Characterizing xylan-degrading enzymes from a putative Xylan Utilization System derived from termite gut metagenome

Haiyang Wu

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Présentée et soutenue par :
Haiyang WU

le vendredi 23 mars 2018

Titre :

Characterizing xylan-degrading enzymes from a putative Xylan Utilization System derived from termite gut metagenome

Caractérisation des enzymes xylanolytiques d'un locus d'utilisation du xylane issu d'un métagénome de termite

---

École doctorale et discipline ou spécialité :
ED SEVAB : Ingénieries microbienne et enzymatique

Unité de recherche :
Laboratoire d'Ingénierie des Système Biologiques et des Procédés (UMR INRA 792 et CNRS 5504)

Directeur/trice(s) de Thèse :

Claire DUMON - Chargée de recherche - INRA de Toulouse
Michael J. O'DONOHUE - Directeur de Recherche - INRA de Toulouse

Jury :

Evelyne FORANO - Directrice de recherche - INRA, Saint-Genès-Champanelle - Rapporteur
Harry J. GILBERT - Professeur - Université de Newcastle upon Tyne - Rapporteur
Bernard HENRISSAT - Directeur de recherche - CNRS, Aix-Marseille Université - Examinateur

NOM : WU Prénom: Haiyang

Titre: Caractérisation des enzymes xylanolytiques d'un locus d'utilisation du xylane issu d'un métagénome de termite

Spécialité: Sciences Ecologique, Vétérinaires, Agronomiques et Bioingénieries, Filière: Ingénieries microbienne et enzymatique

Année: 2018 Lieu: INSA de Toulouse

RESUME

Dans le contexte de la bioéconomie, la découverte et la caractérisation des enzymes capables de dégrader la paroi végétale est particulièrement intéressante pour l'utilisation de la biomasse lignocellulosique dans l'industrie. A cet égard, la métagénomique fonctionnelle est un outil puissant pour découvrir de nouvelles enzymes à partir d'écosystèmes microbiens variés, comme l'illustrent les travaux sur le tube digestif du termite *Pseudacanthotermes militaris*. Cette étude a fourni une mine d'informations et identifié un hypothétique locus d'utilisation du xylane (XUS), codant pour cinq glycosides hydrolases (GH) et une carbohydrate esterase (CE) de *Bacteroidales*.

Le XUS du métagénome de *Pseudacanthotermes militaris* contient une xylanase de la famille GH10 qui possède une organisation modulaire complexe dans laquelle la séquence du domaine GH10 est interrompue par une insertion de deux carbohydrate binding modules (CBM). Des travaux préliminaires ont montré que cette enzyme modulaire, désignée *Pm*25, est active sur xylane. Par conséquent, un des objectifs de cette étude a été la caractérisation détaillée des propriétés biochimiques et catalytiques de *Pm*25. Le rôle des CBM a également été examiné en quantifiant les interactions protéines-sucres et permettant ainsi une meilleure compréhension du rôle spécifique de ces modules, les résultats obtenus permettent de cerner l'impact de la modularité de *Pm*25 sur ses propriétés fonctionnelles.

Dans une deuxième partie de l'étude, nous avons entrepris d'étudier la fonction de *Pm*25 dans le contexte du cluster XUS. Pour ce faire, nous avons étudié les enzymes adjacentes à *Pm*25 sur le locus, une autre GH10, une GH11, une GH115 et une GH43. La comparaison des paramètres cinétiques et une étude détaillée des produits d'hydrolyse ont été analysés par spectrométrie de masse et ont révélé que la GH10 et la GH11 étaient les enzymes clefs de la dépolymérisation en étant 20 fois plus efficaces que *Pm*25.

En parallèle, nous avons développé un protocole pour l'utilisation de la micro-thermophorèse (MST) pour quantifier les interactions CBM-sucres, une approche intéressante qui nécessite peut d'échantillon et de ligand contrairement à d'autres méthodes biophysiques.

Dans l'ensemble, cette étude a révélé le rôle important de *Pm*25 et ses homologues dans les locus d'utilisation des xylanes chez les *Bacteroidetes* et a permis d'identifié le sens de cette architecture particulière.

Mots clefs:

Xylanase, GH10, CBM, MST, XUS, Biomasse lignocellulosique, Bacteroidetes, métagénomique, bioeconomie

Ecole doctorale : Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Unité de recherche : Laboratoire d'Ingénierie des Système Biologiques et des Procédés (UMR CNRS 5504, UMR INRA 792) de l'INSA de Toulouse

Last Name: WU First Name: Haiyang

Title: Characterizing xylan-degrading enzymes from a putative Xylan Utilization System derived from termite gut metagenome.

