **D. Tableau présentant les quantités de xylooligosaccharides libérés lors de l'hydrolyse du RAX et du WAX par GH*.**

Finalement l'analyse des produits d'hydrolyse du xylane de hêtre, du RAX et du WAX par HPAEC-PAD a permis d'affirmer que des activités exoglycosidases telles que α -L-arabinofuranosidase ou β -xylosidase sont négligeables pour GH*. Ces résultats nous ont permis de confirmer que GH* possède une activité endo- β -1,4-xylanase.

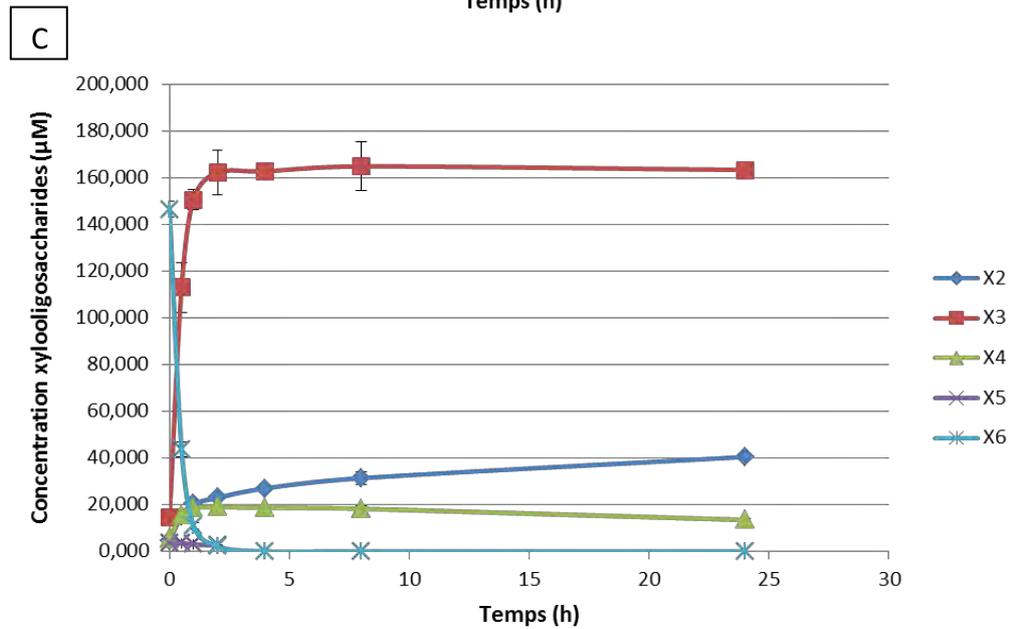
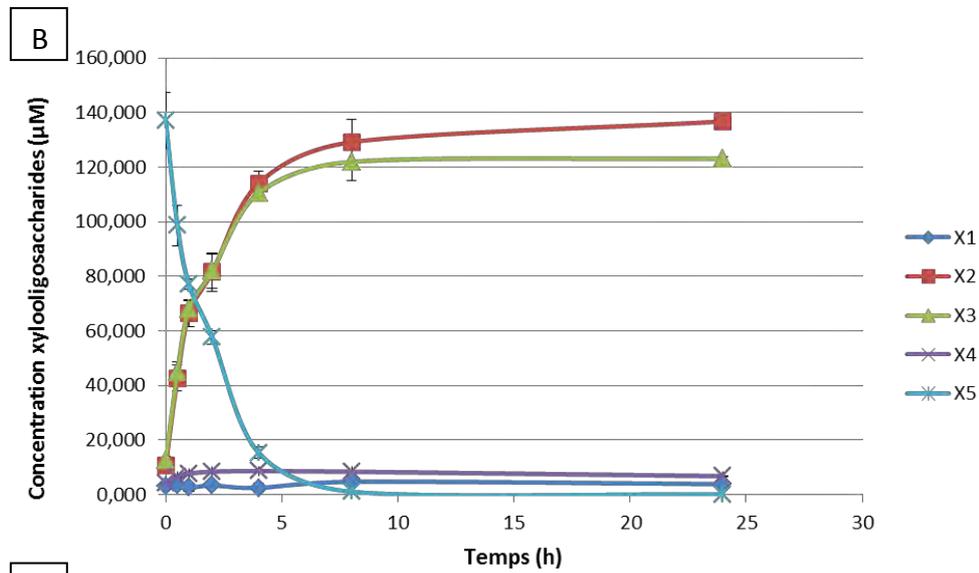
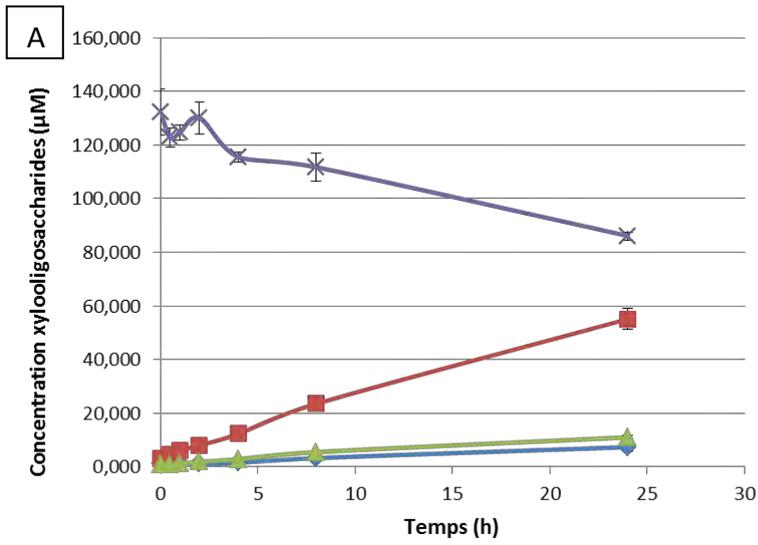
VIII. Analyse des produits d'hydrolyse de xylooligosaccharides par GH*

Afin de confirmer l'activité xylanase de GH*, des réactions d'hydrolyse par GH* de xylooligosaccharides allant de X3 à X6 ont été réalisées. Les produits d'hydrolyse ont ensuite été analysés par HPAEC-PAD.

Alors qu'aucune hydrolyse n'a pu être détectée à partir de X3, le X4 est hydrolysé par GH* mais la réaction est lente. En effet, seulement 1/3 du X4 est hydrolysé en 24 h. Par contre il est intéressant de noter qu'une molécule de X4 donne très majoritairement du X2 et quasiment pas de X1 + X3 (figure 57A).

En revanche, la totalité du X5 est hydrolysé par GH* en 8 h et donne du X3 + X2 comme produits majoritaires et pas d'apparition de X4 ou de X1 (figure 57B).

La réaction la plus rapide est celle sur X6 puisque la totalité du substrat est hydrolysée en 2 h, générant majoritairement 2 molécules de X3. La libération de X4 et de X2 est observée de façon très minoritaire et il n'y a pas de libération de X5 + X1. De plus, il est intéressant de voir que la courbe du X2 continue d'augmenter alors que celle correspondant au X4 diminue après 8 h d'hydrolyse. Il semblerait donc que lorsqu'il n'y a plus de X6 dans le milieu, GH* hydrolyse le X4 formé en X2 (figure 57C).



D	
	AS ($\mu\text{mol}/\text{min}/\text{mg}$)
X3	$0,00 \pm 0,00$
X4	$0,15 \pm 0,01$
X5	$2,30 \pm 0,20$
X6	$11,28 \pm 0,29$

Figure 57 : Chromatogrammes présentant l'hydrolyse du X4 (A), du X5 (B) et du X6 (C) par GH* et tableau présentant l'activité spécifique de l'enzyme sur X3, X4, X5 et X6 (D). Les activités ont été déterminées à partir des pentes des courbes d'hydrolyse (avant d'avoir atteint un plateau) de X6, X5, X4 et X3, obtenues en HPAEC-PAD.

L'activité endo- β -1,4-xylanase de GH* a été confirmée par les expériences d'HPAEC-PAD sur xylooligosaccharides. De plus, il semble que l'enzyme soit plus active sur xylooligosaccharides que sur RAX ($0,023 \pm 0,002 \mu\text{mol}/\text{min}/\text{mg}$) et que celle-ci préfère accommoder des xylooligosaccharides de DP supérieur à 4 car son activité spécifique augmente avec la taille des oligosaccharides (figure 57D). Des xylanases de la famille GH10 actives sur polysaccharide et sur oligosaccharide ont déjà été caractérisées mais la plupart de ces enzymes, contrairement à GH*, sont capables d'hydrolyser des xylooligosaccharides de petite taille tels que le X3 et libèrent des X1 et X2 (Alvarez, Goldbeck, dos Santos, et al. 2013 ; Gong et al. 2013 ; Ko et al. 2016).

IX. Conclusion

Grâce à l'optimisation de l'expression de GH*, des quantités suffisantes d'enzymes ont pu être produites pour permettre la caractérisation fonctionnelle et structurale de l'enzyme. L'optimisation des étapes de purification a permis d'obtenir une enzyme pure.

La recherche d'activité de l'enzyme a permis d'identifier le RAX comme étant le meilleur substrat (parmi ceux testés) pour la caractérisation de l'enzyme. Avec ce substrat les paramètres optimaux de l'enzyme ont été déterminés. Le pH optimal de fonctionnement de GH* est de 5,5 et la température optimale est de 45 à 50 °C. Les conditions d'utilisation de l'enzyme ont donc été fixées en tampon phosphate-citrate (43 mM citrate, 114 mM phosphate) à pH 5,5 à 45 °C.

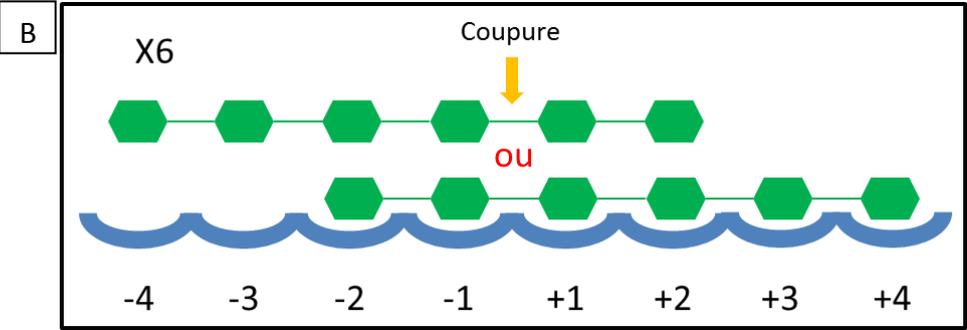
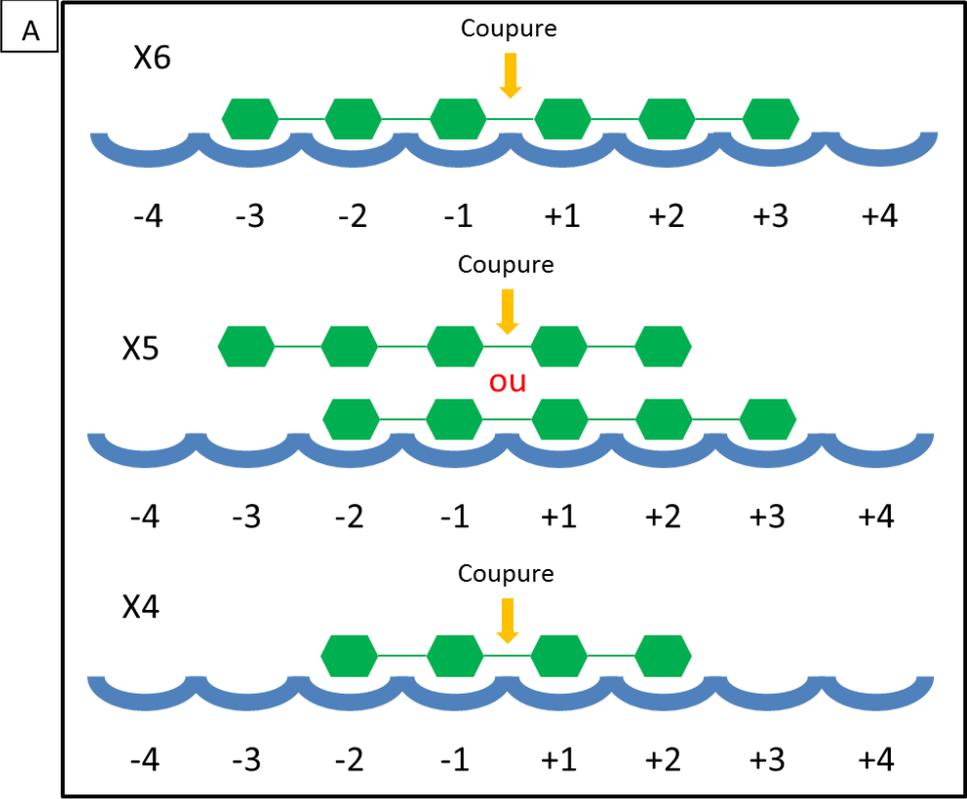
L'analyse des produits d'hydrolyse du RAX par GH* nous a permis d'identifier l'activité de GH*. Il s'agit d'une endo- β -1,4-xylanase. Et l'analyse de l'hydrolyse de xylooligosaccharides par GH* nous a permis d'émettre l'hypothèse que GH* aurait une préférence pour les xylooligosaccharides de DP supérieurs à 4.

Cependant l'activité de GH* reste 10 à 2000 fois plus faible que l'activité de nombreuses xylanase déjà caractérisées (Paës et al. 2007 ; Paës, Berrin, et Beaugrand 2012 ; Gong et al. 2013 ; Alvarez, Goldbeck, dos Santos, et al. 2013 ; Song et al. 2014 ; Ko et al. 2016).

L'analyse des produits d'hydrolyse des xylooligosaccharides permet de faire une hypothèse quant à l'organisation du site catalytique de l'enzyme et la nécessité d'avoir certains sous-sites occupés pour l'activité de GH* (figure 58).

En effet, l'enzyme produit majoritairement 2 molécules de X3 à partir d'une molécule de X6 et peu de X4 + X2. D'autre part, il y a exclusivement du X3 + X2 produit à partir de X5. Enfin à partir du X4, on obtient majoritairement du X2. Le X3 quant à lui n'est pas hydrolysé par GH*. Tout cela implique que les sous-sites -2 et +2 doivent être occupés car il n'y pas de libération de X1 quel que soit le substrat hydrolysé.

L'occupation du site -3 et du site +3 doit aussi jouer un rôle important car on voit que le X4 n'est pas le substrat préférentiel puisqu'après 24 h d'hydrolyse seulement le tiers du X4 de départ a été hydrolysé.



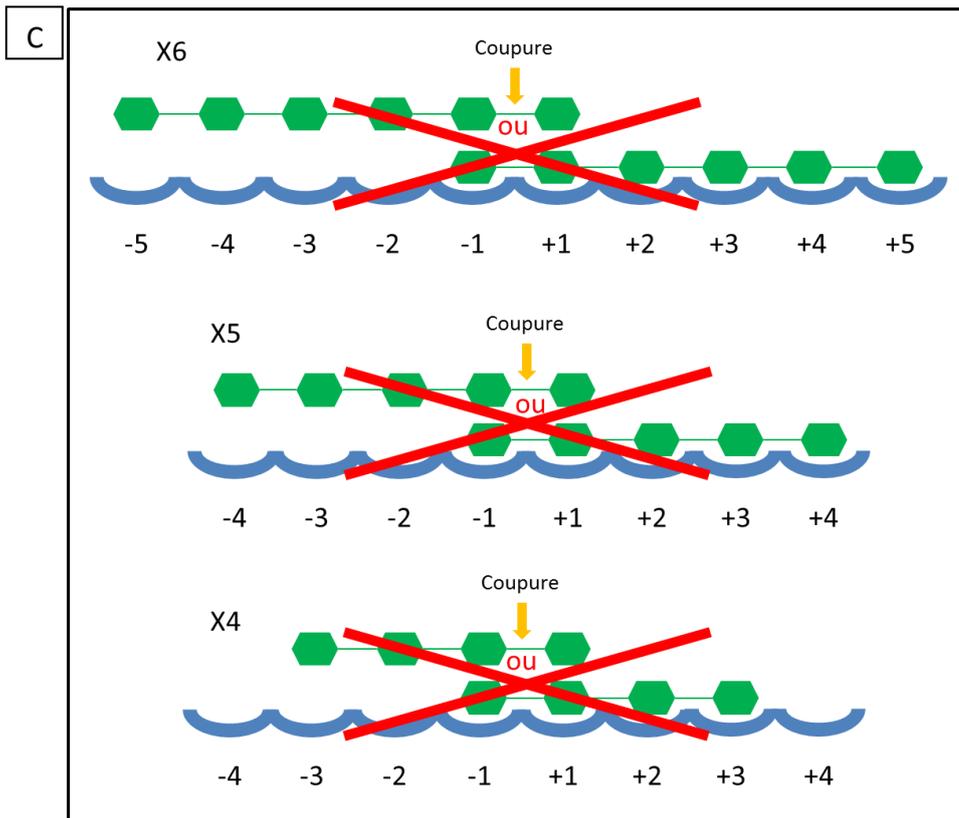


Figure 58 : Schéma des sous-sites de fixation des ligands hypothétiques de GH*. Positions favorisées (A), positions possibles (B) et positions défavorables (C) des xylooligosaccharides dans le site de fixation de GH*. Les sous-sites sont représentés en bleu, les xyloses en vert et le site où se fait l'hydrolyse est pointé par la flèche jaune. L'extrémité réductrice des oligosaccharides se situe du côté des sous-sites positifs et l'extrémité non-réductrice du côté des sous-sites négatifs.

Analyser les 2 pics non attribués (Unk 1 et 2) issus de l'hydrolyse du RAX par GH* afin de les identifier pourrait nous permettre d'avoir plus d'informations sur le site catalytique de l'enzyme. En effet, si ces produits sont linéaires, cela peut impliquer la présence d'autres sous-sites dans la continuité de ceux déjà identifiés. Si ces produits sont ramifiés, cela va impliquer la présence de sous-sites différents. La nature et la localisation des ramifications peut avoir un effet sur l'activité de GH*.

Tester des xylooligosaccharides plus longs tels que les X7 et X8 et des xylooligosaccharides longs et ramifiés pourrait permettre de confirmer ces hypothèses et aussi de mieux comprendre le mécanisme de l'enzyme.

Enfin obtenir la structure cristallographique d'un mutant inactif de GH* en présence d'un substrat tel que le X6 ou le X5 nous permettrait aussi de mieux comprendre le mode d'action de l'enzyme.

Finalement, quand on revient au fosmide sur lequel GH* a été identifiée, on constate que les autres GH présentes pourraient participer à la dégradation de substrats avec GH*. En effet, les autres protéines sont une GH1 (β -glucosidases majoritairement) et une GH4 (6-phospho- β -glucosidases majoritairement). Ces familles d'enzymes regroupent d'autres activités enzymatiques telles que β -xylosidase pour la famille 1 et α -glucuronidase pour la famille 4. Une β -xylosidase pourrait prendre en charge les X2 et X3 produits par GH* pour finir leur dégradation. Une α -glucuronidase pourrait hydrolyser les ramifications constituées d'acide glucuronique présentes sur les xylanes et ainsi permettre à GH* d'hydrolyser la chaîne principale de ces polysaccharides.

Toutefois, le fosmide a été identifié par sa capacité à hydrolyser le X-cellobioside, cette activité d'hydrolyse ne provenant pas de GH*, elle pourrait provenir de la GH1 (β -glucosidase). De nombreux transporteurs à hexoses-phosphate ont également été annotés et peuvent être mis en relation avec l'activité 6-phospho- β -glucosidase de la GH4. Ces différentes activités sont difficiles à mettre en relation avec celle de GH* à moins que la GH1 ne possède une double activité β -glucosidase et β -xylosidase, ce qui a déjà été observé et qui est rendu possible par la proximité structurale qui existe entre les xylooligosaccharides et les glucooligosaccharides (Mattéotti et al. 2011).

Chapitre III : Etude structurale de GH*

I. Introduction

La cristallographie est, depuis de nombreuses années, une des techniques les plus précises pour résoudre la structure tridimensionnelle de protéines. De nombreuses études structure-fonction ont permis de comprendre comment l'organisation tridimensionnelle d'une protéine peut être impliquée dans l'activité de celle-ci. Les études structurales de protéines ont notamment permis d'identifier les acides aminés impliqués dans la reconnaissance et/ou le maintien du substrat ainsi que ceux responsables de l'activité de la protéine (Cartmell et al. 2011b ; Zheng et al. 2012). Des études structurales ont aussi permis d'identifier les différentes étapes de mécanismes réactionnels (Mark et al. 2001 ; Amaya et al. 2004).