Speciality: Ecological, Veterinary, Agronomic Sciences and Bioengineering; Field: Enzymatic and Microbial engineering

Year: 2018 Place: INSA de Toulouse

SUMMARY

In the context of bioeconomy, the discovery and study of plant-cell wall degrading enzymes is particularly relevant for the use of lignocellulosic biomass for industrial purposes. In this respect, functional metagenomics has proven to be a powerful tool to discover new enzymes from a variety of microbial ecosystems, as exemplified by work performed on the gut of the termite *Pseudacanthotermes militaris*. This study provided a wealth of information and identified an interesting hypothetical xylan utilization system, encoding five glycoside hydrolases (GH) and one carbohydrate esterase (CE) annotated from *bacteroidales*.

The *Pseudacanthotermes militaris*-derived putative XUS cluster contains a GH10 xylanase that displays a quite complex modular arrangement wherein the GH10 catalytic module contains two insertional carbohydrate binding modules (CBM). During the preliminary work, this modular enzyme, designated *Pm*25, was shown to be active on xylan, thus in the present research we set out to more thoroughly characterize its biochemical and catalytic properties. The role of the CBM was also investigated, quantifying protein-carbohydrate interactions and thus providing better insight into the specific role of the modules. Taken together, the results obtained provide insight into how *Pm*25 modularity translates into functional properties.

In second part of our study, we set out investigate the function of *Pm*25 in the context of the XUS cluster. To achieve this we studied a xylan utilization system, which is constituted by another GH10, GH11, GH115 and GH43. The comparison of kinetic parameters and a detailed end product analysis by mass spectrometry showed that *GH*10 and *GH*11 outweigh over 20 fold *Pm*25 catalytic efficiency.

In parallel, we developed the use of MicroScale Thermophoresis (MST) to quantify CBM - carbohydrates interactions, an interesting approach requiring smaller concentration of proteins and ligands compared to other biophysical methods.

Overall this study highlighted the important role of *Pm*25 homologs in the xylan utilization system in *Bacteroidetes*, and pinpointed the meaning of its unusual architecture.

Keywords:

Xylanase, GH10, CBM, MST, XUS, lignocellulosic biomass, Bacteroidetes, metagenomic, bioeconomy

Doctoral school : Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Research unit : Laboratoire d'Ingénierie des Système Biologiques et des Procédés (UMR CNRS 5504, UMR INRA 792) de l'INSA de Toulouse

PUBLICATIONS

Published

Wu H., Montanier C.Y., Dumon C. (2017) Quantifying CBM Carbohydrate Interactions Using Microscale Thermophoresis. In: Abbott D., Lammerts van Bueren A. (eds) Protein-Carbohydrate Interactions. Methods in Molecular Biology, vol 1588. Humana Press, New York, NY

To be submitted

Wu H., Ioannou E, Montanier C.Y., O'Donohue M., Dumon C. Understanding the multimodularity of a Xyn10C-like protein found in termite gut.

Wu H., Montanier C.Y., O'Donohue M., Dumon C. Investigation of the function of a putative xylan utilization locus from termite gut microbiome.

COMMUNICATIONS ORALES

1. INRA-Séminaire BILI (BIoraffinerie des LIgnocelluloses), March 11, 2016, Paris, France.
   Oral presentation on topic *Characterization of an unusual GH10 xylanase from termite gut.*

2. INSA-CIMEs (Catalysis and Enzyme Molecular Engineering Team) meeting, February 2nd, 2017, Toulouse, France.
   Oral presentation on topic Characterization of a multi-modular xylanase belonging to a potential microbial xylan utilization system.

3. 12th Carbohydrate Bioengineering Meeting, April 23-26, 2017, Vienna, Austria.
   Oral presentation on topic Characterization of an unusual GH10 xylanase from termite gut.

COMMUNICATIONS PAR POSTERS

1. Characterization of an unusual GH10 xylanase from termite gut.
   Wu H., Ioannou E, Montanier C., Arnal G., Fernandez-Fuentes N., O'Donohue M., Dumon C. ; Summer Course Glycosciences (14th European Training Course on Carbohydrates) , 12-16 June 2016, Groningen, The Netherlands.
2. Characterization of an unusual multimodular GH10 xylanase
   Ioannou E, **Wu H.**, Cioci G., Guyez B., Arnal G., Dumon C, Bryant D, Fernandez-Fuentes N., O'Donohue M.; 11st Carbohydrate Bioengineering Meeting, May 10-13, 2015, Espoo, Finland.

To my parents and to my soul mate Lei.

ACKNOWLEDGEMENTS

This manuscript is the miniature of three and a half year's thesis. Without the help and support from all the kind people here, I can't manage to go through it by myself.

I would like to thank my jury members, Evelyne Forano, Harry J Gilbert, and Bernard Henrissat. It is my great honor to have you in my thesis defense.

Claire and Mike, I could not to find a word to express how much I am grateful to you two. Claire, the way you are talking comforted me when I am too nervous to do a presentation, and the way you showed me to do the schedule, simplified my life. Every problem can be solved if you just start to tackle it step by step. What is more, your rigorous experimental design and cautious conclusion impressed me a lot. Mike, thank you for always broadening my vision by guiding me a new world in science. You let me realize everything could be a scientific question if you deeply look into it. Furthermore, I can't forget in my life the scene that you wrote in capitals about all the aspects about my thesis project when the first day I came into the lab, which is like an exciting opening ceremony of my thesis. And I would like to thank you two for your immense support in my career, patience in supervising me, and your confidence in me.

Cédric, thank you for being there whenever I met the problem with the rotor and your humor to make me relax coping with these kinds of problems. Thank you for your great effort in the investigation of CBMs with the MST technology, which helped me have my first book chapter. Every time you never gave up explaining a complex abstract concept to me until I fully understood, and I am impressed with your extremely careful and precise experimental performance.

Sophie, thank you for being in my monthly meeting and giving your invaluable comments. I will never forget your helping me register in LISBP. Your encouragement as well as hearty laugh comforted and motivated me a lot.

Régis, thank you for every "Bonjour & How are you" in the lab and office, this routine means a lot to me. Thank you for your helping me ameliorate the manuscript and slides for the CBM 12 meeting. Thanks to you, I won't forget to pay attention to every single punctuation and font style in the manuscript.

Gianluca, I would like to specially thank you for everything, far more than your assistance in the Akta platform, circular dichroism, DSF et al. But also because your standing point is always at other people's place and you may not notice how your kindness deeply affected others and motivated their career and life.

Thanks to Nelly for her help on the analytical platform, and especially for her patience in solving instrumental problems for me. Besides, the brochure introducing the manipulation of the device is amazing, thank you for your delicate work.

My heartfelt thanks to Bernard Herissat, thank you for sharing your time to solve our scientific problem with your amazing CAZy database. In less than three hours, your meticulous thinking, never give up spirits, high curiosity, and organization skills fully explained me the definition of a great man in academia. You are our idol.

My dear CIME labmates, thank you, every one of you for your accompany for these three years and sharing my pain and happiness in the experimental work, conference and so on. Recalling the very first day when I joined this big family, Nathalie, Lisa, Yvonne, Coralline, Pablo, Barbara, Julien, Louise, Eleni, Alex, Yanick, I got a lot of help from you and appreciate it so much. I am grateful for the atmosphere you contributed. Many activites, *Jeudi Bière*, *Recipe exchange*, *discover new flavor in Toulouse* and so on, I never expect that my PhD life could be so colorful. Jiao, Zhi, Ao, thank you so much for making us the Chinese team in EAD1, our sharing knowledge in science, sharing feeling in study in foreign countries, sharing Chinese food with each other, did make me feel at home and let my last year in PhD less stressful. Lénaik, thank you very much for your assistance in my project. You are the best, sweetest intern ever! I am so lucky to have you involve in my PhD.

I gratefully acknowlege the support of the Chinese Scholarship Council that funded my 42-months of thesis.

Finally I would like to thank my family. My father and mother, thank you for letting me immerse in your love, and feel free to pursue my dream. You are the greatest parents in the whole world. The love you give me is the strongest power in me, which help me conquer any difficulties and setbacks. 爸爸妈妈，我爱你们。Thanks to my beautiful sister, you are there helping me do a daughter's duty and always have ways to make our parents happy. Many thanks to my grandparents, who cannot see me get the highest degree. You are the brightest stars in the night sky, and accompany me wherever I am. Last but not least, my darling, Lei. Without you, I am not brave enough to go abroad alone. Thank you for sharing the ups and downs in life together with me.