La connaissance de la structure tridimensionnelle d'une protéine pour laquelle on n'a que très peu d'informations peut permettre d'émettre un certain nombre d'hypothèses. Si la protéine a un repliement classique de familles d'enzymes, cela peut permettre d'identifier les résidus catalytiques hypothétiques ou encore orienter vers un type de substrat. Ces hypothèses peuvent être énoncées après analyse comparative entre la protéine d'intérêt et des protéines présentant une certaine homologie structurale ou en analysant la protéine elle-même (répartition des charges, présence de sillons ou de cavités, localisation de boucles potentiellement flexibles...).

II. Rappels sur GH*

A. Connaissances générales

Après annotation fonctionnelle par CAZy et selon Bernard Henrissat, GH* est une glycoside hydrolase appartenant au clan GH-A mais à aucune des 135 familles présentes à ce jour dans la classification CAZy. Toutes les enzymes du clan GH-A partagent le même repliement en tonneau $(\beta/\alpha)_8$ et ont pour résidus catalytiques deux acides glutamiques, l'un

jouant le rôle d'acide/base et l'autre de nucléophile. Toutes ces enzymes possèdent aussi un mécanisme d'action par rétention de la configuration absolue du carbone anomérique (cf. introduction bibliographique) (Vuong et Wilson 2010 ; Dilokpimol et al. 2011b).

La famille d'enzymes dont GH* est la plus proche est celle des GH5 (appartenant elle aussi au clan des GH-A) mais GH* ne peut pas être assignée à cette famille car il lui manque des signatures caractéristiques de cette famille, et notamment celle contenant le résidu catalytique jouant le rôle de nucléophile. Par contre l'acide glutamique jouant le rôle d'acide/base semble être présent dans la séquence (cf chapitre II).

B. Activité

Nous avons pu voir dans le chapitre II que GH* est une xylanase active sur arabinoxylane. En effet GH* libère des xylooligosaccharides allant du xylobiose au xylohexaose avec majoritairement du xylotriose à partir d'arabinoxylane de blé et de seigle. Il a aussi été démontré que l'enzyme est capable d'accommoder et d'hydrolyser des xylooligosaccharides avec une préférence pour les oligosaccharides de DP supérieur à 4 puisque l'activité est plus forte sur xylohexaose que sur xylopentaose et que sur xylohexaose. L'activité la plus élevée déterminée à ce jour est donc sur xylohexaose mais reste faible par rapport à des xylanases classiques des familles GH10 ou 11. Comme l'activité est meilleure sur des xylooligosaccharides de DP supérieur à 4 et que celle-ci augmente avec la longueur de la chaîne, il est possible que l'activité de GH* soit encore plus grande sur des xylooligosaccharides plus long tels que des X7, X8, voire plus longs encore. L'activité n'a pas pu être testée sur des xylooligosaccharides linéaires de DP supérieur à 6, ni sur des xylooligosaccharides ayant un DP de 5 minimum et ramifiés car nous n'avons pas accès à de tels substrats non commerciaux.

III. Conditions de cristallisation

Une fois produite et purifiée en deux étapes de chromatographie successives (affinité et filtration sur gel) (cf. chapitre II), GH* est concentrée par centrifugation sur concentrateur ayant un seuil de coupure de 10 kDa. La protéine est conditionnée dans un tampon 20 mM citrate de sodium, 100 mM NaCl, pH 5,0. Ce tampon a été identifié par DSF comme étant le

tampon dans lequel la protéine a le T_m le plus élevé (58,5 °C). Or, plus le T_m est élevé et plus la probabilité d'obtenir des cristaux de protéine est élevée (Ericsson et al. 2006 ; Reinhard et al. 2013). Pour les premiers essais de cristallisation, GH* a été concentrée à 12,5 mg/ml afin de pouvoir mener suffisamment d'essais de cristallisation. Ainsi 4 cribles commerciaux (JCSG I à IV, Qiagen) ont été utilisés pour cribler une large gamme de conditions. Les premiers essais ont donné des petits cristaux dans plusieurs conditions. Les plus intéressants ont poussé dans la condition contenant une solution de cristallisation 0,2 M d'acétate de calcium et 20 % p/v de PEG 3350 (figure 59A) (tableau 13).

Comme de nombreuses gouttes contenaient des précipités, ce qui peut indiquer une trop grande concentration en protéine, de nouveaux essais donc ont été menés en diminuant la concentration en protéine de 12,5 à 8 mg/ml, en utilisant les mêmes cribles commerciaux. De nouvelles conditions de cristallisation notamment avec 0,2M d'acétate de magnésium, 20 % p/v PEG 3350, sont apparues (figure 59C) (tableau 13) et des cristaux moins nombreux mais plus gros ont été obtenus dans les conditions déjà identifiées lors des premiers essais (0,2 M acétate de calcium, 20 % p/v PEG 3350) (figure 59B) (tableau 13). Diminuer la concentration en protéine a donc permis d'obtenir des cristaux de même forme mais dont les dimensions sont plus grandes.

Par la suite, de nombreux essais manuels ont été menés, en faisant varier la concentration en sel et en PEG 3350 (tableau 14), afin d'optimiser les conditions de cristallisation mais n'ont pas permis d'accroître la taille des cristaux.

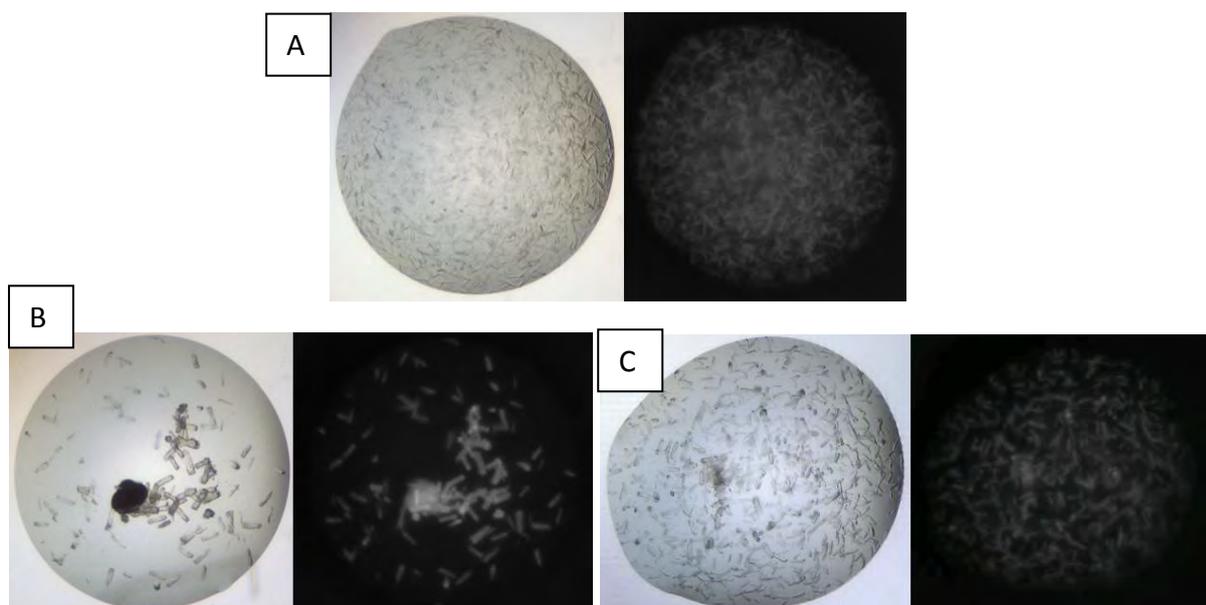


Figure 59 : Microphotographie des gouttes de cristallisation observées en lumière visible (image de gauche) et en UV (image de droite) pour différentes conditions. A. Cristaux de GH* obtenus à 12,5 mg/ml en présence de tampon 0,2 M acétate de calcium, 20% p/v PEG 3350. **B.** Cristaux de GH* obtenus à 8 mg/ml en 0,2 M acétate de calcium, 20 % p/v PEG 3350. **C.** Cristaux de GH* obtenus à 8mg/ml en 0,2 M acétate de magnésium, 20 % p/v PEG 3350.

Concentration GH*	Nom du kit + puits	Condition	Cristaux
12,5 mg/ml	JCSG I A10	0,2M acétate de Ca 0,1M imidazole pH8 20% PEG 1000	Amas de cristaux
12,5 mg/ml	JCSG I B8	0,2M acétate de Ca 20% PEG 3350	Petites baguettes (figure 1A)
12,5 mg/ml	JCSG I C2	0,2M LiNO3 20% PEG 3350	Petites baguettes très fines partant d'un même point
12,5 mg/ml	JCSG I C11	0,2M KNO3 20% PEG 3350	Petites baguettes très fines partant d'un même point
12,5 mg/ml	JCSG I F2	0,2M MgSO4 20% PEG 3350	Petites baguettes
12,5 mg/ml	JCSG II B3	0,2M acétate de Ca 0,1M imidazole pH8 20% PEG 8000	Amas de cristaux
8mg/ml	JCSG I A12	0,2M acétate de Mg 20% PEG 3350	Petites baguettes (figure 1C)
8mg/ml	JCSG I B8	0,2M acétate de Ca 20% PEG 3350	Baguettes (figure 1B)
8mg/ml	JCSG I F2	0,2M MgSo4 20% PEG 3350	Baguettes
8mg/ml	JCSG I F3	0,2M Mg formate 20% PEG 3350	Baguettes
8mg/ml	JCSG II B3	0,2M acétate de Ca 0,1M imidazole pH8 20% PEG 8000	Baguettes
3,4mg/ml	JCSG I E2	0,2M NH4Cl 20% PEG 3350	Grande baguette (figure 2B)
3,4mg/ml	JCSG IV B9	0,2M MgCl2 0,1M Tris pH8,5 30% PEG 4000	Petites baguettes (figure 2A)

Tableau 13 : Conditions de cristallisation des cribles commerciaux ayant permis l'obtention des premiers cristaux.

Boite n°	Conditions en Sel	Conditions en PEG	Concentration GH*
1	Acétate de calcium à 0,10; 0,15; 0,20; 0,25; 0,30 et 0,35M	PEG 3350 à 15, 20, 25 et 30%	GH* à 15 et 7,5mg/ml
2	Acétate de magnésium à 0,10; 0,15; 0,20; 0,25; 0,30 et 0,35M	PEG 3350 à 15, 20, 25 et 30%	GH* à 15 et 7,5mg/ml
3	Acétate de calcium à 0,10; 0,15; 0,20; 0,25; 0,30 et 0,35M	PEG 3350 à 15, 20, 25 et 30%	GH* à 8, 4 et 2mg/ml
4	Acétate de magnésium à 0,10; 0,15; 0,20; 0,25; 0,30 et 0,35M	PEG 3350 à 15, 20, 25 et 30%	GH* à 8, 4 et 2mg/ml
5	Acétate de calcium à 0,10; 0,15 et 0,20M	PEG 3350 à 25%	GH* à 4,6mg/ml

Tableau 14 : Conditions de cristallisation optimisés manuellement en boîtes 24 puits.

Différents cristaux obtenus avec les cribles commerciaux présentés en figure 59 ont donc été prélevés et congelés en utilisant la solution de réservoir supplémentée en PEG 400 à une concentration finale de 15 % (v/v) afin de les cryoprotéger. Les cristaux ont été testés sur le site synchrotron de l'ESRF (Grenoble, France). Ces cristaux présentaient un pouvoir diffractant faible compris entre 5 et 3,5 Å.

De nouveaux essais ont donc été menés en diminuant encore la concentration en protéine à 3,4 mg/ml. De plus, comme entre-temps, des activités faibles avaient été identifiées pour GH*, les essais ont été réalisés en présence de L-arabinose ou de D-galactose, pour lesquels une faible activité sur pNP-monosaccharide avait été observée. De nombreux cribles (JCSG I à IV, PEG I et II, pHclear I et II, Morpheus et Nucleix) ont été utilisés par Claire Raingeval, lors de son stage de M1, et ont permis d'obtenir des cristaux. En présence de D-galactose, les cristaux ont la même forme que précédemment mais sont plus grands (figures 60A) (tableau 13) et n'ont pas été obtenus dans les mêmes conditions : 0,2 M chlorure de magnésium, 0,1 M Tris pH 8,5, 30 % p/v PEG 4000. En présence de L-arabinose et à plus faible concentration en protéine, des cristaux ont poussé dans une nouvelle condition (0,2 M chlorure d'ammonium, 20 % p/v PEG 3350). Ces cristaux n'ont pas la même forme que ceux obtenus jusqu'alors (figure 60B) (tableau 13).

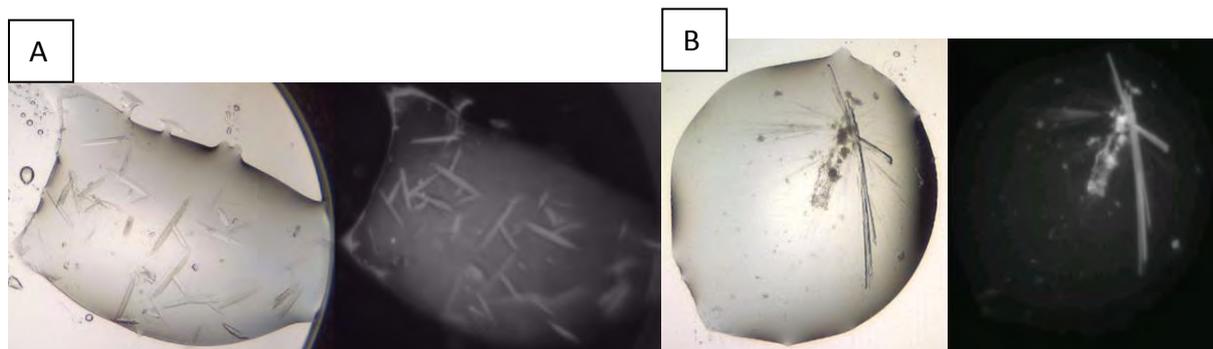


Figure 60 : A. Cristaux obtenus avec GH* à 3,4 mg/ml en présence de 0,2 M chlorure de magnésium, 0,1 M Tris pH 8,5 et 30 % p/v PEG 4000. B. Cristaux obtenus avec GH* à 3,4 mg/ml en présence de 0,2 M chlorure d'ammonium, 20 % p/v PEG 3350. A gauche la goutte visualisée en lumière visible et à droite en UV.

Ces nouveaux cristaux ont été prélevés et congelés en utilisant la solution de cristallisation contenant 100 mM de L-arabinose ou de D-galactose permettant d'assurer leur cryoprotection. Ils ont été testés sur site synchrotron (ESRF, Grenoble, France) et un cristal a permis d'obtenir un jeu de données ayant une résolution maximale de 1,55 Å. Ce cristal est un fragment de la baguette présentée sur la figure 60B.

Ces données semblent indiquer que le L-arabinose joue un rôle dans la cristallisation. Deux autres éléments qui semblent essentiels à la cristallisation sont la présence de PEG (3350 ou 4000) à 20% p/v (voire 30% p/v) et la présence de sel à 0,2M car toutes les conditions dans lesquelles les meilleurs cristaux ont été obtenus comportent du PEG et du sel.

IV. Résolution de la structure

A. Données de diffraction

Les données de diffraction ont été collectées sur la ligne de lumière ID23-eh2 de l'ESRF (Grenoble, France) et indexées, intégrées et mises à l'échelle à l'aide du logiciel XDS (Kabsch 1993). Les statistiques du traitement des données sont présentées dans le tableau 15.

GH* cristallise dans un groupe d'espace orthorhombique $P2_12_12_1$ et présente une seule molécule dans l'unité asymétrique avec un pourcentage de solvant de 35 %. Le cristal diffracte à une résolution maximale de 1,55 Å (tableau 15).

Les cristaux diffractant à haute résolution et en l'absence de protéine de structure connue suffisamment homologue, nous avons donc envisagé l'utilisation du signal anomal du soufre dans des expériences de phasage S-SAD. En effet, dans la séquence de GH* on compte trois cystéines et sept méthionines dont deux sur la séquence de l'étiquette à 6 histidines qui a été conservée pour la cristallisation. Plusieurs jeux de données ont ainsi été collectés sur la ligne de lumière ID30A de l'ESRF (Grenoble, France) mais le signal anomal du soufre s'est avéré insuffisant pour permettre la détermination de la structure par cette méthode.

B. Obtention de la structure

Longueur d'onde (Å)	0,87260
Gamme de résolution (Å)	49,04 - 1,55 (1,64 - 1,55)
Groupe d'espace	P 2 ₁ 2 ₁ 2 ₁
Paramètres de maille	a=61,66Å b=63,46Å c=77,23Å α=β=γ=90°
Nombre total de réflexions	122283 (18392)
Nombre de réflexions uniques	43486 (6756)
Multiplicité	2,81 (2,72)
Complétude (%)	96,9 (94,8)
Moyenne I/sigma(I)	7,11 (1,41)
B-facteur de Wilson (Å ²)	14,81
R-merge	0,098 (0,717)
R-meas	0,12 (0,884)
CC1/2	99,5 (52,5)
Nombre de réflexions pour calculer le R-free	2124
R-work	0,1657 (0,3124)
R-free	0,1994 (0,3373)
Nombre d'atomes (hors hydrogène)	2869
protéines	2470
eau	399
Nombre de résidus dans la protéine	298
RMS(liaisons)	0,007Å
RMS(angles)	1,07°
Ramachandran favorable (%)	97
Ramachandran autorisé (%)	3
Ramachandran non autorisé (%)	0
B-facteur moyen (Å ²)	20,90
protéines (Å ²)	18,80
solvant (Å ²)	33,90
Statistiques de la couronne externe entre parenthèses	

Tableau 15 : Présentation des statistiques de la résolution de la structure de GH*.

La structure cristallographique de GH* a pu être obtenue par remplacement moléculaire en utilisant comme modèle l'endo-β-1,4-glucanase de *Fervidobacterium nodosum* FnCel5A

(code PDB : 3NCO) appartenant à la famille GH5 (Zheng et al. 2012). Une solution de remplacement moléculaire correcte a pu être obtenue à l'aide du logiciel Phaser (McCoy et al. 2007) malgré la très faible identité de séquence (14 %) avec GH*. Le modèle a été construit manuellement dans les cartes de densité électroniques 2mFo-DFc pondérées par les SigmaA en utilisant WinCoot (Emsley et Cowtan 2004) et plusieurs cycles d'affinement ont été effectués avec Refmac5 (Murshudov et al. 2011). Comme le montre l'alignement de séquences (figure 60) entre GH* et FnCel5A, la conservation des éléments de structure secondaires des deux protéines a certainement aidé lors du remplacement moléculaire malgré la faible identité de séquence. Le Rfacteur et le Rlibre sont respectivement de 0,165 et 0,200 en fin d'affinement. Le modèle final possède 97% des résidus dans les régions les plus favorables du diagramme de Ramachandran et les 3% restant se situent dans des régions autorisées.

GH* adopte un repliement de type tonneau (β/α)₈ constitué d'un tonneau de brins β central entouré par des hélices α . L'analyse par PISA (Krissinel et Henrick 2007) indique que la protéine est monomérique dans le cristal. Ce résultat est en accord avec ce qui a été observé par chromatographie d'exclusion-diffusion puisque le pic majoritaire du chromatogramme présente un volume d'élution compatible avec une protéine de 40 kDa ($V_e = 77$ ml, figure 50C). Seules les fractions correspondant à ce pic d'élution ont été regroupées et utilisées pour la cristallisation.

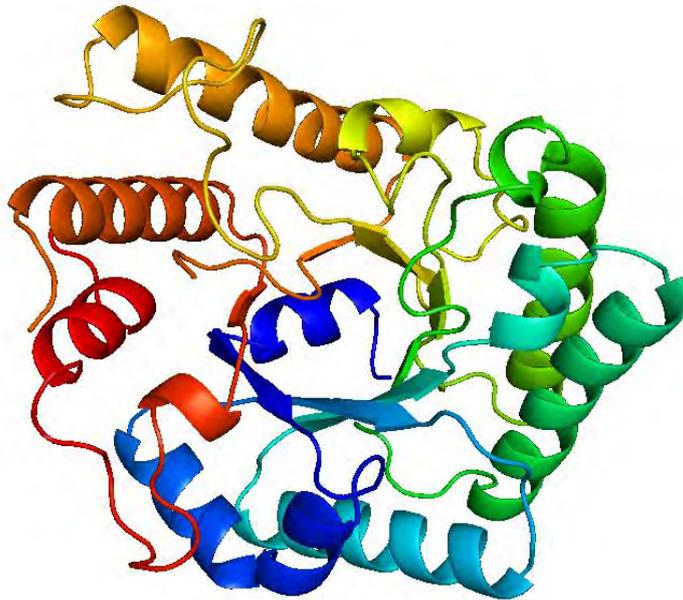


Figure 62 : Représentation en cartoon de la structure tridimensionnelle de GH*. Le dégradé de couleur commence à l'extrémité N-terminale de la protéine avec la couleur bleue et termine à l'extrémité C-terminale par la couleur rouge.

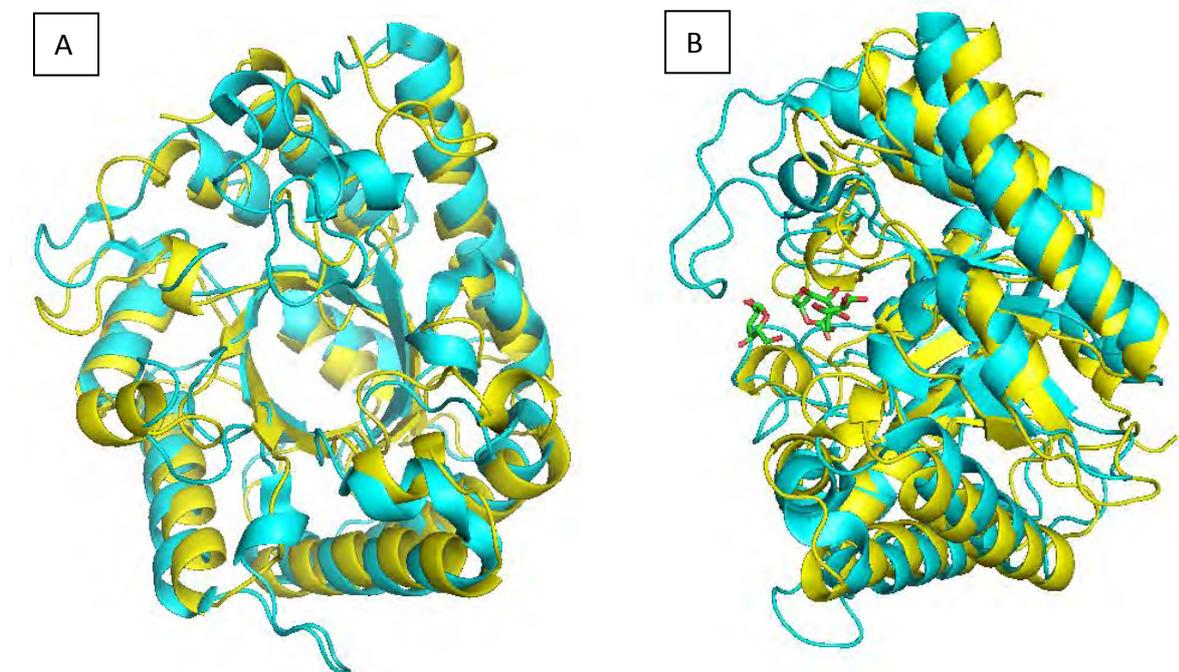
V. Analyse de la structure de GH*

GH* présente un repliement en tonneau (β/α)₈ caractéristique des enzymes appartenant au clan GH-A. Quand on superpose GH* et l'endo- β -1,4-glucanase de *Fervidobacterium nodosum* (FnCel5A) on peut voir que les éléments de structure secondaire ne sont pas tout à fait superposés (figure 63A). En effet, lorsqu'on calcule le RMSD sur les carbones α , on se rend compte que l'alignement n'est pas très bon avec une valeur de 2,4 Å sur 143 carbones α alignés. Les brins β sont bien superposés malgré quelques décalages des extrémités C-

terminales de chaque brin. En effet, le RMSD calculé uniquement sur l'alignement des résidus composant les tonneaux β donne une valeur de 0,6 Å sur 40 carbones α alignés. Les éléments présentant le plus de divergence sont les boucles qui sont de tailles différentes et donc qui peuvent présenter des différences de flexibilité. Un pont disulfure entre les cystéines 252 (sur l'hélice α 8) et 293 (sur la boucle après l'hélice α 9) participe au repliement de GH*. Ce pont disulfure n'est pas retrouvé chez FnCel5A.

On voit aussi sur la représentation de profil de la superposition (figure 63B et C) qu'il y a une longue boucle chez FnCel5A qui referme le site catalytique en haut et qui n'est pas retrouvée dans la structure de GH* (figure 61 encadré bleu). Cette boucle participe au positionnement et au maintien du substrat dans le site catalytique de la GH5, notamment par l'intermédiaire du tryptophane 240. Cette boucle existe dans d'autres GH5 et présente aussi un tryptophane capable d'interagir avec le substrat (figure 63D).

La superposition de la structure de GH* avec celle de FnCel5A en complexe avec un substrat a permis de confirmer la localisation du site catalytique putatif de GH* déterminé par alignement de séquence et l'identification du résidu jouant le rôle d'acide/base chez GH* (E117).



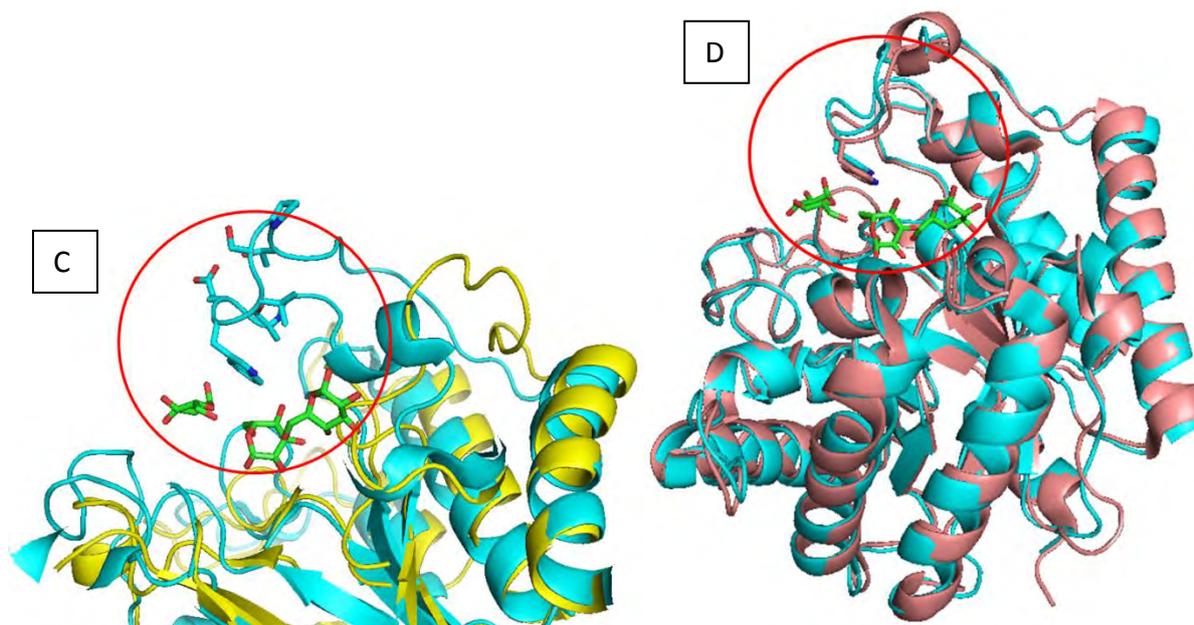


Figure 63 : Superposition des structures 3D de GH* et FnCel5A. Vue de face de la superposition (A). Vue de profil de la superposition en présence d'un substrat (glucose) de l'endo- β -1,4-glucanase de *Fervidobacterium nodosum* FnCel5A (code PDB : 3RJY) (B). Zoom sur la boucle supplémentaire de FnCel5A (C). Superposition de FnCel5A avec une autre GH5 (une endo- β -1,4-glucanase de *Thermotoga maritima*) (D). L'interaction entre la boucle et le substrat est entourée en rouge. GH* est représenté en jaune, FnCel5A en bleu et la GH5 en rose. La représentation en bâtons vert et rouge correspond aux 3 D-glucopyranoses.

Quand on compare les représentations de surface électrostatique des deux protéines, on constate que le site catalytique de FnCel5A est globalement électronégatif (figure 64A), ce qui est le cas chez les différentes GH de la famille 5 (figure 65), alors que le site catalytique putatif de GH*, déterminé par la superposition des deux structures, est globalement électropositif en surface en raison de la présence de sept arginines (R17, R20, R23, R42, R44, R156 et R209) (figure 64B et C). Cette différence de charges pourrait être impliquée dans la sélection, le chargement et l'orientation du substrat, et nous a orienté vers une activité de GH* sur des oligo- et polysaccharides chargés négativement. Plusieurs substrats (héparosane, acide aldo-uronique, alginate et carraghénane) ont donc été testés pour répondre à cette hypothèse, cependant aucune activité n'a pu être détectée. Il est intéressant de noter que les enzymes appartenant au clan GH-A possèdent aussi un site catalytique électronégatif (figure 66). GH* serait donc la seule enzyme de ce clan à posséder un site catalytique présentant une zone très électropositive en surface.

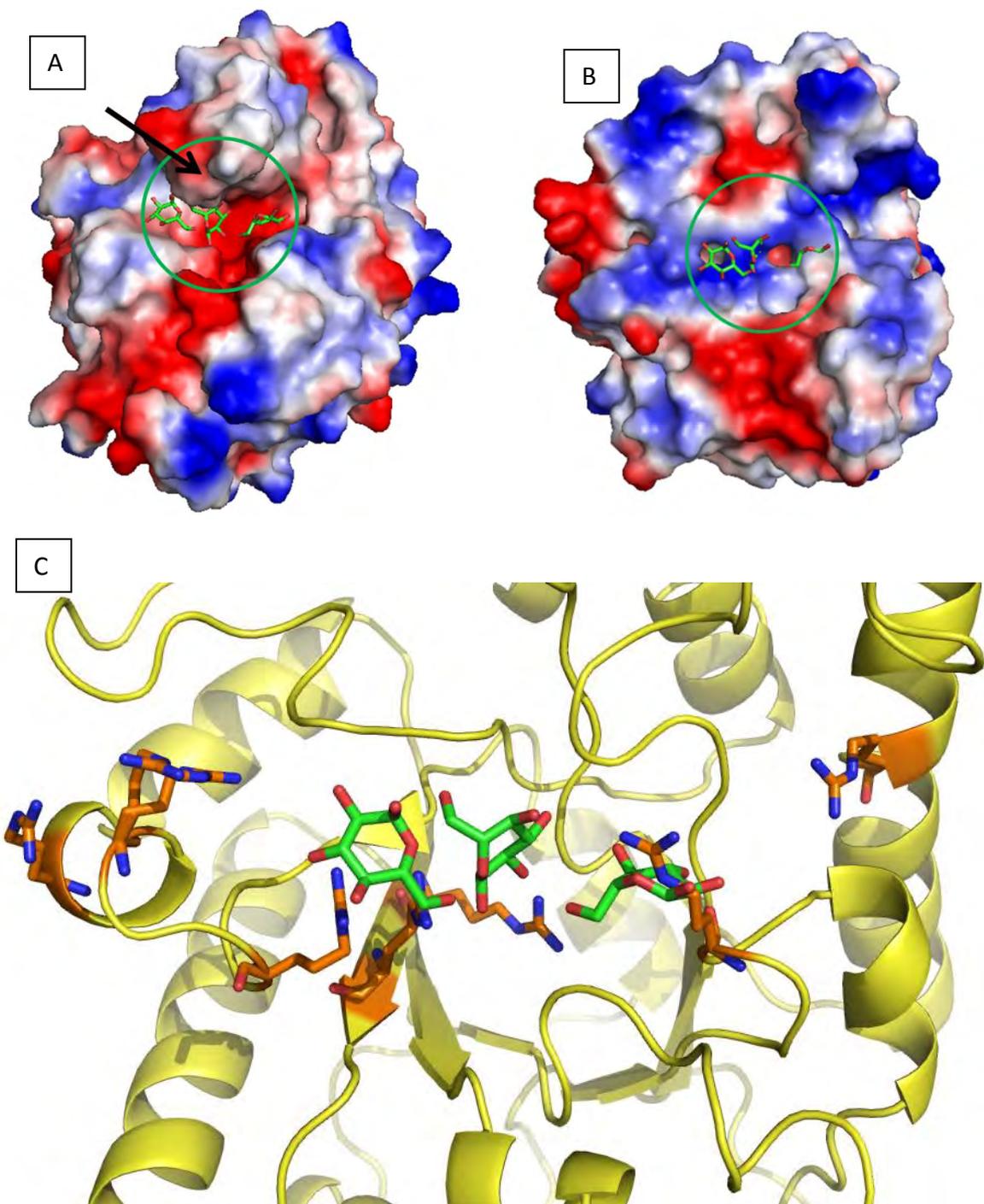


Figure 64 : Représentation de surface électrostatique de l'endo-β-1,4-glucanase de *Fervidobacterium nodosum* FvCel5A (A) et de GH* (B). Le site catalytique de FvCel5A et le site catalytique putatif de GH* sont entourés en vert. Les zones chargées négativement sont visualisées en rouge et celles chargées positivement sont visualisées en bleu. Les zones neutres sont en gris. Les D-glucopyranoses issus de la structure de FvCel5A sont représentés en bâtonnets verts (carbones) et rouges (oxygènes). En (A) la flèche noire pointe sur la boucle de FvCel5A qui ferme le site catalytique

en haut. (C) Localisation des 7 arginines notées R, responsables de l'électropositivité du site catalytique hypothétique de GH*.

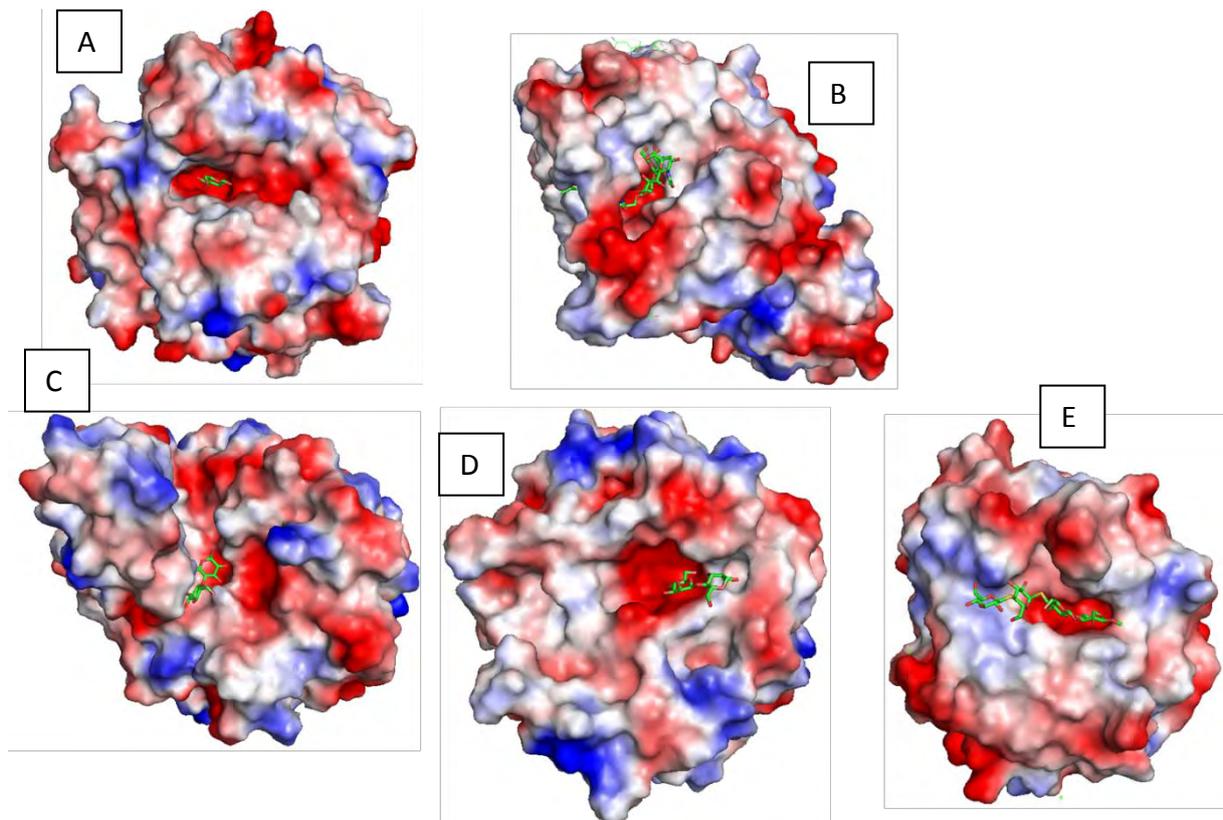


Figure 65 : Représentation de surface de plusieurs GH5. A : endo- β -1,3-glucanase de *Candida albicans*. B : endoglycocéramidase II de *Rhodococcus sp.* C : endo- β -1,4-glucanase de *Ruminiclostridium thermocellum*. D : β -mannanase de *Trichoderma reesei*. E : endo- β -1,4-glucanase de *Bacillus agaradhaerens*. Les zones chargées négativement sont visualisées en rouge et celles chargées positivement sont visualisées en bleu. Les zones neutres sont en gris. Les substrats des enzymes sont représentés en bâtonnets verts (carbone) et rouges (oxygène).

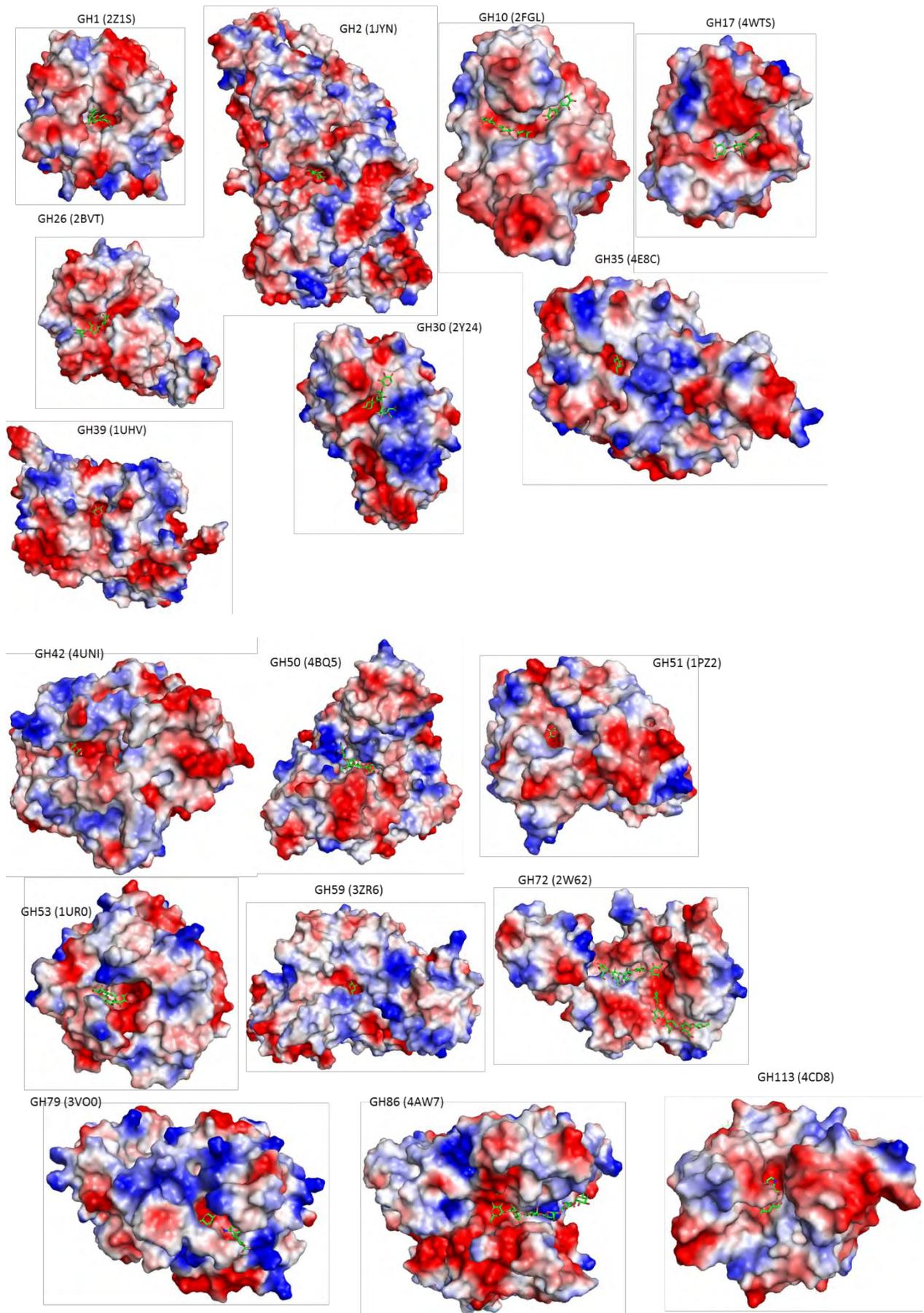


Figure 66 : Représentation de surface d'un membre de chaque famille de GH appartenant au clan des GH-A. GH1 : β -glucosidase de *Paenibacillus polymyxa*, GH2 : β -galactosidase de *E. coli*, GH10 : endo- β -1,4-xylanase de *Bacillus sp.*, GH17 : β -1,3-glucanosyltransferase de *Rhizomucor miehei*, GH26 : β -mannanase de *Cellulomonas fimi*, GH30 : glucuronoxylanase de *Dickeya chrysanthemi*, GH35 : β -1,3-galactosidase de *Streptococcus pneumoniae*, GH39 : β -xylosidase de *Thermoanaerobacterium saccharolyticum*, GH42 : β -1,6/1,3-galactosidase de *Bifidobacterium animalis subsp.* GH50 : exo- β -agarase de *Saccharophagus degradans*, GH51 : α -L-arabinofuranosidase de *Geobacillus stearothermophilus*, GH52 : endo- β -1,4-galactanase de *Bacillus licheniformis*, GH59 : galactocerebrosidase de *Mus musculus*, GH72 : β -1,3-glucanosyltransférase de *Saccharomyces cerevisiae*, GH79 : β -glucuronidase de *Acidobacterium capsulatum*, GH86 : β -porphyranase de *Bacteroides plebeius* et GH113 : endo- β -1,4-mananase de *Alicyclobacillus acidocaldarius*. Les zones chargées négativement sont visualisées en rouge et celles chargées positivement sont visualisées en bleu. Les zones neutres sont en gris. Les substrats des enzymes sont représentés en bâtonnets verts (carbones) et rouges (oxygènes).

Sur la représentation de surface électrostatique, on note aussi que le site catalytique de FnCel5A présente une poche dans laquelle vient se fixer le substrat. On retrouve aussi la boucle (indiquée par la flèche noire de la figure 64A) qui vient fermer le haut de cette poche. Quand on regarde la représentation de surface de GH*, on constate que la poche catalytique hypothétique de GH* est fermée par les arginines 20 et 156 (figure 64B et C). L'arginine 20 a deux positions alternatives qui ont été observées. D'autre part, on peut voir sur la figure 64C que l'arginine 156 se situe au même endroit qu'un des D-glucopyranoses appartenant à la structure de FnCel5A. Il est donc possible qu'afin d'accommoder le substrat, la chaîne latérale de cette R156 soit aussi capable de pivoter. Ces mouvements pourraient aussi impliquer un réarrangement d'autres résidus du site catalytique lors de la fixation du substrat.

Si on regarde un peu plus en détail la superposition des sites catalytiques de l'endo- β -1,4-glucanase de *Fervidobacterium nodosum* FnCel5A et GH*, on peut identifier les résidus qui pourraient être impliqués dans l'activité de GH* (figure 67).

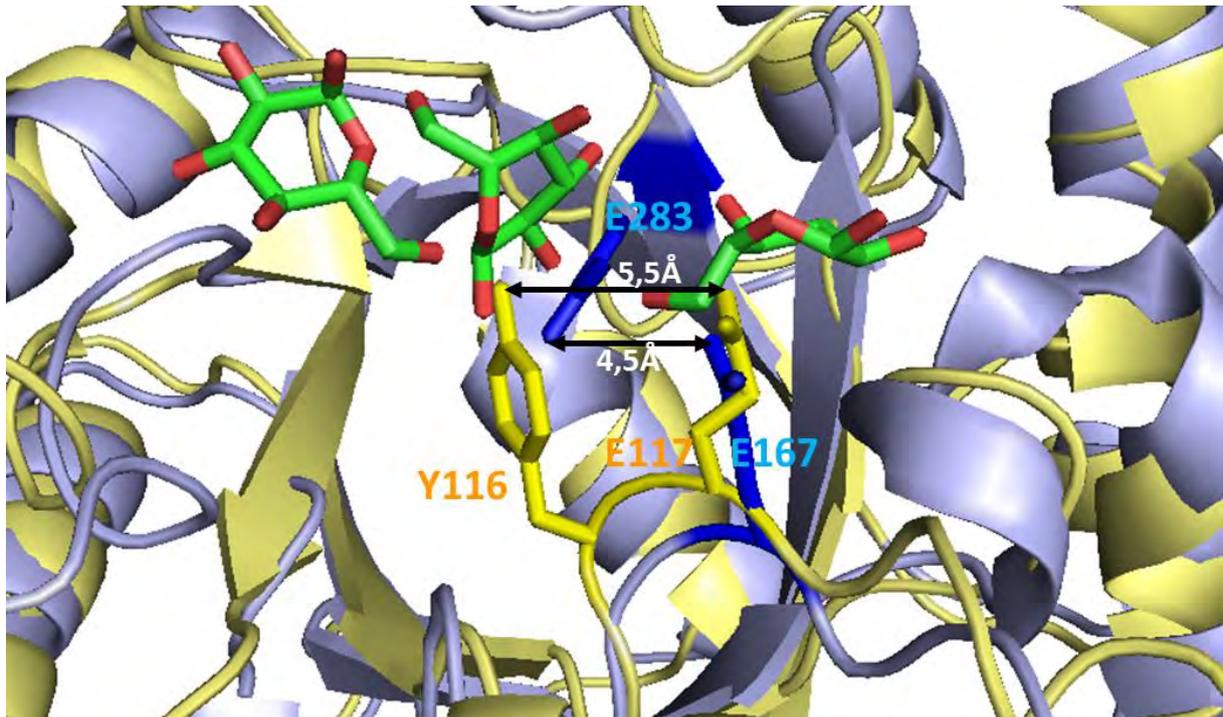


Figure 67 : Superposition du site catalytique de FnCel5A et du site catalytique hypothétique de GH*. FnCel5A est représentée en bleu pâle et GH* en jaune clair. Les résidus catalytiques de FnCel5A sont représentés en bleu et les résidus catalytiques hypothétiques de GH* en jaune. Les D-glucopyranoses issus de la structure de FnCel5A sont représentés en bâtonnets verts (carbone) et rouges (oxygène). Les flèches noires symbolisent les distances entre E167 et E283 ainsi qu'entre E117 et Y116.

Les résidus catalytiques de l'endo- β -1,4-glucanase de *Fervidobacterium nodosum* FnCel5A sont deux glutamates. Le glutamate E167 joue le rôle d'acide/base et le glutamate E283 joue le rôle de nucléophile. Les groupements carboxyles de ces résidus sont espacés de 4,5 Å. Cette distance est significativement plus courte que celle proposée dans un mécanisme de rétention de configuration (environ 5,5 Å) (McCarter et Stephen Withers 1994 ; Davies et Henrissat 1995).

Dans la structure de GH*, un glutamate est présent en position proche du E167 de FnCel5A. Il s'agit de E117 et ce résidu pourrait donc être l'acide/base de GH* qui avait été prédit comme présent par B. Henrissat et confirmé par l'alignement de séquence entre GH* et des GH5. E117 se situe sur la boucle présente entre le brin β 4 et l'hélice α 6. Par contre, l'observation de la structure de GH* ne permet pas d'identifier de résidu chargé porteur de groupement carboxylique (glutamate ou aspartate) chez GH* à proximité spatiale de E283

de l'endo- β -1,4-glucanase de *Fervidobacterium nodosum* FnCel5A pouvant jouer le rôle de nucléophile. Seules deux glycine sont présentes dans cet environnement. En s'éloignant un peu de cette position en amont et en aval sur la séquence, on se rend compte qu'il n'y a pas de résidus pouvant jouer le rôle de nucléophile. Cette absence de résidu nucléophile chez GH* avait été prédite sur la base des alignements de séquences. Pour identifier un résidu capable de jouer le rôle de nucléophile, l'angle de recherche a été changé, plutôt que se déplacer sur la chaîne principale nous avons regardé dans l'environnement du groupement fonctionnel donc à l'extrémité de la chaîne latérale de E283 de FnCel5A. Dans la structure de GH*, dans la région correspondant au groupement carboxylique de E283 de FnCel5A, on constate la présence d'un groupement hydroxyle porté par la chaîne latérale de la tyrosine Y116. La distance entre l'oxygène du groupement hydroxyle de Y116 et le carboxyle de E117 de GH* est de 5,5 Å. Cette distance est compatible avec un mécanisme de rétention de configuration. Sachant qu'un tel résidu jouant le rôle de nucléophile a déjà été observé dans la famille des sialidases GH33 et GH34, nous proposons que cette tyrosine puisse être le résidu nucléophile.

D'autre part, quand on regarde le reste du site catalytique putatif de GH*, on constate que d'autres résidus sont bien conservés et notamment l'histidine 180 de GH* qui se superpose très bien avec l'histidine 226 de l'endo- β -1,4-glucanase de *Fervidobacterium nodosum* FnCel5A. Cette histidine est impliquée dans le maintien du substrat dans le site catalytique selon Zheng et collaborateurs (Zheng et al. 2012). Pour vérifier si l'histidine 180 de GH* joue aussi un rôle dans le maintien du substrat, nous avons envisagé de faire le mutant H180A. La présence du tryptophane 184 de GH* situé en position proche de W240 de FnCel5A a aussi été remarquée. Ce type de résidu pourrait être impliqué dans le maintien du substrat dans le site catalytique. Un autre résidu identifié chez FnCel5A comme étant impliqué dans le maintien du substrat est l'histidine 127. Chez GH* en position proche de cette histidine se trouve la glutamine 76 qui pourrait aussi être impliquée dans le maintien du substrat.

Si ce résidu est essentiel à la catalyse enzymatique alors, la mutation de ce résidu devrait se traduire par une perte significative d'activité de GH*.

Différents mutants de GH* ont donc été envisagés pour vérifier l'implication de différents acides aminés dont le résidu prédit comme étant l'acide/base E117 et le résidu que nous supposons être le nucléophile Y116. Les mutants E117A, E117Q et Y116A ont été produits et purifiés comme GH* sauvage et ont été testés sur arabinoxylane de seigle. Ces 3 mutants sont inactifs avec une perte de 100% de l'activité pour le mutant Y116A et d'environ 97% de l'activité pour les mutants de E117 (figure 68). Ces résultats prouvent l'implication et le rôle essentiel de Y116 et de E117 dans le mécanisme catalytique de GH*. En revanche ces expériences ne suffisent pas à prouver que Y116 de GH* est bien le résidu nucléophile.



Figure 68 : Graphique présentant l'activité spécifique relative de GH* sauvage et des mutants Y116A, E117A et E117Q de GH* sur RAX. L'enzyme a été incubée 3 h en présence de 2 mg/ml de RAX dans un tampon phosphate-citrate (43 mM citrate, 114 mM phosphate) pH 5,5 et à 45 °C. GH* sauvage (GH* wt).

VI. Mécanisme catalytique

Le fait que la tyrosine 116 de GH* semble être le résidu nucléophile implique la nécessité de confirmer le mécanisme d'action par rétention de la configuration caractéristique du clan GH-A. En effet, même si la distance entre les groupements portés par Y116 et E117 est la distance caractéristique d'un mécanisme par rétention de configuration (environ 5,5 Å), il reste à identifier la position du substrat car lorsque nous superposons la structure d'un complexe de l'endo-β-1,4-glucanase de *Fervidobacterium nodosum* FnCel5A sur celle de

GH*, les 2 résidus catalytiques de cette dernière se trouveraient du même côté du substrat. Dans ce type de mécanisme, ils se situent de part et d'autre du plan de la liaison à hydrolyser (Mark et al. 2001) (Vuong et Wilson 2010).

Dans la base de données CAZy, il existe un mécanisme catalytique impliquant une tyrosine (Watson et al. 2003 ; Amaya et al. 2004 ; Watts et al. 2006). La tyrosine joue le rôle de nucléophile et est activée par un acide glutamique situé à environ 2,55 Å (Amaya et al. 2004 ; Watts et al. 2006). Le résidu acide/base est un acide glutamique ou un acide aspartique. Ce mécanisme est retrouvé dans les familles de sialidases GH33 et 34 mais ces enzymes ont un repliement totalement différent puisqu'il s'agit d'un repliement en hélice à 6 pales (6 fold β -propeller). Chez GH*, il n'y a pas de glutamate ou d'aspartate retrouvé en face de la tyrosine et de l'acide glutamique. Par contre, un aspartate présent un peu plus loin pourrait jouer ce rôle d'activateur du nucléophile par l'intermédiaire d'une molécule d'eau, il s'agit de D12 (figure 69).

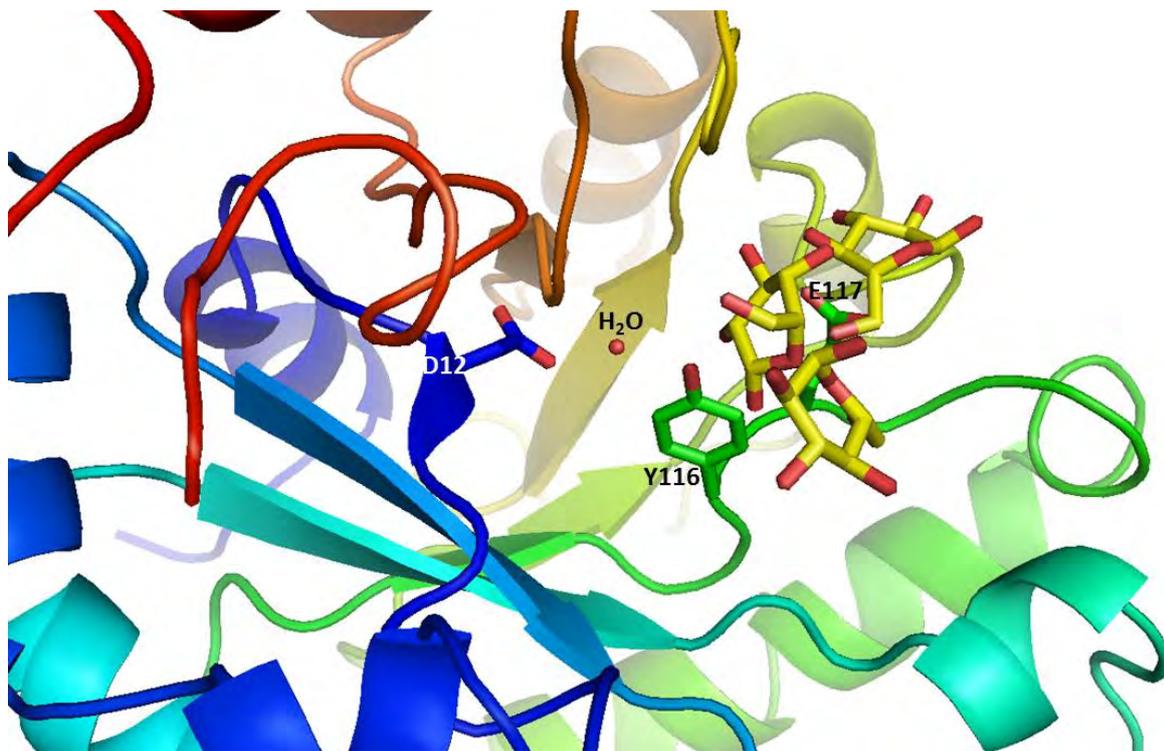


Figure 69 : Zoom sur le site catalytique hypothétique de GH*. La molécule d'eau pouvant être l'intermédiaire entre D12 et Y116 est entourée en violet. Les D-glucopyranoses issus de la structure de FnCel5A sont représentés en bâtonnets verts (carbones) et rouges (oxygènes).

Afin d'obtenir une structure avec le substrat cocristallisé et pouvoir confirmer les hypothèses et découvrir d'autres acides aminés importants dans l'interaction avec le substrat, des essais de cristallisation des mutants de GH* Y116A et E117Q en présence de X6 ont été réalisés par Franck Moncassin. Des cristaux ont été obtenus en présence de 0,2 M chlorure de magnésium, 0,1 M Tris pH 8,5 et 30 % PEG 4000 (figure 70) mais n'ont pas permis d'obtenir la structure de GH* en complexe avec un substrat car la diffraction maximale des jeux de données enregistrés sur le site synchrotron de Alba (Barcelone, Espagne) ne dépasse pas 4 Å. Toutefois, à partir de ces conditions, des optimisations manuelles vont pouvoir être envisagées.



Figure 70 : Cristaux obtenus avec GH* Y116A à 3,4 mg/ml et X6 à 30 mM en présence de 0,2 M chlorure de magnésium, 0,1 M Tris pH 8,5 et 30 % PEG 4000. A gauche la goutte visualisée en lumière visible et à droite en UV.

VII. Conclusions et perspectives

La structure de GH* a pu être résolue par remplacement moléculaire avec une GH5, l'endo- β -1,4-glucanase de *Fervidobacterium nodosum* FnCel5A, comme modèle. GH* a un repliement classique des GH du clan A.

On a pu voir que contrairement aux GH5, GH* a un site catalytique putatif électropositif dû à la présence de nombreuses arginines à sa surface.

Deux résidus ont pu être identifiés comme étant impliqués dans l'activité catalytique de GH*. Il s'agit de la tyrosine 116 et du glutamate 117. L'hypothèse émise est que le glutamate jouerait le rôle de l'acide/base car il est bien conservé par rapport aux GH5, et que la tyrosine jouerait le rôle de nucléophile. La 2^{ème} partie de l'hypothèse semble moins

vraisemblable car les deux résidus catalytiques sont côte à côte sur la même chaîne polypeptidique.

Si ce glutamate et cette tyrosine sont réellement les résidus catalytiques, cela implique un mécanisme différent de ceux connus à ce jour.

La résolution de la structure des mutants de GH* en complexe avec du X6 pourra nous donner plus de renseignements quant à l'orientation du substrat dans le site catalytique ainsi que sur les différents résidus impliqués dans le maintien du substrat et dans l'hydrolyse.

Afin de confirmer ou d'infirmer le mécanisme de rétention de configuration suspecté pour GH*, des études par RMN du proton en utilisant un analogue de X5, le pNP-X4, permettront de déterminer précisément le mécanisme mis en jeu lors de l'hydrolyse d'un xylooligosaccharide.

De plus, caractériser structurellement et biochimiquement GH* avec d'autres substrats pourrait permettre d'identifier d'autres résidus impliqués et notamment ceux présents dans les sous-sites -3 et +3 qui sont essentiels à l'activité de l'enzyme.

Conclusions générales

Les objectifs de ces travaux de thèse se divisaient en 2 grandes parties. La première consistait en la comparaison du potentiel hémicellulolytique de deux espèces de termites *N. corniger* et *T. hispaniolae* par l'analyse de leur métagénome. La seconde était de caractériser du point de vue fonctionnel et structural une nouvelle glycoside hydrolase hypothétique provenant du métagénome d'un ver de Madagascar.

Dans la première partie de cette thèse, afin de comparer le potentiel hémicellulolytique des termites *N. corniger* et *T. hispaniolae*, il fallait traiter les données de séquençage de chaque métagénome. La majeure partie de l'annotation fonctionnelle et taxonomique du métagénome du termite *N. corniger* a été effectuée par Gregory Arnal pendant ses travaux de thèse (Arnal 2014). La totalité de l'annotation fonctionnelle et taxonomique du métagénome du termite *T. hispaniolae* a été effectuée dans ces travaux de thèse. Toutes ces annotations ont pu être effectuées grâce au pipeline mis en place dans l'équipe lors des travaux de thèse de Géraldine Bastien avec l'aide de Sandrine Laguerre (Bastien-Uluis 2012). Ces analyses ont permis de découvrir un grand nombre de nouvelles CAZymes qui sont pour la plus grande majorité des GH, mais aussi des lyases et des modules de liaison aux glucides. Dans le métagénome de *N. corniger*, 27 familles de GH différentes ont pu être identifiées et 23 dans celui de *T. hispaniolae*. On a aussi pu voir que les annotations des banques sont en bonne corrélation avec les annotations faites dans d'autres études (Warnecke et al. 2007 ; He et al. 2013).

Une fois les annotations faites, tout le travail de comparaison a pu être effectué. Les 2 termites n'ont pas la même alimentation, *N. corniger* est xylophage strict alors que *T. hispaniolae* est humivore et xylophage. Lorsque le crible fonctionnel sur substrat chromogénique a été réalisé sur chacune des banques métagénomiques, les activités trouvées dans les deux banques étaient différentes. Une tendance semblait alors se dégager avec des activités plutôt endoglycosidases pour *N. corniger* alors que pour *T. hispaniolae*, les activités étaient plutôt exoglycosidases. L'analyse des résultats de l'annotation a permis de

confirmer cette tendance même si, finalement, les rapports entre activité endoglycosidase et activité exoglycosidase tendent à s'équilibrer dans chacune des 2 banques.

D'autre part, l'annotation taxonomique montre une grande différence dans les microbiotes des 2 espèces de termites. Le microbiote intestinal de *N. corniger* est majoritairement composé de bactéries appartenant aux embranchements *Spirochaetes* et *Fibrobacteres*, avec essentiellement les genres *Treponema* et *Fibrobacter*, respectivement. Ces résultats sont aussi en accord avec la littérature (Warnecke et al. 2007 ; He et al. 2013).

Le microbiote intestinal de *T. hispaniolae* est, quant à lui, majoritairement composé de bactéries appartenant aux embranchements des *Firmicutes* et des *Bacteroidetes*. L'annotation taxonomique n'a pas permis d'aller plus loin que l'embranchement pour ce terme. Par contre l'annotation taxonomique a montré une plus grande variété d'embranchements dans le microbiote intestinal de *T. hispaniolae* que dans celui de *N. corniger*. La plus grande variété du microbiote de *T. hispaniolae* semble en accord avec l'alimentation plus variée de ce termite.

Pour aller plus loin dans ce travail, la caractérisation biochimique de certaines protéines est envisageable. Certaines associations bi- ou tri-modulaires, comme par exemple l'association CE3-GH5-CBM16 identifiée dans la banque métagénomique de *T. hispaniolae*, seraient intéressantes à étudier fonctionnellement et structurellement afin de mieux comprendre le fonctionnement de ces enzymes ainsi que les stratégies mises en place par les termites et leur microbiote pour dégrader la biomasse végétale. Par exemple, au laboratoire, une protéine possédant une organisation originale fait l'objet de deux thèses. Il s'agit d'une protéine issue de la banque métagénomique de *Pseudacanthotermes militaris*. C'est une GH10 présentant deux CBM4 au milieu de sa séquence catalytique : GH10a-CBM4_1-CBM4_2-GH10b. Cette protéine est active et les données de modélisation moléculaire semblent indiquer que les deux parties de la GH10 se rapprochent pour reformer le repliement classique en tonneau (β/α)₈ des GH10. Il semblerait aussi que cette enzyme soit plus active en présence d'autres enzymes (cocktail enzymatique) (résultats des thèses en cours de H. Wu et E. Ioannou).

La deuxième grande partie de ces travaux de thèse portait sur l'étude structurale et fonctionnelle d'une enzyme annotée comme étant une glycoside hydrolase n'appartenant à

aucune famille connue appelée GH*. Cette protéine est issue de la banque métagénomique du ver de Madagascar *P. corethrurus*. Comme GH* n'appartient à aucune famille, il n'y avait aucune piste pour son activité potentielle et il fallait donc prouver que cette enzyme était réellement active. La totalité des substrats disponibles au laboratoire et/ou commercialisés ont été criblés et une activité endo-xylanase, bien que faible par rapport à des xylanases de la famille des GH10 (Alvarez, Goldbeck, dos Santos, et al. 2013 ; Gong et al. 2013 ; Ko et al. 2016), a été obtenue. GH* a donc une activité endo-xylanase avec une préférence pour les arabinoxylanes et libère des xylooligosaccharides ayant un degré de polymérisation supérieur ou égal à 2. L'activité de l'enzyme a aussi été testée sur xylooligosaccharides et est la meilleure observée à ce jour, notamment sur xylooligosaccharides de DP supérieur à 4, en notant toutefois que des oligosaccharides de DP supérieur à 6 n'ont pas pu être testés car non commercialisés. Malgré la faible activité, les paramètres optimaux de l'enzyme ont pu être établis. L'obtention des activités sur xylooligosaccharides a permis d'établir qu'il faut au minimum que les sous-sites catalytiques -2 à +2 soient occupés pour que l'enzyme soit active et que, pour avoir la meilleure activité, il faut que les sous-sites -3 à +3 soient occupés.

D'autre part, des essais de cristallisation de l'enzyme ont été conduits et ont permis l'obtention de cristaux diffractant à 1,6 Å de résolution. A partir de ces cristaux un jeu de données a pu être collecté. La structure 3D de GH* a ainsi été résolue par remplacement moléculaire avec la GH5 FxCel5A comme modèle.

L'analyse de la structure a permis d'identifier le site catalytique hypothétique de GH* ainsi que les résidus d'acide aminé potentiellement impliqués dans le maintien et l'hydrolyse du substrat. Deux résidus ont été identifiés comme pouvant être les résidus catalytiques : l'acide glutamique 117 et la tyrosine 116. Les groupements fonctionnels de ces résidus sont espacés de 5,5 Å, ce qui va dans le sens d'un mécanisme avec rétention de la configuration, et est en accord avec le fait que GH* avait été annotée comme appartenant au clan GH-A.

Des mutants de GH* pour chacun des résidus catalytiques supposés ont été obtenus et montrent une perte de plus de 97 % de l'activité de l'enzyme. Cela prouve que ces résidus jouent un rôle crucial dans l'activité de GH* mais pas qu'ils sont, de façon certaine, les acides aminés catalytiques. Une tyrosine jouant le rôle du résidu catalytique nucléophile existe dans les familles de sialidases GH33 et 34, mais est activée par un acide glutamique (Newstead et al. 2008 ; Park et al. 2013). Par contre, dans les GH33 et 34 les résidus catalytiques se font face comme dans les mécanismes classiques de rétention de la

configuration, ce qui n'est pas le cas chez GH*. De plus, les GH33 et 34 ont un repliement en hélice à 6 pales différent de celui de GH*. Cependant, dans la structure de GH* l'aspartate 12 pourrait jouer ce rôle d'activateur du résidu nucléophile par l'intermédiaire d'une molécule d'eau. Pour vérifier si ce D12 est impliqué dans le mécanisme de GH*, un mutant D12A peut être envisagé.

Pour aller plus loin dans ce travail, il est maintenant nécessaire de déterminer les paramètres cinétiques de GH*. Il peut être intéressant aussi de générer d'autres mutants pour voir l'implication d'autres acides aminés qui ont été identifiés comme pouvant jouer un rôle dans le mécanisme de GH*.

D'autre part, des cristaux des mutants de GH* en complexe avec un xylohexaose ont été obtenus mais n'ont pas suffisamment diffracté pour résoudre la structure du complexe. Il reste donc à résoudre la structure de complexes entre la protéine sauvage ou des mutants inactifs et différents ligands xylooligosaccharidiques afin de vérifier l'emplacement du site catalytique et si les hypothèses formulées par rapport aux résidus impliqués se vérifient.

Enfin, il faudrait prouver que le mécanisme de GH* est bien un mécanisme avec rétention de la configuration (Wilson et al. 2014). Pour cela, des expériences de RMN vont être menées rapidement.

Enfin, on sait qu'il existe d'autres GH notées étoile (n'appartenant à aucunes familles de GH connues) dans la base de données CAZy. Une collaboration avec l'équipe de Bernard Henrissat peut être envisagée afin de chercher si d'autres GH ont une activité et un mécanisme similaires à ceux de GH* et donc si une nouvelle famille de GH peut être créée.

Bibliographie :

- Aachmann, Finn L., Morten Sørli, Gudmund Skjåk-Bræk, Vincent G. H. Eijsink, et Gustav Vaaje-Kolstad. 2012. « NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions ». *Proceedings of the National Academy of Sciences of the United States of America* 109 (46): 18779- 84. doi:10.1073/pnas.1208822109.
- Abbott, D. Wade, et Alisdair B. Boraston. 2007. « The Structural Basis for Exopolygalacturonase Activity in a Family 28 Glycoside Hydrolase ». *Journal of Molecular Biology* 368 (5): 1215- 22. doi:10.1016/j.jmb.2007.02.083.
- Abe, Y., David Edward Bignell, et T. Higashi. 2014. *Termites: Evolution, Sociality, Symbioses, Ecology*. Springer.
- Adelsberger, Helmuth, Christian Hertel, Erich Glawischnig, Vladimir V. Zverlov, et Wolfgang H. Schwarz. 2004. « Enzyme system of *Clostridium stercoarium* for hydrolysis of arabinoxylan: reconstitution of the in vivo system from recombinant enzymes ». *Microbiology* 150 (7): 2257- 66. doi:10.1099/mic.0.27066-0.
- Aist, J. R. 1976. « Papillae and Related Wound Plugs of Plant Cells ». *Annual Review of Phytopathology* 14 (1): 145 - 63. doi:10.1146/annurev.py.14.090176.001045.
- Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, et David J. Lipman. 1990. « Basic local alignment search tool ». *Journal of Molecular Biology* 215 (3): 403- 10. doi:10.1016/S0022-2836(05)80360-2.
- Alvarez, Thabata M., Rosana Goldbeck, Camila Ramos dos Santos, Douglas A. A. Paixão, Thiago A. Gonçalves, João Paulo L. Franco Cairo, Rodrigo Ferreira Almeida, et al. 2013. « Development and Biotechnological Application of a Novel Endoxylanase Family GH10 Identified from Sugarcane Soil Metagenome ». *PLOS ONE* 8 (7): e70014. doi:10.1371/journal.pone.0070014.
- Alvarez, Thabata M., Rosana Goldbeck, Camila Ramos dos Santos, Douglas A. A. Paixão, Thiago A. Gonçalves, João Paulo L. Franco Cairo, Rodrigo Ferreira Almeida, et al. 2013. « Development and Biotechnological Application of a Novel Endoxylanase Family GH10 Identified from Sugarcane Soil Metagenome ». *PLoS ONE* 8 (7). doi:10.1371/journal.pone.0070014.
- Amaya, María Fernanda, Andrew G. Watts, Iben Damager, Annemarie Wehenkel, Tong Nguyen, Alejandro Buschiazzi, Gastón Paris, A. Carlos Frasch, Stephen G. Withers, et Pedro M. Alzari. 2004. « Structural Insights into the Catalytic Mechanism of *Trypanosoma Cruzi* Trans-Sialidase ». *Structure* 12 (5): 775- 84. doi:10.1016/j.str.2004.02.036.
- Arantes, Valdeir, et Adriane Maria Ferreira Milagres. 2006. « Degradation of Cellulosic and Hemicellulosic Substrates Using a Chelator-Mediated Fenton Reaction ». *Journal of Chemical Technology & Biotechnology* 81 (3): 413- 19. doi:10.1002/jctb.1417.

- Arnal, Gregory. 2014. « Discovery and characterization of biomass-degrading enzymes and enzyme systems in termite gut microbial ecosystems. » Toulouse, INSA. <http://www.theses.fr/2014ISAT0040>.
- Aspeborg, Henrik, Pedro M Coutinho, Yang Wang, Harry Brumer, et Bernard Henrissat. 2012. « Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5) ». *BMC Evolutionary Biology* 12 (septembre): 186. doi:10.1186/1471-2148-12-186.
- Barth, Heidi, Christiane Schäfer, Mohammed I. Adah, Fuming Zhang, Robert J. Linhardt, Hidenao Toyoda, Akiko Kinoshita-Toyoda, et al. 2003. « Cellular Binding of Hepatitis C Virus Envelope Glycoprotein E2 Requires Cell Surface Heparan Sulfate ». *Journal of Biological Chemistry* 278 (42): 41003- 12. doi:10.1074/jbc.M302267200.
- Bashline, Logan, Lei Lei, Shundai Li, et Ying Gu. 2014. « Cell Wall, Cytoskeleton, and Cell Expansion in Higher Plants ». *Molecular Plant* 7 (4): 586- 600. doi:10.1093/mp/ssu018.
- Bastien, Géraldine, Grégory Arnal, Sophie Bozonnet, Sandrine Laguerre, Fernando Ferreira, Régis Fauré, Bernard Henrissat, et al. 2013. « Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics ». *Biotechnology for Biofuels* 6 (mai): 78. doi:10.1186/1754-6834-6-78.
- Bastien-Uluis, Geraldine. 2012. « Découverte de nouvelles enzymes de dégradation des polysaccharides végétaux par métagénomique fonctionnelle ». Toulouse, INSA. <http://www.theses.fr/2012ISAT0005>.
- Bayer, Edward A., Jean-Pierre Belaich, Yuval Shoham, et Raphael Lamed. 2004. « The Cellulosomes: Multienzyme Machines for Degradation of Plant Cell Wall Polysaccharides ». *Annual Review of Microbiology* 58 (1): 521- 54. doi:10.1146/annurev.micro.57.030502.091022.
- Bayer, Edward A., Linda J. W. Shimon, Yuval Shoham, et Raphael Lamed. 1998. « Cellulosomes—Structure and Ultrastructure ». *Journal of Structural Biology* 124 (2): 221- 34. doi:10.1006/jsbi.1998.4065.
- Béguin, Pierre, et Marc Lemaire. 1996. « The Cellulosome: An Exocellular, Multiprotein Complex Specialized in Cellulose Degradation ». *Critical Reviews in Biochemistry and Molecular Biology* 31 (3): 201- 36. doi:10.3109/10409239609106584.
- Bekele, Aschalew Z., Satoshi Koike, et Yasuo Kobayashi. 2011. « Phylogenetic Diversity and Dietary Association of Rumen *Treponema* Revealed Using Group-Specific 16S rRNA Gene-Based Analysis ». *FEMS Microbiology Letters* 316 (1): 51- 60. doi:10.1111/j.1574-6968.2010.02191.x.
- Biely, Peter. 2012. « Microbial carbohydrate esterases deacetylating plant polysaccharides ». *Biotechnology Advances*, Special issue on ACB 2011, 30 (6): 1575- 88. doi:10.1016/j.biotechadv.2012.04.010.

- Bisgaard-Frantzen, Henrik, Allan Svendsen, Barrie Norman, Sven Pedersen, Søren Kjaerulff, Helle Outtrup, et Torben V. Borchert. 1999. « Development of Industrially Important α -Amylases ». *Journal of Applied Glycoscience* 46 (2): 199- 206. doi:10.5458/jag.46.199.
- Bjursell, Magnus K., Eric C. Martens, et Jeffrey I. Gordon. 2006. « Functional Genomic and Metabolic Studies of the Adaptations of a Prominent Adult Human Gut Symbiont, *Bacteroides thetaiotaomicron*, to the Suckling Period ». <http://www.jbc.org.gate1.inist.fr/content/281/47/36269.long>.
- Boerjan, Wout, John Ralph, et Marie Baucher. 2003. « Lignin Biosynthesis ». *Annual Review of Plant Biology* 54 (1): 519- 46. doi:10.1146/annurev.arplant.54.031902.134938.
- Bolam, David N, et Nicole M Koropatkin. 2012. « Glycan recognition by the Bacteroidetes Sus-like systems ». *Current Opinion in Structural Biology, Carbohydrates and glycoconjugates ? Biophysical methods*, 22 (5): 563- 69. doi:10.1016/j.sbi.2012.06.006.
- Boraston, Alisdair B., David N. Bolam, Harry J. Gilbert, et Gideon J. Davies. 2004. « Carbohydrate-binding modules: fine-tuning polysaccharide recognition ». *Biochemical Journal* 382 (Pt 3): 769- 81. doi:10.1042/BJ20040892.
- Bornscheuer, Uwe, Klaus Buchholz, et Jürgen Seibel. 2014. « Enzymatic Degradation of (Ligno)cellulose ». *Angewandte Chemie International Edition* 53 (41): 10876- 93. doi:10.1002/anie.201309953.
- Bos, Martine P., Viviane Robert, et Jan Tommassen. 2007. « Biogenesis of the Gram-Negative Bacterial Outer Membrane ». *Annual Review of Microbiology* 61 (1): 191- 214. doi:10.1146/annurev.micro.61.080706.093245.
- Brennan, YaLi, Walter N. Callen, Leif Christoffersen, Paul Dupree, Florence Goubet, Shaun Healey, Myrian Hernández, et al. 2004. « Unusual Microbial Xylanases from Insect Guts ». *Applied and Environmental Microbiology* 70 (6): 3609- 17. doi:10.1128/AEM.70.6.3609-3617.2004.
- Brune, Andreas. 1998. « Termite guts: the world's smallest bioreactors ». *Trends in Biotechnology* 16 (1): 16- 21. doi:10.1016/S0167-7799(97)01151-7.
- . 2014. « Symbiotic Digestion of Lignocellulose in Termite Guts ». *Nature Reviews Microbiology* 12 (3): 168- 80. doi:10.1038/nrmicro3182.
- Bujacz, Anna, Marzena Jedrzejczak-Krzepkowska, Stanislaw Bielecki, Izabela Redzynia, et Grzegorz Bujacz. 2011. « Crystal Structures of the Apo Form of β -Fructofuranosidase from *Bifidobacterium Longum* and Its Complex with Fructose ». *FEBS Journal* 278 (10): 1728- 44. doi:10.1111/j.1742-4658.2011.08098.x.

- Burton, Rachel A., Michael J. Gidley, et Geoffrey B. Fincher. 2010. « Heterogeneity in the Chemistry, Structure and Function of Plant Cell Walls ». *Nature Chemical Biology* 6 (10): 724- 32. doi:10.1038/nchembio.439.
- Cantarel, Brandi L., Pedro M. Coutinho, Corinne Rancurel, Thomas Bernard, Vincent Lombard, et Bernard Henrissat. 2009. « The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics ». *Nucleic Acids Research* 37 (Database issue): D233- 38. doi:10.1093/nar/gkn663.
- Capounova, Dana, et Milan Drdak. 2002. « Comparison of some Commercial Pectic Enzyme Preparations Applicable in Wine Technology ». *Czech Journal of Food Science* 20 (4): 131- 34.
- Cartmell, Alan, Lauren S. McKee, Maria J. Peña, Johan Larsbrink, Harry Brumer, Satoshi Kaneko, Hitomi Ichinose, et al. 2011a. « The Structure and Function of an Arabinan-Specific α -1,2-Arabinofuranosidase Identified from Screening the Activities of Bacterial GH43 Glycoside Hydrolases ». *Journal of Biological Chemistry* 286 (17): 15483- 95. doi:10.1074/jbc.M110.215962.
- . 2011b. « The Structure and Function of an Arabinan-specific α -1,2-Arabinofuranosidase Identified from Screening the Activities of Bacterial GH43 Glycoside Hydrolases ». *The Journal of Biological Chemistry* 286 (17): 15483- 95. doi:10.1074/jbc.M110.215962.
- « CAZy - Home ». 2016. Consulté le mai 2. <http://www.cazy.org/>.
- « CAZypedia ». 2016. Consulté le juin 1. https://www.cazypedia.org/index.php/Main_Page.
- Chan, Jordi, Magdalena Eder, Elizabeth Faris Crowell, Janet Hampson, Grant Calder, et Clive Lloyd. 2011. « Microtubules and CESA Tracks at the Inner Epidermal Wall Align Independently of Those on the Outer Wall of Light-Grown Arabidopsis Hypocotyls ». *J Cell Sci* 124 (7): 1088- 94. doi:10.1242/jcs.086702.
- Charnock, Simon J, et Barry V McCleary. 2005. « Les enzymes : applications industrielles et analytiques. » https://www.megazyme.com/docs/default-source/analytical-applications-downloads/enzymes_indsutrial_and_analytical_appliation_fra.pdf?sfvrsn=4.
- Chávez, Renato, Paulina Bull, et Jaime Eyzaguirre. 2006. « The xylanolytic enzyme system from the genus *Penicillium* ». *Journal of Biotechnology* 123 (4): 413- 33. doi:10.1016/j.jbiotec.2005.12.036.
- Cherubini, Francesco. 2010. « The biorefinery concept: Using biomass instead of oil for producing energy and chemicals ». *Energy Conversion and Management* 51 (7): 1412- 21. doi:10.1016/j.enconman.2010.01.015.
- Chundawat, Shishir P. S., Gregg T. Beckham, Michael E. Himmel, et Bruce E. Dale. 2011. « Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals ». *Annual Review*

of Chemical and Biomolecular Engineering 2 (1): 121- 45. doi:10.1146/annurev-chembioeng-061010-114205.

- Claesson, Marcus J., Orla O'Sullivan, Qiong Wang, Janne Nikkilä, Julian R. Marchesi, Hauke Smidt, Willem M. de Vos, R. Paul Ross, et Paul W. O'Toole. 2009. « Comparative Analysis of Pyrosequencing and a Phylogenetic Microarray for Exploring Microbial Community Structures in the Human Distal Intestine ». *PLOS ONE* 4 (8): e6669. doi:10.1371/journal.pone.0006669.
- Coen, Judith A., et B. A. Dehority. 1970. « Degradation and Utilization of Hemicellulose from Intact Forages by Pure Cultures of Rumen Bacteria 1 ». *Applied Microbiology* 20 (3): 362- 68.
- Cosgrove, Daniel J. 2005. « Growth of the Plant Cell Wall ». *Nature Reviews Molecular Cell Biology* 6 (11): 850- 61. doi:10.1038/nrm1746.
- Coutinho, Pedro M., Emeline Deleury, Gideon J. Davies, et Bernard Henrissat. 2003. « An Evolving Hierarchical Family Classification for Glycosyltransferases ». *Journal of Molecular Biology* 328 (2): 307- 17. doi:10.1016/S0022-2836(03)00307-3.
- Couturier, Marie, Alain Roussel, Anna Rosengren, Philippe Leone, Henrik Stålbrand, et Jean-Guy Berrin. 2013. « Structural and Biochemical Analyses of Glycoside Hydrolase Families 5 and 26 β -(1,4)-Mannanases from *Podospora Anserina* Reveal Differences upon Manno-Oligosaccharide Catalysis ». *Journal of Biological Chemistry* 288 (20): 14624- 35. doi:10.1074/jbc.M113.459438.
- Couturier, Marie, Narumon Tangthirasunun, Xie Ning, Sylvain Brun, Valérie Gautier, Chloé Bennati-Granier, Philippe Silar, et Jean-Guy Berrin. 2016. « Plant biomass degrading ability of the coprophilic ascomycete fungus *Podospora anserina* ». *Biotechnology Advances* 34 (5): 976- 83. doi:10.1016/j.biotechadv.2016.05.010.
- Cragg, Simon M, Gregg T Beckham, Neil C Bruce, Timothy DH Bugg, Daniel L Distel, Paul Dupree, Amaia Green Etxabe, et al. 2015. « Lignocellulose degradation mechanisms across the Tree of Life ». *Current Opinion in Chemical Biology, Energy • Mechanistic biology*, 29 (décembre): 108- 19. doi:10.1016/j.cbpa.2015.10.018.
- Crowell, Elizabeth Faris, Hélène Timpano, Thierry Desprez, Tiny Franssen-Verheijen, Anne-Mie Emons, Herman Höfte, et Samantha Vernhettes. 2011. « Differential Regulation of Cellulose Orientation at the Inner and Outer Face of Epidermal Cells in the *Arabidopsis Hypocotyl* ». *The Plant Cell* 23 (7): 2592 - 2605. doi:10.1105/tpc.111.087338.
- Cuskin, Fiona, Arnaud Baslé, Simon Ladevèze, Alison M. Day, Harry J. Gilbert, Gideon J. Davies, Gabrielle Potocki-Véronèse, et Elisabeth C. Lowe. 2015. « The GH130 Family of Mannoside Phosphorylases Contains Glycoside Hydrolases That Target β -1,2-Mannosidic Linkages in *Candida Mannan* ». *The Journal of Biological Chemistry* 290 (41): 25023- 33. doi:10.1074/jbc.M115.681460.

- Cuskin, Fiona, James E. Flint, Tracey M. Gloster, Carl Morland, Arnaud Baslé, Bernard Henrissat, Pedro M. Coutinho, et al. 2012. « How Nature Can Exploit Nonspecific Catalytic and Carbohydrate Binding Modules to Create Enzymatic Specificity ». *Proceedings of the National Academy of Sciences* 109 (51): 20889- 94. doi:10.1073/pnas.1212034109.
- Czjzek, Mirjam, Alon Ben David, Tsafirir Bravman, Gil Shoham, Bernard Henrissat, et Yuval Shoham. 2005. « Enzyme–Substrate Complex Structures of a GH39 β -Xylosidase from *Geobacillus stearothermophilus* ». *Journal of Molecular Biology* 353 (4): 838- 46. doi:10.1016/j.jmb.2005.09.003.
- Dashtban, Mehdi, Heidi Schraft, Tarannum A. Syed, et Wensheng Qin. 2010. « Fungal biodegradation and enzymatic modification of lignin ». *International Journal of Biochemistry and Molecular Biology* 1 (1): 36- 50.
- Davies, David R., Michael K. Theodorou, Michelle I. G. Lawrence, et Anthony P. J. Trinci. 1993. « Distribution of anaerobic fungi in the digestive tract of cattle and their survival in faeces ». *Microbiology* 139 (6): 1395- 1400. doi:10.1099/00221287-139-6-1395.
- Davies, Gideon, et Bernard Henrissat. 1995. « Structures and Mechanisms of Glycosyl Hydrolases ». *Structure* 3 (9): 853- 59. doi:10.1016/S0969-2126(01)00220-9.
- D’Elia, J. N., et A. A. Salyers. 1996. « Effect of Regulatory Protein Levels on Utilization of Starch by *Bacteroides Thetaiotaomicron*. » *Journal of Bacteriology* 178 (24): 7180- 86.
- Dilokpimol, Adiphol, Hiroyuki Nakai, Charlotte H. Gotfredsen, Martin J. Baumann, Natsuko Nakai, Maher Abou Hachem, et Birte Svensson. 2011a. « Recombinant production and characterisation of two related GH5 endo- β -1,4-mannanases from *Aspergillus nidulans* FGSC A4 showing distinctly different transglycosylation capacity ». *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1814 (12): 1720- 29. doi:10.1016/j.bbapap.2011.08.003.
- . 2011b. « Recombinant production and characterisation of two related GH5 endo- β -1,4-mannanases from *Aspergillus nidulans* FGSC A4 showing distinctly different transglycosylation capacity ». *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1814 (12): 1720- 29. doi:10.1016/j.bbapap.2011.08.003.
- Divne, C, J Stahlberg, T Reinikainen, L Ruohonen, G Pettersson, J. Knowles, T. Teeri, et T. Jones. 1994. « The Three-Dimensional Crystal Structure of the Catalytic Core of Cellobiohydrolase I from *Trichoderma Reesei* ». *Science* 265 (5171): 524- 28. doi:10.1126/science.8036495.
- Donovan, S. E., P. Eggleton, et D. E. Bignell. 2001. « Gut Content Analysis and a New Feeding Group Classification of Termites ». *Ecological Entomology* 26 (4): 356- 66. doi:10.1046/j.1365-2311.2001.00342.x.

- Driouich, A., L. Faye, et L. A. Staehelin. 1993. « The Plant Golgi Apparatus: A Factory for Complex Polysaccharides and Glycoproteins ». *Trends in Biochemical Sciences* 18 (6): 210- 14.
- Driouich, Azeddine, Marie-Laure Follet-Gueye, Sophie Bernard, Sumaira Kousar, Laurence Chevalier, Maïté Vicré-Gibouin, et Olivier Lerouxel. 2012. « Golgi-Mediated Synthesis and Secretion of Matrix Polysaccharides of the Primary Cell Wall of Higher Plants ». *Frontiers in plant science* 3 (avril). doi:10.3389/fpls.2012.00079.
- Dumon, Claire, Letian Song, Sophie Bozonnet, Régis Fauré, et Michael J. O'Donohue. 2012. « Progress and future prospects for pentose-specific biocatalysts in biorefining ». *Process Biochemistry*, Biomass-derived pentoses, 47 (3): 346- 57. doi:10.1016/j.procbio.2011.06.017.
- Eckburg, Paul B., Elisabeth M. Bik, Charles N. Bernstein, Elizabeth Purdom, Les Dethlefsen, Michael Sargent, Steven R. Gill, Karen E. Nelson, et David A. Relman. 2005. « Diversity of the Human Intestinal Microbial Flora ». *Science (New York, N.Y.)* 308 (5728): 1635- 38. doi:10.1126/science.1110591.
- Emsley, Paul, et Kevin Cowtan. 2004. « Coot: model-building tools for molecular graphics ». <http://journals.iucr.org/d/issues/2004/12/01/ba5070/ba5070.pdf>.
- Engelbrektson, Anna, Victor Kunin, Kelly C. Wrighton, Natasha Zvenigorodsky, Feng Chen, Howard Ochman, et Philip Hugenholtz. 2010. « Experimental Factors Affecting PCR-Based Estimates of Microbial Species Richness and Evenness ». *The ISME Journal* 4 (5): 642- 47. doi:10.1038/ismej.2009.153.
- Ericsson, Ulrika B., B. Martin Hallberg, George T. DeTitta, Niek Dekker, et Pär Nordlund. 2006. « Thermofluor-based high-throughput stability optimization of proteins for structural studies ». *Analytical Biochemistry* 357 (2): 289- 98. doi:10.1016/j.ab.2006.07.027.
- Fauré, Régis, Christophe M. Courtin, Jan A. Delcour, Claire Dumon, Craig B. Faulds, Geoffrey B. Fincher, Sébastien Fort, et al. 2009. « A Brief and Informationally Rich Naming System for Oligosaccharide Motifs of Heteroxylans Found in Plant Cell Walls ». *Australian Journal of Chemistry* 62 (6): 533. doi:10.1071/CH08458.
- Finlay, B. B., et S. Falkow. 1997. « Common Themes in Microbial Pathogenicity Revisited. » *Microbiology and Molecular Biology Reviews* 61 (2): 136- 69.
- Flint, Harry J., Edward A. Bayer, Marco T. Rincon, Raphael Lamed, et Bryan A. White. 2008. « Polysaccharide Utilization by Gut Bacteria: Potential for New Insights from Genomic Analysis ». *Nature Reviews Microbiology* 6 (2): 121- 31. doi:10.1038/nrmicro1817.
- Forsberg, Zarah, Cassandra E. Nelson, Bjørn Dalhus, Sophanit Mekasha, Jennifer S. M. Loose, Lucy I. Crouch, Åsmund K. Røhr, Jeffrey G. Gardner, Vincent G. H. Eijsink, et Gustav Vaaje-Kolstad. 2016. « Structural and Functional Analysis of a Lytic Polysaccharide

- Monooxygenase Important for Efficient Utilization of Chitin in *Cellvibrio Japonicus* ». *Journal of Biological Chemistry* 291 (14): 7300- 7312. doi:10.1074/jbc.M115.700161.
- Forsberg, Zarah, Gustav Vaaje-Kolstad, Bjørge Westereng, Anne C Bunæs, Yngve Stenstrøm, Alasdair MacKenzie, Morten Sørli, Svein J Horn, et Vincent GH Eijsink. 2011. « Cleavage of cellulose by a CBM33 protein ». *Protein Science : A Publication of the Protein Society* 20 (9): 1479- 83. doi:10.1002/pro.689.
- Frandsen, Kristian E. H., Thomas J. Simmons, Paul Dupree, Jens-Christian N. Poulsen, Glyn R. Hemsworth, Luisa Ciano, Esther M. Johnston, et al. 2016. « The Molecular Basis of Polysaccharide Cleavage by Lytic Polysaccharide Monooxygenases ». *Nature Chemical Biology* 12 (4): 298- 303. doi:10.1038/nchembio.2029.
- Fujii, Katsuhiko, Kana Ikeda, et Seo Yoshida. 2012. « Isolation and characterization of aerobic microorganisms with cellulolytic activity in the gut of endogeic earthworms ». <http://revistes.iec.cat/index.php/IM/article/viewFile/64745/20.1501.01.165>.
- Fushinobu, Shinya, Masafumi Hidaka, Yuji Honda, Takayoshi Wakagi, Hirofumi Shoun, et Motomitsu Kitaoka. 2005. « Structural Basis for the Specificity of the Reducing End Xylose-Releasing Exo-Oligoxylanase from *Bacillus Halodurans* C-125 ». *Journal of Biological Chemistry* 280 (17): 17180- 86. doi:10.1074/jbc.M413693200.
- Gibson, Robert P., Tracey M. Gloster, Shirley Roberts, R. Anthony J. Warren, Isabel Storch de Gracia, Ángela García, Jose L. Chiara, et Gideon J. Davies. 2007. « Molecular Basis for Trehalase Inhibition Revealed by the Structure of Trehalase in Complex with Potent Inhibitors ». *Angewandte Chemie International Edition* 46 (22): 4115- 19. doi:10.1002/anie.200604825.
- Golan, Gali, Dalia Shallom, Anna Teplitsky, Galia Zaide, Smadar Shulami, Timor Baasov, Vivian Stojanoff, Andy Thompson, Yuval Shoham, et Gil Shoham. 2004. « Crystal Structures of *Geobacillus Stearothermophilus* α -Glucuronidase Complexed with Its Substrate and Products MECHANISTIC IMPLICATIONS ». *Journal of Biological Chemistry* 279 (4): 3014- 24. doi:10.1074/jbc.M310098200.
- Gong, X., R. J. Gruninger, R. J. Forster, R. M. Teather, et T. A. McAllister. 2013. « Biochemical Analysis of a Highly Specific, pH Stable Xylanase Gene Identified from a Bovine Rumen-Derived Metagenomic Library ». *Applied Microbiology and Biotechnology* 97 (6): 2423- 31. doi:10.1007/s00253-012-4088-y.
- Gorshkova, T. A., L. V. Kozlova, et P. V. Mikshina. 2013. « Spatial Structure of Plant Cell Wall Polysaccharides and Its Functional Significance ». *Biochemistry (Moscow)* 78 (7): 836- 53. doi:10.1134/S0006297913070146.
- Gruber, Karl, Gerd Klintschar, Marianne Hay, Anton Schlacher, Walter Steiner, et Christoph Kratky. 1998. « Thermophilic Xylanase from *Thermomyces lanuginosus* : High-Resolution X-ray Structure and Modeling Studie ». *Biochemistry* 37 (39): 13475- 85. doi:10.1021/bi980864l.

- Gruninger, Robert J., Anil K. Puniya, Tony M. Callaghan, Joan E. Edwards, Noha Youssef, Sumit S. Dagar, Katerina Fliegerova, et al. 2014. « Anaerobic Fungi (Phylum Neocallimastigomycota): Advances in Understanding Their Taxonomy, Life Cycle, Ecology, Role and Biotechnological Potential ». *FEMS Microbiology Ecology* 90 (1): 1- 17. doi:10.1111/1574-6941.12383.
- Guérin, Laurence, Diane-Hélène Sutter, Adeline Demois, Mélanie Chereau, et Gabriel Trandafir. 2009. « Determination of Activity Profiles of the Main Commercial Enzyme Preparations Used in Winemaking ». *American Journal of Enology and Viticulture* 60 (3): 322- 31.
- Guillén, Daniel, Sergio Sánchez, et Romina Rodríguez-Sanoja. 2009. « Carbohydrate-Binding Domains: Multiplicity of Biological Roles ». *Applied Microbiology and Biotechnology* 85 (5): 1241- 49. doi:10.1007/s00253-009-2331-y.
- Hall, Justine R., Kendra R. Mitchell, Olan Jackson-Weaver, Ara S. Kooser, Brandi R. Cron, Laura J. Crossey, et Cristina D. Takacs-Vesbach. 2008. « Molecular Characterization of the Diversity and Distribution of a Thermal Spring Microbial Community by Using rRNA and Metabolic Genes ». *Applied and Environmental Microbiology* 74 (15): 4910- 22. doi:10.1128/AEM.00233-08.
- Hara, Atsushi, Akihiro Imamura, Hiromune Ando, Hideharu Ishida, et Makoto Kiso. 2013. « A New Chemical Approach to Human ABO Histo-Blood Group Type 2 Antigens ». *Molecules* 19 (1): 414- 37. doi:10.3390/molecules19010414.
- He, Shaomei, Natalia Ivanova, Edward Kirton, Martin Allgaier, Claudia Bergin, Rudolf H. Scheffrahn, Nikos C. Kyrpides, Falk Warnecke, Susannah G. Tringe, et Philip Hugenholtz. 2013. « Comparative Metagenomic and Metatranscriptomic Analysis of Hindgut Paunch Microbiota in Wood- and Dung-Feeding Higher Termites ». *PLOS ONE* 8 (4): e61126. doi:10.1371/journal.pone.0061126.
- Henrissat, B, et A Bairoch. 1996. « Updating the sequence-based classification of glycosyl hydrolases. » *Biochemical Journal* 316 (Pt 2): 695- 96.
- Henrissat, Bernard, et Gideon Davies. 1997. « Structural and sequence-based classification of glycoside hydrolases ». *Current Opinion in Structural Biology* 7 (5): 637- 44. doi:10.1016/S0959-440X(97)80072-3.
- Heux, S., I. Meynial-Salles, M. J. O'Donohue, et C. Dumon. 2015. « White biotechnology: State of the art strategies for the development of biocatalysts for biorefining ». *Biotechnology Advances* 33 (8): 1653- 70. doi:10.1016/j.biotechadv.2015.08.004.
- Himmel, Michael E., Shi-You Ding, David K. Johnson, William S. Adney, Mark R. Nimlos, John W. Brady, et Thomas D. Foust. 2007. « Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production ». *Science* 315 (5813): 804- 7. doi:10.1126/science.1137016.
- Höltje, J. V. 1996. « Lysozyme Substrates ». *EXS* 75: 105- 10.

- Hongoh, Yuichi, Pinsurang Deevong, Tetsushi Inoue, Shigeharu Moriya, Savitr Trakulnaleamsai, Moriya Ohkuma, Charunee Vongkaluang, Napavarn Noparatnaraporn, et Toshiaki Kudo. 2005. « Intra- and Interspecific Comparisons of Bacterial Diversity and Community Structure Support Coevolution of Gut Microbiota and Termite Host ». *Applied and Environmental Microbiology* 71 (11): 6590- 99. doi:10.1128/AEM.71.11.6590-6599.2005.
- Hsieh, Yin-Cheng, Yue-Jin Wu, Tzu-Ying Chiang, Chueh-Yuan Kuo, Keshab Lal Shrestha, Cheng-Fu Chao, Yen-Chieh Huang, et al. 2010. « Crystal Structures of Bacillus Cereus NCTU2 Chitinase Complexes with Chitooligomers Reveal Novel Substrate Binding for Catalysis A CHITINASE WITHOUT CHITIN BINDING AND INSERTION DOMAINS ». *Journal of Biological Chemistry* 285 (41): 31603- 15. doi:10.1074/jbc.M110.149310.
- Huson, Daniel H., Alexander F. Auch, Ji Qi, et Stephan C. Schuster. 2007. « MEGAN analysis of metagenomic data ». *Genome Research* 17 (3): 377- 86. doi:10.1101/gr.5969107.
- Ishii, Tadashi. 1997. « Structure and functions of feruloylated polysaccharides ». *Plant Science* 127 (2): 111- 27. doi:10.1016/S0168-9452(97)00130-1.
- Ito, Tasuku, Kyo Saikawa, Seonah Kim, Kiyotaka Fujita, Akihiro Ishiwata, Sophon Kaeothip, Takatoshi Arakawa, et al. 2014. « Crystal structure of glycoside hydrolase family 127 β -l-arabinofuranosidase from Bifidobacterium longum ». *Biochemical and Biophysical Research Communications* 447 (1): 32- 37. doi:10.1016/j.bbrc.2014.03.096.
- Jami, Elie, Adi Israel, Assaf Kotser, et Itzhak Mizrahi. 2013. « Exploring the bovine rumen bacterial community from birth to adulthood ». *The ISME Journal* 7 (6): 1069- 79. doi:10.1038/ismej.2013.2.
- Jayani, Ranveer Singh, Shivalika Saxena, et Reena Gupta. 2005. « Microbial pectinolytic enzymes: A review ». *Process Biochemistry* 40 (9): 2931- 44. doi:10.1016/j.procbio.2005.03.026.
- Jenkins, John, Leila Lo Leggio, Gillian Harris, et Richard Pickersgill. 1995. « β -Glucosidase, β -Galactosidase, Family A Cellulases, Family F Xylanases and Two Barley Glycanases Form a Superfamily of Enzymes With 8-Fold β/α Architecture and with Two Conserved Glutamates near the Carboxy-Terminal Ends of β -Strands Four and Seven ». *FEBS Letters* 362 (3): 281- 85. doi:10.1016/0014-5793(95)00252-5.
- Joshi, Manish D., Gary Sidhu, Isabelle Pot, Gary D. Brayer, Stephen G. Withers, et Lawrence P. McIntosh. 2000. « Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase1 ». *Journal of Molecular Biology* 299 (1): 255- 79. doi:10.1006/jmbi.2000.3722.
- Kabsch, W. 1993. « Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants ». *Journal of Applied Crystallography* 26 (6): 795- 800. doi:10.1107/S0021889893005588.

- Kambhampati, Srinivas, et Paul Eggleton. 2000. « Taxonomy and Phylogeny of Termites ». In *Termites: Evolution, Sociality, Symbioses, Ecology*, édité par Takuya Abe, David Edward Bignell, et Masahiko Higashi, 1- 23. Dordrecht: Springer Netherlands. http://link.springer.com/10.1007/978-94-017-3223-9_1.
- Kamra, D. N. 2005. « Rumen microbial ecosystem ». <http://www.ashdin.com/journals/MG/235571.pdf>.
- Kang, Kang, Shaowen Wang, Guohong Lai, Gang Liu, et Miao Xing. 2013. « Characterization of a novel swollenin from *Penicillium oxalicum* in facilitating enzymatic saccharification of cellulose ». *BMC Biotechnology* 13 (mai): 42. doi:10.1186/1472-6750-13-42.
- Karnaouri, Anthi, Evangelos Topakas, Io Antonopoulou, et Paul Christakopoulos. 2014. « Genomic insights into the fungal lignocellulolytic system of *Myceliophthora thermophila* ». *Frontiers in Microbiology* 5 (juin). doi:10.3389/fmicb.2014.00281.
- Kirk, Ole, Torben Vedel Borchert, et Claus Crone Fuglsang. 2002. « Industrial enzyme applications ». *Current Opinion in Biotechnology* 13 (4): 345- 51. doi:10.1016/S0958-1669(02)00328-2.
- Ko, Ja Kyong, Hyeokjin Ko, Kyoung Heon Kim, et In-Geol Choi. 2016. « Characterization of the Biochemical Properties of Recombinant Xyn10C from a Marine Bacterium, *Saccharophagus Degradans* 2-40 ». *Bioprocess and Biosystems Engineering* 39 (4): 677- 84. doi:10.1007/s00449-016-1548-2.
- Koropatkin, Nicole M., Elizabeth A. Cameron, et Eric C. Martens. 2012. « How Glycan Metabolism Shapes the Human Gut Microbiota ». *Nature Reviews Microbiology* 10 (5): 323- 35. doi:10.1038/nrmicro2746.
- Kozich, James J., Sarah L. Westcott, Nielson T. Baxter, Sarah K. Highlander, et Patrick D. Schloss. 2013. « Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform ». *Applied and Environmental Microbiology* 79 (17): 5112- 20. doi:10.1128/AEM.01043-13.
- Krause, Denis O., Stuart E. Denman, Roderick I. Mackie, Mark Morrison, Ann L. Rae, Graeme T. Attwood, et Christopher S. McSweeney. 2003. « Opportunities to Improve Fiber Degradation in the Rumen: Microbiology, Ecology, and Genomics ». *FEMS Microbiology Reviews* 27 (5): 663- 93. doi:10.1016/S0168-6445(03)00072-X.
- Křišťůfek, Václav, Kinga Ravasz, et Václav Pižl. 1992. « Changes in densities of bacteria and microfungi during gut transit in *Lumbricus rubellus* and *Aporrectodea caliginosa* (Oligochaeta: Lumbricidae) ». *Soil Biology and Biochemistry* 24 (12): 1499- 1500. doi:10.1016/0038-0717(92)90139-O.

- Krissinel, Evgeny, et Kim Henrick. 2007. « Inference of Macromolecular Assemblies from Crystalline State ». *Journal of Molecular Biology* 372 (3): 774- 97. doi:10.1016/j.jmb.2007.05.022.
- Kulkarni, Neeta, Abhay Shendye, et Mala Rao. 1999. « Molecular and Biotechnological Aspects of Xylanases ». *FEMS Microbiology Reviews* 23 (4): 411- 56. doi:10.1111/j.1574-6976.1999.tb00407.x.
- Ladevèze, Simon, Laurence Tarquis, Davide A. Cecchini, Juliette Bercovici, Isabelle André, Christopher M. Topham, Sandrine Morel, et al. 2013. « Role of Glycoside Phosphorylases in Mannose Foraging by Human Gut Bacteria ». *The Journal of Biological Chemistry* 288 (45): 32370- 83. doi:10.1074/jbc.M113.483628.
- Lagaert, Stijn, Annick Pollet, Christophe M. Courtin, et Guido Volckaert. 2014. « β -Xylosidases and α -L-arabinofuranosidases: Accessory enzymes for arabinoxylan degradation ». *Biotechnology Advances* 32 (2): 316- 32. doi:10.1016/j.biotechadv.2013.11.005.
- Lammerts van Bueren, Alicia, Suzie Otani, Esben P. Friis, Keith S. Wilson, et Gideon J. Davies. 2012. « Three-dimensional structure of a thermophilic family GH11 xylanase from *Thermobifida fusca* ». *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 68 (Pt 2): 141- 44. doi:10.1107/S1744309111049608.
- Lavelle, P., I. Barois, I. Cruz, C. Fragoso, A. Hernandez, A. Pineda, et P. Rangel. 1987. « Adaptive Strategies of *Pontoscolex Corethrurus* (Glossoscolecidae, Oligochaeta), a Peregrine Geophagous Earthworm of the Humid Tropics ». *Biology and Fertility of Soils* 5 (3): 188- 94. doi:10.1007/BF00256899.
- Lee, Sun Hwa, et Yong-Eok Lee. 2014. « Cloning, expression, and characterization of a thermostable GH51 α -L-arabinofuranosidase from *Paenibacillus* sp. DG-22. » <http://www.jmb.or.kr/journal/viewJournal.html?doi=10.4014/jmb.1308.08078>.
- Levasseur, Anthony, Elodie Drula, Vincent Lombard, Pedro M Coutinho, et Bernard Henrissat. 2013a. « Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes ». *Biotechnology for Biofuels* 6 (mars): 41. doi:10.1186/1754-6834-6-41.
- . 2013b. « Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes ». *Biotechnology for Biofuels* 6 (mars): 41. doi:10.1186/1754-6834-6-41.
- Ley, Ruth E., Daniel A. Peterson, et Jeffrey I. Gordon. 2006. « Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine ». *Cell* 124 (4): 837- 48. doi:10.1016/j.cell.2006.02.017.
- Li, Luen-Luen, Safiyh Taghavi, Sean M McCorkle, Yian-Biao Zhang, Michael G Blewitt, Roman Brunecky, William S Adney, et al. 2011. « Bioprospecting metagenomics of decaying

- wood: mining for new glycoside hydrolases ». *Biotechnology for Biofuels* 4 (août): 23. doi:10.1186/1754-6834-4-23.
- Li, Xin, William T. Beeson, Christopher M. Phillips, Michael A. Marletta, et Jamie H. D. Cate. 2012. « Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases ». *Structure (London, England : 1993)* 20 (6): 1051- 61. doi:10.1016/j.str.2012.04.002.
- Lombard, Vincent, Thomas Bernard, Corinne Rancurel, Harry Brumer, Pedro M. Coutinho, et Bernard Henrissat. 2010. « A Hierarchical Classification of Polysaccharide Lyases for Glycogenomics ». *Biochemical Journal* 432 (3): 437- 44. doi:10.1042/BJ20101185.
- Lombard, Vincent, Hemalatha Golaconda Ramulu, Elodie Drula, Pedro M. Coutinho, et Bernard Henrissat. 2014. « The carbohydrate-active enzymes database (CAZY) in 2013 ». *Nucleic Acids Research* 42 (Database issue): D490- 95. doi:10.1093/nar/gkt1178.
- Lyu, Qianqian, Song Wang, Wenhua Xu, Baoqin Han, Wanshun Liu, David N. M. Jones, et Weizhi Liu. 2014. « Structural Insights into the Substrate-Binding Mechanism for a Novel Chitosanase ». *Biochemical Journal* 461 (2): 335- 45. doi:10.1042/BJ20140159.
- Mark, Brian L., David J. Vocadlo, Spencer Knapp, Barbara L. Triggs-Raine, Stephen G. Withers, et Michael N. G. James. 2001. « Crystallographic Evidence for Substrate-Assisted Catalysis in a Bacterial β -Hexosaminidase ». *Journal of Biological Chemistry* 276 (13): 10330- 37. doi:10.1074/jbc.M011067200.
- Martens, Eric C., Nicole M. Koropatkin, Thomas J. Smith, et Jeffrey I. Gordon. 2009. « Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm ». *Journal of Biological Chemistry* 284 (37): 24673- 77. doi:10.1074/jbc.R109.022848.
- Martens, Eric C., Elisabeth C. Lowe, Herbert Chiang, Nicholas A. Pudlo, Meng Wu, Nathan P. McNulty, D. Wade Abbott, et al. 2011. « Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts ». *PLoS Biology* 9 (12). doi:10.1371/journal.pbio.1001221.
- Mattéotti, Christel, Eric Haubruge, Philippe Thonart, Frédéric Francis, Edwin De Pauw, Daniel Portetelle, et Micheline Vandenberg. 2011. « Characterization of a New β -Glucosidase/ β -Xylosidase from the Gut Microbiota of the Termite (*Reticulitermes Santonensis*) ». *FEMS Microbiology Letters* 314 (2): 147- 57. doi:10.1111/j.1574-6968.2010.02161.x.
- McCarter, John D., et G. Stephen Withers. 1994. « Mechanisms of enzymatic glycoside hydrolysis ». *Current Opinion in Structural Biology* 4 (6): 885- 92. doi:10.1016/0959-440X(94)90271-2.

- McCleary, Barry V., et Paraic McGeough. 2015. « A Comparison of Polysaccharide Substrates and Reducing Sugar Methods for the Measurement of endo-1,4- β -Xylanase ». *Applied Biochemistry and Biotechnology* 177 (5): 1152- 63. doi:10.1007/s12010-015-1803-z.
- McCoy, Airlie J., Ralf W. Grosse-Kunstleve, Paul D. Adams, Martyn D. Winn, Laurent C. Storoni, et Randy J. Read. 2007. « Phaser Crystallographic Software ». *Journal of Applied Crystallography* 40 (Pt 4): 658- 74. doi:10.1107/S0021889807021206.
- McKee, Lauren S., Maria J. Peña, Artur Rogowski, Adam Jackson, Richard J. Lewis, William S. York, Kristian B. R. M. Krogh, et al. 2012. « Introducing Endo-Xylanase Activity into an Exo-Acting Arabinofuranosidase That Targets Side Chains ». *Proceedings of the National Academy of Sciences* 109 (17): 6537- 42. doi:10.1073/pnas.1117686109.
- Medini, Duccio, Davide Serruto, Julian Parkhill, David A. Relman, Claudio Donati, Richard Moxon, Stanley Falkow, et Rino Rappuoli. 2008. « Microbiology in the Post-Genomic Era ». *Nature Reviews Microbiology* 6 (6): 419- 30. doi:10.1038/nrmicro1901.
- Meereis, Florian, et Michael Kaufmann. 2008. « Extension of the COG and arCOG databases by amino acid and nucleotide sequences ». *BMC Bioinformatics* 9 (novembre): 479. doi:10.1186/1471-2105-9-479.
- Megson, ZA, A Koerdt, H Schuster, R Ludwig, B Janesch, A Frey, K Naylor, et al. 2015. « Characterization of an α -l-fucosidase from the periodontal pathogen *Tannerella forsythia* ». *Virulence* 6 (3): 282- 92. doi:10.1080/21505594.2015.1010982.
- Mewis, Keith, Nicolas Lenfant, Vincent Lombard, et Bernard Henrissat. 2016. « Dividing the Large Glycoside Hydrolase Family 43 into Subfamilies: A Motivation for Detailed Enzyme Characterization ». *Applied and Environmental Microbiology*, janvier, AEM.03453-15. doi:10.1128/AEM.03453-15.
- Mitra, Suparna, Bernhard Klar, et Daniel H. Huson. 2009. « Visual and Statistical Comparison of Metagenomes ». *Bioinformatics* 25 (15): 1849- 55. doi:10.1093/bioinformatics/btp341.
- Mizuno, Masahiro, Takashi Tonozuka, Saori Suzuki, Rie Uotsu-Tomita, Shigehiro Kamitori, Atsushi Nishikawa, et Yoshiyuki Sakano. 2004. « Structural Insights into Substrate Specificity and Function of Glucodextranase ». *Journal of Biological Chemistry* 279 (11): 10575- 83. doi:10.1074/jbc.M310771200.
- Mohnen, Debra. 2008. « Pectin structure and biosynthesis ». *Current Opinion in Plant Biology, Physiology and Metabolism* - Edited by Markus Pauly and Kenneth Keegstra, 11 (3): 266- 77. doi:10.1016/j.pbi.2008.03.006.
- Montanier, Cedric, Alicia Lammerts van Bueren, Claire Dumon, James E. Flint, Marcia A. Correia, Jose A. Prates, Susan J. Firbank, et al. 2009. « Evidence That Family 35 Carbohydrate Binding Modules Display Conserved Specificity but Divergent Function ». *Proceedings of the National Academy of Sciences* 106 (9): 3065- 70. doi:10.1073/pnas.0808972106.

- Murshudov, Garib N., Pavol Skubák, Andrey A. Lebedev, Navraj S. Pannu, Roberto A. Steiner, Robert A. Nicholls, Martyn D. Winn, Fei Long, et Alexei A. Vagin. 2011. « REFMAC5 for the refinement of macromolecular crystal structures ». *Acta Crystallographica Section D: Biological Crystallography* 67 (Pt 4): 355- 67. doi:10.1107/S0907444911001314.
- Muzard, M., N. Aubry, R. Plantier-Royon, M. O'Donohue, et C. Rémond. 2009. « Evaluation of the transglycosylation activities of a GH 39 β -d-xylosidase for the synthesis of xylose-based glycosides ». *Journal of Molecular Catalysis B: Enzymatic* 58 (1-4): 1- 5. doi:10.1016/j.molcatb.2008.10.004.
- Neutelings, Godfrey. 2011. « Lignin variability in plant cell walls: Contribution of new models ». *Plant Science* 181 (4): 379- 86. doi:10.1016/j.plantsci.2011.06.012.
- Newstead, Simon L., Jane A. Potter, Jennifer C. Wilson, Guogang Xu, Chin-Hsiang Chien, Andrew G. Watts, Stephen G. Withers, et Garry L. Taylor. 2008. « The Structure of *Clostridium Perfringens* NanI Sialidase and Its Catalytic Intermediates ». *Journal of Biological Chemistry* 283 (14): 9080- 88. doi:10.1074/jbc.M710247200.
- Ni, Jinfeng, et Gaku Tokuda. 2013. « Lignocellulose-degrading enzymes from termites and their symbiotic microbiota ». *Biotechnology Advances*, « Bioenergy and Biorefinery from Biomass » through innovative technology development, 31 (6): 838- 50. doi:10.1016/j.biotechadv.2013.04.005.
- Niesen, Frank H., Helena Berglund, et Masoud Vedadi. 2007. « The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability ». *Nature Protocols* 2 (9): 2212- 21. doi:10.1038/nprot.2007.321.
- Noguchi, Hideki, Jungho Park, et Toshihisa Takagi. 2006. « MetaGene: prokaryotic gene finding from environmental genome shotgun sequences ». *Nucleic Acids Research* 34 (19): 5623- 30. doi:10.1093/nar/gkl723.
- Nurizzo, Didier, Tibor Nagy, Harry J Gilbert, et Gideon J Davies. 2002. « The Structural Basis for Catalysis and Specificity of the *Pseudomonas cellulosa* α -Glucuronidase, GlcA67A ». *Structure* 10 (4): 547- 56. doi:10.1016/S0969-2126(02)00742-6.
- Ohkuma, M. 2003. « Termite Symbiotic Systems: Efficient Bio-Recycling of Lignocellulose ». *Applied Microbiology and Biotechnology* 61 (1): 1- 9. doi:10.1007/s00253-002-1189-z.
- Okazawa, Yuka, Takatsugu Miyazaki, Gaku Yokoi, Yuichi Ishizaki, Atsushi Nishikawa, et Takashi Tonozuka. 2015. « Crystal Structure and Mutational Analysis of Isomalto-Dextranase, a Member of Glycoside Hydrolase Family 27 ». *Journal of Biological Chemistry* 290 (43): 26339- 49. doi:10.1074/jbc.M115.680942.
- Orpin, C. G. 1975. « Studies on the Rumen Flagellate *Neocallimastix frontalis* ». *Microbiology* 91 (2): 249- 62. doi:10.1099/00221287-91-2-249.

- Osborne, J. M., et B. A. Dehority. 1989. « Synergism in Degradation and Utilization of Intact Forage Cellulose, Hemicellulose, and Pectin by Three Pure Cultures of Ruminant Bacteria ». *Applied and Environmental Microbiology* 55 (9): 2247- 50.
- Paës, Gabriel, Jean-Guy Berrin, et Johnny Beaugrand. 2012. « GH11 xylanases: Structure/function/properties relationships and applications ». *Biotechnology Advances* 30 (3): 564- 92. doi:10.1016/j.biotechadv.2011.10.003.
- Paës, Gabriel, Lars K. Skov, Michael J. O'Donohue, Caroline Rémond, Jette S. Kastrup, Michael Gajhede, et Osman Mirza. 2008. « The Structure of the Complex between a Branched Pentasaccharide and Thermobacillus xylanilyticus GH-51 Arabinofuranosidase Reveals Xylan-Binding Determinants and Induced Fit ». *Biochemistry* 47 (28): 7441- 51. doi:10.1021/bi800424e.
- Paës, Gabriel, Vinh Tran, Masayuki Takahashi, Imen Boukari, et Michael J. O'Donohue. 2007. « New Insights into the Role of the Thumb-like Loop in GH-11 Xylanases ». *Protein Engineering Design and Selection* 20 (1): 15- 23. doi:10.1093/protein/gzl049.
- Paice, M. G., R. Bernier, et L. Jurasek. 1988. « Viscosity-Enhancing Bleaching of Hardwood Kraft Pulp with Xylanase from a Cloned Gene ». *Biotechnology and Bioengineering* 32 (2): 235- 39. doi:10.1002/bit.260320214.
- Paper, Janet M., John S. Scott-Craig, David Cavalier, Ahmed Faik, Richard E. Wiemels, Melissa S. Borrusch, Mareike Bongers, et Jonathan D. Walton. 2012. « α -Fucosidases with Different Substrate Specificities from Two Species of Fusarium ». *Applied Microbiology and Biotechnology* 97 (12): 5371- 80. doi:10.1007/s00253-012-4423-3.
- Park, Jong-Tae, Hyung-Nam Song, Tae-Yang Jung, Myoung-Hee Lee, Sung-Goo Park, Eui-Jeon Woo, et Kwan-Hwa Park. 2013. « A novel domain arrangement in a monomeric cyclodextrin-hydrolyzing enzyme from the hyperthermophile Pyrococcus furiosus ». *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, Special Section: Chemistry and mechanism of phosphatases, diesterases and triesterases, 1834 (1): 380- 86. doi:10.1016/j.bbapap.2012.08.001.
- Parthasarathi, K., Ranganathan L. S., Anandi V., et Zeyer J. 2007. « Diversity of microflora in the gut and casts of tropical composting earthworms reared on different substrates. »
<http://web.a.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&authType=crawler&jrnl=02548704&AN=25233382&h=wZYdMOdp4c%2fXtgchJZftHxnF%2bZX1FlfMXrQBVszLxFKOs7VfEqZr09gxZn%2fso2sbay8rsnYUt5SqHf4GYhGc2A%3d%3d&crl=c&resultNs=AdminWebAuth&resultLocal=ErrCrlNotAuth&crlhashurl=login.aspx%3fdirect%3dtrue%26profile%3dehost%26scope%3dsite%26authType%3dcrawler%26rnl%3d02548704%26AN%3d25233382>.
- Pell, Gavin, Lóránd Szabo, Simon J. Charnock, Hefang Xie, Tracey M. Gloster, Gideon J. Davies, et Harry J. Gilbert. 2004. « Structural and Biochemical Analysis of Cellvibrio Japonicus Xylanase 10C HOW VARIATION IN SUBSTRATE-BINDING CLEFT INFLUENCES

THE CATALYTIC PROFILE OF FAMILY GH-10 XYLANASES ». *Journal of Biological Chemistry* 279 (12): 11777- 88. doi:10.1074/jbc.M311947200.

Pérez, J., J. Muñoz-Dorado, T. de la Rubia, et J. Martínez. 2002. « Biodegradation and Biological Treatments of Cellulose, Hemicellulose and Lignin: An Overview ». *International Microbiology* 5 (2): 53- 63. doi:10.1007/s10123-002-0062-3.

Peters, Christoph W. B., Ulrich Kruse, Renate Pollwein, Karl-Heinz Grzeschik, et Albrecht E. Sippel. 1989. « The Human Lysozyme Gene ». *European Journal of Biochemistry* 182 (3): 507- 16. doi:10.1111/j.1432-1033.1989.tb14857.x.

Pettersen, Roger C. 1984. « The Chemical Composition of Wood ». In .
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.458.4426&rep=rep1&type=pdf>.

Pollet, Annick, Jan A. Delcour, et Christophe M. Courtin. 2010. « Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families ». *Critical Reviews in Biotechnology* 30 (3): 176- 91. doi:10.3109/07388551003645599.

Pollet, Annick, Stijn Lagaert, Elena Eneyskaya, Anna Kulminkaya, Jan A. Delcour, et Christophe M. Courtin. 2010. « Mutagenesis and subsite mapping underpin the importance for substrate specificity of the aglycon subsites of glycoside hydrolase family 11 xylanases ». *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1804 (4): 977- 85. doi:10.1016/j.bbapap.2010.01.009.

Pujol, Anaïs. 2010. « Synthèse et étude de nouveaux chélateurs sélectifs du cuivre(I) pour les maladies de type Wilson ». <https://tel.archives-ouvertes.fr/tel-00598705/document>.

Quinlan, R. Jason, Matt D. Sweeney, Leila Lo Leggio, Harm Otten, Jens-Christian N. Poulsen, Katja Salomon Johansen, Kristian B. R. M. Krogh, et al. 2011. « Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components ». *Proceedings of the National Academy of Sciences of the United States of America* 108 (37): 15079- 84. doi:10.1073/pnas.1105776108.

Quiroz-Castañeda, Rosa Estela, et Jorge Luis Folch-Mallol. 2013. « Hydrolysis of Biomass Mediated by Cellulases for the Production of Sugars ». In *Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization*, 119- 55. InTech. <http://cdn.intechopen.com/pdfs-wm/44414.pdf>.

Ragauskas, Arthur J., Kathleen M. Poll, et Anthony J. Cesternino. 1993. « Effect of xylanase pretreatment procedures for nonchlorine bleaching. » *Enzyme & Microbial Technology* 482. <https://smartech.gatech.edu/bitstream/handle/1853/2037/tps-482.pdf;jsessionid=325A25179256CFBF237455FE1149EFB7.smartech?sequence=1>.

Razzauti, Maria, Maxime Galan, Maria Bernard, Sarah Maman, Christophe Klopp, Nathalie Charbonnel, Muriel Vayssier-Taussat, Marc Eloit, et Jean-François Cosson. 2015. « A Comparison between Transcriptome Sequencing and 16S Metagenomics for

- Detection of Bacterial Pathogens in Wildlife ». *PLoS Neglected Tropical Diseases* 9 (8). doi:10.1371/journal.pntd.0003929.
- Reinhard, Linda, Hubert Mayerhofer, Arie Geerlof, Jochen Mueller-Dieckmann, et Manfred S. Weiss. 2013. « Optimization of Protein Buffer Cocktails Using Thermofluor ». *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications* 69 (Pt 2): 209- 14. doi:10.1107/S1744309112051858.
- Ridley, Brent L, Malcolm A O'Neill, et Debra Mohnen. 2001. « Pectins: structure, biosynthesis, and oligogalacturonide-related signaling ». *Phytochemistry* 57 (6): 929- 67. doi:10.1016/S0031-9422(01)00113-3.
- Rio Segade, Susana, Caterina Pace, Fabrizio Torchio, Simone Giacosa, Vincenzo Gerbi, et Luca Rolle. 2015. « Impact of maceration enzymes on skin softening and relationship with anthocyanin extraction in wine grapes with different anthocyanin profiles - ImpactOfMacerationEnzymes2015.pdf ». <http://www.edsvs.u-bordeaux2.fr/Fictifs15/Oenologie/ImpactOfMacerationEnzymes2015.pdf>.
- Robert, Céline, et Annick Bernalier-Donadille. 2003. « The Cellulolytic Microflora of the Human Colon: Evidence of Microcrystalline Cellulose-Degrading Bacteria in Methane-Excreting Subjects ». *FEMS Microbiology Ecology* 46 (1): 81- 89. doi:10.1016/S0168-6496(03)00207-1.
- Rogowski, Artur, Arnaud Baslé, Cristiane S. Farinas, Alexandra Solovyova, Jennifer C. Mortimer, Paul Dupree, Harry J. Gilbert, et David N. Bolam. 2014. « Evidence That GH115 α -Glucuronidase Activity, Which Is Required to Degrade Plant Biomass, Is Dependent on Conformational Flexibility ». *Journal of Biological Chemistry* 289 (1): 53- 64. doi:10.1074/jbc.M113.525295.
- Roisin, Yves, Alain Dejean, Bruno Corbara, Jérôme Orivel, Mirna Samaniego, et Maurice Leponce. 2006. « Vertical Stratification of the Termite Assemblage in a Neotropical Rainforest ». *Oecologia* 149 (2): 301- 11.
- Rossmassler, Karen, Carsten Dietrich, Claire Thompson, Aram Mikaelyan, James O. Nonoh, Rudolf H. Scheffrahn, David Sillam-Dussès, et Andreas Brune. 2015. « Metagenomic Analysis of the Microbiota in the Highly Compartmented Hindguts of Six Wood- or Soil-Feeding Higher Termites ». *Microbiome* 3: 56. doi:10.1186/s40168-015-0118-1.
- Rye, Carl S., et Stephen G. Withers. 2000. « Glycosidase mechanisms ». *Current Opinion in Chemical Biology* 4: 573- 80.
- Sapag, Amalia, Johan Wouters, Christophe Lambert, Pablo de Ioannes, Jaime Eyzaguirre, et Eric Depiereux. 2002. « The endoxylanases from family 11: computer analysis of protein sequences reveals important structural and phylogenetic relationships ». *Journal of Biotechnology* 95 (2): 109- 31. doi:10.1016/S0168-1656(02)00002-0.
- Scheller, Henrik Vibe, et Peter Ulvskov. 2010. « Hemicelluloses ». *Annual Review of Plant Biology* 61 (1): 263- 89. doi:10.1146/annurev-arplant-042809-112315.

- Schleifer, K. H., et O. Kandler. 1972. « Peptidoglycan Types of Bacterial Cell Walls and Their Taxonomic Implications. » *Bacteriological Reviews* 36 (4): 407- 77.
- Schloss, Patrick D., Sarah L. Westcott, Thomas Ryabin, Justine R. Hall, Martin Hartmann, Emily B. Hollister, Ryan A. Lesniewski, et al. 2009. « Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities ». *Applied and Environmental Microbiology* 75 (23): 7537- 41. doi:10.1128/AEM.01541-09.
- Schoeler, Constantin, Klara H. Malinowska, Rafael C. Bernardi, Lukas F. Milles, Markus A. Jobst, Ellis Durner, Wolfgang Ott, et al. 2014. « Ultrastable cellulosome-adhesion complex tightens under load ». *Nature Communications* 5 (décembre). doi:10.1038/ncomms6635.
- Shallom, Dalia, et Yuval Shoham. 2003. « Microbial Hemicellulases ». *Current Opinion in Microbiology* 6 (3): 219- 28. doi:10.1016/S1369-5274(03)00056-0.
- Sieiro, Carmen, Belen Garcia-Fraga, Jacobo Lopez-Seijas, Abigail F. Da Silva, et Tomas G. Villa. 2012. « Microbial_pectic_enzymes_in_the_food_and_wine_industry.pdf ». In *Food Industrial Processes – Methods and Equipment*, 201- 18. InTech. http://cdn.intechopen.com/pdfs/29160/InTech-Microbial_pectic_enzymes_in_the_food_and_wine_industry.pdf.
- Smaali, Issam, Caroline Rémond, et Michael J. O'Donohue. 2006. « Expression in Escherichia Coli and Characterization of β -Xylosidases GH39 and GH-43 from Bacillus Halodurans C-125 ». *Applied Microbiology and Biotechnology* 73 (3): 582- 90. doi:10.1007/s00253-006-0512-5.
- Somerville, Chris, Stefan Bauer, Ginger Brininstool, Michelle Facette, Thorsten Hamann, Jennifer Milne, Erin Osborne, et al. 2004. « Toward a Systems Approach to Understanding Plant Cell Walls ». *Science* 306 (5705): 2206- 11. doi:10.1126/science.1102765.
- Song, Letian, Claire Dumon, Béatrice Siguier, Isabelle André, Elena Eneyskaya, Anna Kulminskaya, Sophie Bozonnet, et Michael Joseph O'Donohue. 2014. « Impact of an N-terminal extension on the stability and activity of the GH11 xylanase from Thermobacillus xylanilyticus ». *Journal of Biotechnology* 174 (mars): 64- 72. doi:10.1016/j.jbiotec.2014.01.004.
- Stahl, Martin, Lorna M. Friis, Harald Nothaft, Xin Liu, Jianjun Li, Christine M. Szymanski, et Alain Stintzi. 2011. « L-Fucose Utilization Provides Campylobacter Jejuni with a Competitive Advantage ». *Proceedings of the National Academy of Sciences* 108 (17): 7194- 99. doi:10.1073/pnas.1014125108.
- Subramanian, V., S. Ramachandran, C. Klein, J. R. Wellen, S. Shenoy, W. C. Chapman, et T. Mohanakumar. 2012. « ABO-Incompatible Organ Transplantation ». *International Journal of Immunogenetics* 39 (4): 282- 90. doi:10.1111/j.1744-313X.2012.01101.x.

- Sun, Ye, et Jiayang Cheng. 2002. « Hydrolysis of Lignocellulosic Materials for Ethanol Production: A Review ». *Bioresource Technology* 83 (1): 1- 11.
- Sunna, A., et G. Antranikian. 1997. « Xylanolytic Enzymes from Fungi and Bacteria ». *Critical Reviews in Biotechnology* 17 (1): 39- 67. doi:10.3109/07388559709146606.
- Tajima, Kiyoshi, Roustam I Aminov, Takafumi Nagamine, Koretsugu Ogata, Mutsumi Nakamura, Hiroki Matsui, et Yoshimi Benno. 1999. « Rumen Bacterial Diversity as Determined by Sequence Analysis of 16S rDNA Libraries ». *FEMS Microbiology Ecology* 29 (2): 159- 69. doi:10.1111/j.1574-6941.1999.tb00607.x.
- Tasse, Lena, Juliette Bercovici, Sandra Pizzut-Serin, Patrick Robe, Julien Tap, Christophe Klopp, Brandi L. Cantarel, et al. 2010. « Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes ». *Genome Research* 20 (11): 1605- 12. doi:10.1101/gr.108332.110.
- Tatusov, Roman L., Eugene V. Koonin, et David J. Lipman. 1997. « A Genomic Perspective on Protein Families ». *Science* 278 (5338): 631- 37. doi:10.1126/science.278.5338.631.
- Taylor, Edward J., Nicola L. Smith, Johan P. Turkenburg, Simone D'Souza, Harry J. Gilbert, et Gideon J. Davies. 2006. « Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from *Clostridium thermocellum* ». *Biochemical Journal* 395 (Pt 1): 31- 37. doi:10.1042/BJ20051780.
- Thomas, Torsten, Jack Gilbert, et Folker Meyer. 2012. « Metagenomics - a Guide from Sampling to Data Analysis ». *Microbial Informatics and Experimentation* 2 (1): 3. doi:10.1186/2042-5783-2-3.
- Underwood, William. 2012. « The Plant Cell Wall: A Dynamic Barrier Against Pathogen Invasion ». *Frontiers in plant science* 3 (mai). doi:10.3389/fpls.2012.00085.
- Vaaje-Kolstad, Gustav, Bjørge Westereng, Svein J. Horn, Zhanliang Liu, Hong Zhai, Morten Sørlie, et Vincent G. H. Eijsink. 2010. « An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides ». *Science* 330 (6001): 219- 22. doi:10.1126/science.1192231.
- Valentine, P J, P Arnold, et A A Salyers. 1992. « Cloning and partial characterization of two chromosomal loci from *Bacteroides ovatus* that contain genes essential for growth on guar gum. » *Applied and Environmental Microbiology* 58 (5): 1541- 48.
- Vardakou, Maria, Claire Dumon, James W. Murray, Paul Christakopoulos, David P. Weiner, Nathalie Juge, Richard J. Lewis, Harry J. Gilbert, et James E. Flint. 2008. « Understanding the Structural Basis for Substrate and Inhibitor Recognition in Eukaryotic GH11 Xylanases ». *Journal of Molecular Biology* 375 (5): 1293- 1305. doi:10.1016/j.jmb.2007.11.007.
- Vardakou, Maria, James Flint, Paul Christakopoulos, Richard J. Lewis, Harry J. Gilbert, et James W. Murray. 2005. « A Family 10 *Thermoascus aurantiacus* Xylanase Utilizes

- Arabinose Decorations of Xylan as Significant Substrate Specificity Determinants ». *Journal of Molecular Biology* 352 (5): 1060- 67. doi:10.1016/j.jmb.2005.07.051.
- Varki, Ajit. 1993. « Biological Roles of Oligosaccharides: All of the Theories Are Correct ». *Glycobiology* 3 (2): 97- 130. doi:10.1093/glycob/3.2.97.
- Várnai, Anikó, Miia R. Mäkelä, Demi T. Djajadi, Jenni Rahikainen, Annele Hatakka, et Liisa Viikari. 2014. « Chapter Four - Carbohydrate-Binding Modules of Fungal Cellulases: Occurrence in Nature, Function, and Relevance in Industrial Biomass Conversion ». In *Advances in Applied Microbiology*, édité par Sima Sariaslani and Geoffrey Michael Gadd, 88:103- 65. Academic Press.
<http://www.sciencedirect.com/science/article/pii/B9780128002605000048>.
- Velasquez, Silvia M., Martiniano M. Ricardi, Javier Gloazzo Dorosz, Paula V. Fernandez, Alejandro D. Nadra, Laercio Pol-Fachin, Jack Egelund, et al. 2011. « O-Glycosylated Cell Wall Proteins Are Essential in Root Hair Growth ». *Science* 332 (6036): 1401- 3. doi:10.1126/science.1206657.
- Vollmer, Waldemar, Didier Blanot, et Miguel A. De Pedro. 2008. « Peptidoglycan Structure and Architecture ». *FEMS Microbiology Reviews* 32 (2): 149- 67. doi:10.1111/j.1574-6976.2007.00094.x.
- Voragen, Alphons G. J., Gerd-Jan Coenen, René P. Verhoef, et Henk A. Schols. 2009. « Pectin, a versatile polysaccharide present in plant cell walls ». <http://library.wur.nl/WebQuery/wurpubs/fulltext/10879>.
- Vorwerk, Sonja, Shauna Somerville, et Chris Somerville. 2004. « The role of plant cell wall polysaccharide composition in disease resistance ». *Trends in Plant Science* 9 (4): 203- 9. doi:10.1016/j.tplants.2004.02.005.
- Vuong, Thu V., et David B. Wilson. 2010. « Glycoside Hydrolases: Catalytic Base/Nucleophile Diversity ». *Biotechnology and Bioengineering* 107 (2): 195- 205. doi:10.1002/bit.22838.
- Wagschal, Kurt, Diana Franqui-Espiet, Charles C. Lee, George H. Robertson, et Dominic W. S. Wong. 2005. « Enzyme-Coupled Assay for β -Xylosidase Hydrolysis of Natural Substrates ». *Applied and Environmental Microbiology* 71 (9): 5318- 23. doi:10.1128/AEM.71.9.5318-5323.2005.
- Walker, A. W. 2006. « Influence of substrate and environmental factors on human gut microbial ecology and metabolism. »
- Wang, Weijun, Ruoyu Yan, Boguslaw P. Nocek, Thu V. Vuong, Rosa Di Leo, Xiaohui Xu, Hong Cui, et al. 2016. « Biochemical and Structural Characterization of a Five-Domain GH115 Alpha-Glucuronidase from the Marine Bacterium *Saccharophagus Degradans* 2-40T ». *Journal of Biological Chemistry*, avril, jbc.M115.702944. doi:10.1074/jbc.M115.702944.

- Warnecke, Falk, Peter Luginbühl, Natalia Ivanova, Majid Ghassemian, Toby H. Richardson, Justin T. Stege, Michelle Cayouette, et al. 2007. « Metagenomic and Functional Analysis of Hindgut Microbiota of a Wood-Feeding Higher Termite ». *Nature* 450 (7169): 560- 65. doi:10.1038/nature06269.
- Watson, Jacqueline N., Veedeeta Dookhun, Thor Borgford, et Andrew J. Bennet. 2003. « Mutagenesis of the Conserved Active-Site Tyrosine Changes a Retaining Sialidase into an Inverting Sialidase ». <http://pubs.acs.org.gate1.inist.fr/doi/pdf/10.1021/bi035396g>.
- Watts, Andrew G., Pablo Opezzo, Stephen G. Withers, Pedro M. Alzari, et Alejandro Buschiazzi. 2006. « Structural and Kinetic Analysis of Two Covalent Sialosyl-Enzyme Intermediates on Trypanosoma Rangeli Sialidase ». *Journal of Biological Chemistry* 281 (7): 4149- 55. doi:10.1074/jbc.M510677200.
- Wedekind, K J, H R Mansfield, et L Montgomery. 1988. « Enumeration and isolation of cellulolytic and hemicellulolytic bacteria from human feces. » *Applied and Environmental Microbiology* 54 (6): 1530- 35.
- Westereng, Bjørge, Takuya Ishida, Gustav Vaaje-Kolstad, Miao Wu, Vincent G. H. Eijsink, Kiyohiko Igarashi, Masahiro Samejima, Jerry Ståhlberg, Svein J. Horn, et Mats Sandgren. 2011. « The Putative Endoglucanase PcGH61D from Phanerochaete chrysosporium Is a Metal-Dependent Oxidative Enzyme that Cleaves Cellulose ». *PLoS ONE* 6 (11). doi:10.1371/journal.pone.0027807.
- Whitehurst, Robert J., et Maarten van Oort. 2010. *Enzymes in Food Technology*. 2^e éd. Wiley-Blackwell. http://197.14.51.10:81/pmb/AGROALIMENTAIRE/Enzymes_in_Food_Technology.pdf.
- Willats, William G. T., Lesley McCartney, William Mackie, et J. Paul Knox. 2001. « Pectin: Cell Biology and Prospects for Functional Analysis ». *Plant Molecular Biology* 47 (1- 2): 9- 27. doi:10.1023/A:1010662911148.
- Wilson, Jenifer C, Andrew Elohim Laloo, Sanjesh Singh, et Vito Ferro. 2014. « ¹H NMR spectroscopic studies establish that heparanase is a retaining glycosidase ». *Biochemical and Biophysical Research Communications* 443: 185- 88.
- Wong, Dominic W. S. 2005. « Feruloyl Esterase ». *Applied Biochemistry and Biotechnology* 133 (2): 87- 112. doi:10.1385/ABAB:133:2:87.
- Wu, Miao, Gregg T. Beckham, Anna M. Larsson, Takuya Ishida, Seonah Kim, Christina M. Payne, Michael E. Himmel, et al. 2013. « Crystal Structure and Computational Characterization of the Lytic Polysaccharide Monooxygenase GH61D from the Basidiomycota Fungus Phanerochaete chrysosporium ». *The Journal of Biological Chemistry* 288 (18): 12828- 39. doi:10.1074/jbc.M113.459396.
- Xie, Gary, David C. Bruce, Jean F. Challacombe, Olga Chertkov, John C. Detter, Paul Gilna, Cliff S. Han, et al. 2007. « Genome Sequence of the Cellulolytic Gliding Bacterium

- Cytophaga hutchinsonii ». *Applied and Environmental Microbiology* 73 (11): 3536- 46. doi:10.1128/AEM.00225-07.
- Xu, Jian, Magnus K. Bjursell, Jason Himrod, Su Deng, Lynn K. Carmichael, Herbert C. Chiang, Lora V. Hooper, et Jeffrey I. Gordon. 2003. « A Genomic View of the Human-Bacteroides Thetaiotaomicron Symbiosis ». *Science* 299 (5615): 2074- 76. doi:10.1126/science.1080029.
- Yoshimura, Koji, Asae Toibana, et Kazuo Nakahama. 1988. « Human lysozyme: Sequencing of a cDNA, and expression and secretion by *Saccharomyces cerevisiae* ». *Biochemical and Biophysical Research Communications* 150 (2): 794- 801. doi:10.1016/0006-291X(88)90461-5.
- Zheng, Baisong, Wen Yang, Xinyu Zhao, Yuguo Wang, Zhiyong Lou, Zihao Rao, et Yan Feng. 2012. « Crystal Structure of Hyperthermophilic Endo- β -1,4-glucanase ». *The Journal of Biological Chemistry* 287 (11): 8336- 46. doi:10.1074/jbc.M111.266346.