TABLE OF CONTENTS

LIST OF FIGURES

ARTICLE II

ARTICLE III

LIST OF TABLES

BIBLIOGRAPHY

ARTICLE I

ARTICLE II

ABBREVIATIONS

Abfs: α-L-arabinofuranosidase

AGE: affinity gel electrophoresis

BGX: beechwood glucuronoxylan

BSA: bovine serum albumin

CAZy: carbohydrate-Active EnZymes

*Cj*Xyn10C: *Cellvibrio japonicas* Xylanase 10C

*Cj*Xyn10A: *Cellvibrio japonicas* Xylanase 10A

*Cm*Xyn10A: *Cellvibrio mixtus* Xylanase 10A

CBM: carbohydrate binding module

CBP: consolidated bioprocessing

DPs: degrees of polymerization

D-Xyl: D-xylopyanosyl

EG: endoglucanase

GAX: glucuronoarabinoxylan

GX: glucuronoxylan

GH: glycoside hydrolase

GlcA: D-glucuronic acid

HG: homogalacturonan

IUB: International Union of Biochemistry and Molecular Biology

K_d: dissociation constant

L-Araf: L-arabinofuranosyl

MeGlcA: 4-O-methyl-D-glucuronic acid

MW: molecular weight

MST: microscale thermophoresis

PCW: plant cell wall

pI: isoeletric point

PUL: polysaccharide utilization loci

RAX: rye arabinoxylan

RG1: rhamnogalacturonan i

RG2: rhamnogalacturonan II

SSN: sequence similarity network

SUS: starch utilization system

WAX: wheat arabinoxylan

XG: xyloglucan

XOS: xylooligosaccharide

XUS: xylan utilization system

RESUME EN FRANÇAIS

Introduction

La découverte et l'optimisation de biocatalyseurs constitue un levier pour le développement de la bioéconomie et particulièrement dans le domaine de la valorisation des agro-ressources dont les polysaccharides de la paroi végétale constituent la ressource de carbone la plus abondante sur terre. La métagénomique est une technique puissante pour explorer l'ensemble des génomes d'un écosystème en s'affranchissant des conditions de cultures des microorganismes (Handelsman, 2004; Madhavan et al., 2017), cette technique a permis un accroissement considérable des séquences disponibles dans les bases de données mais ne garantit pas l'activité des enzymes annotées. La métagénomique fonctionnelle est basée non pas sur l'exploration des séquences des génomes mais sur l'activité des enzymes codées par ces métagénomes et a permis de découvrir de nouvelles enzymes (Bastien et al., 2013; Fernández-Arrojo et al., 2010; Tasse et al., 2010).

Précédemment au laboratoire une campagne de métagénomique fonctionnelle du métagénome du tube digestif du termite champignonniste *Pseudacanthotermes militaris* a permis de découvrir 63 nouvelles enzymes dédiées à la modification des sucres (Bastien et al., 2013). Un fragment génomique a été identifié lors du criblage sur xylane, l'annotation de ce fragment a montré qu'il contient plusieurs gènes juxtaposés codant pour des activités xylanolytiques potentielles et appartenant aux familles de glycoside hydrolases GH10, GH11, GH43, GH115 et d'une carbohydrate estérase CE1. L'analyse bio-informatique a montré ces enzymes pourraient faire partie d'un locus d'utilisation des polysaccharides, PUL, puisque les GH sont codées après une paire de gène *sucC*/*sucD* rencontrés spécifiquement dans les PUL. De manière générale SusC/SusD sont responsables de la reconnaissance et de l'internalisation des substrats partiellement hydrolysés par les GH codées par les PUL, ce système constitue un paradigme pour l'hydrolyse des polysaccharides végétaux chez les bactéries anaérobies de l'embranchement des *Bacteroidetes* (Martens et al., 2009) initialement décrit pour l'hydrolyse de l'amidon, il a été étendu aux xylanes on parle alors du couple XusC/XusD (xylan utilisation system). La caractérisation des enzymes du PUL *Pm25* (GH10), *Pm115* (GH115), *Pm12* (GH10), *Pm24* (GH11) et *Pm21* (GH43) (Fig I-A) et plus particulièrement *Pm*25, a constitué le cœur de cette thèse.

*Pm*25 possède une activité xylanase (thèse de G. Arnal), serait une xylanase multi modulaire constituée d'un domaine catalytique de la famille GH10 et de deux domaines non catalytiques ou CBM pour carbohydrate binding domain de la famille 4 de la base de données CAZy (Lombard et al., 2014). La particularité de cette enzyme provient de son organisation puisque

les deux CBM4 sont insérés dans la séquence primaire du domaine catalytique GH10. Cette architecture a retenu notre attention et la caractérisation de cette enzyme constitue la première partie du travail présenté dans cette thèse. Les autres glycosides hydrolases codées par ce cluster de gène sont hypothétiquement actives pour la dégradation des xylanes (Fig I-A) soit en dépolymérisant la chaine principale, activité xylanase, soit en hydrolysant les substitutions de cette même chaine, xylosidase, glucuronidase. La caractérisation des enzymes du PUL et les synergies intra et inter-moléculaires sont étudiées dans le chapitre II. Et enfin, micro-thermophorèse a été utilisée pour quantifier l'affinité du CBM4 de *Pm*25, cette mise au point a permis la rédaction d'un chapitre dans Methods in Molecular Biology (Wu et al., 2017) qui est présenté dans le chapitre III.

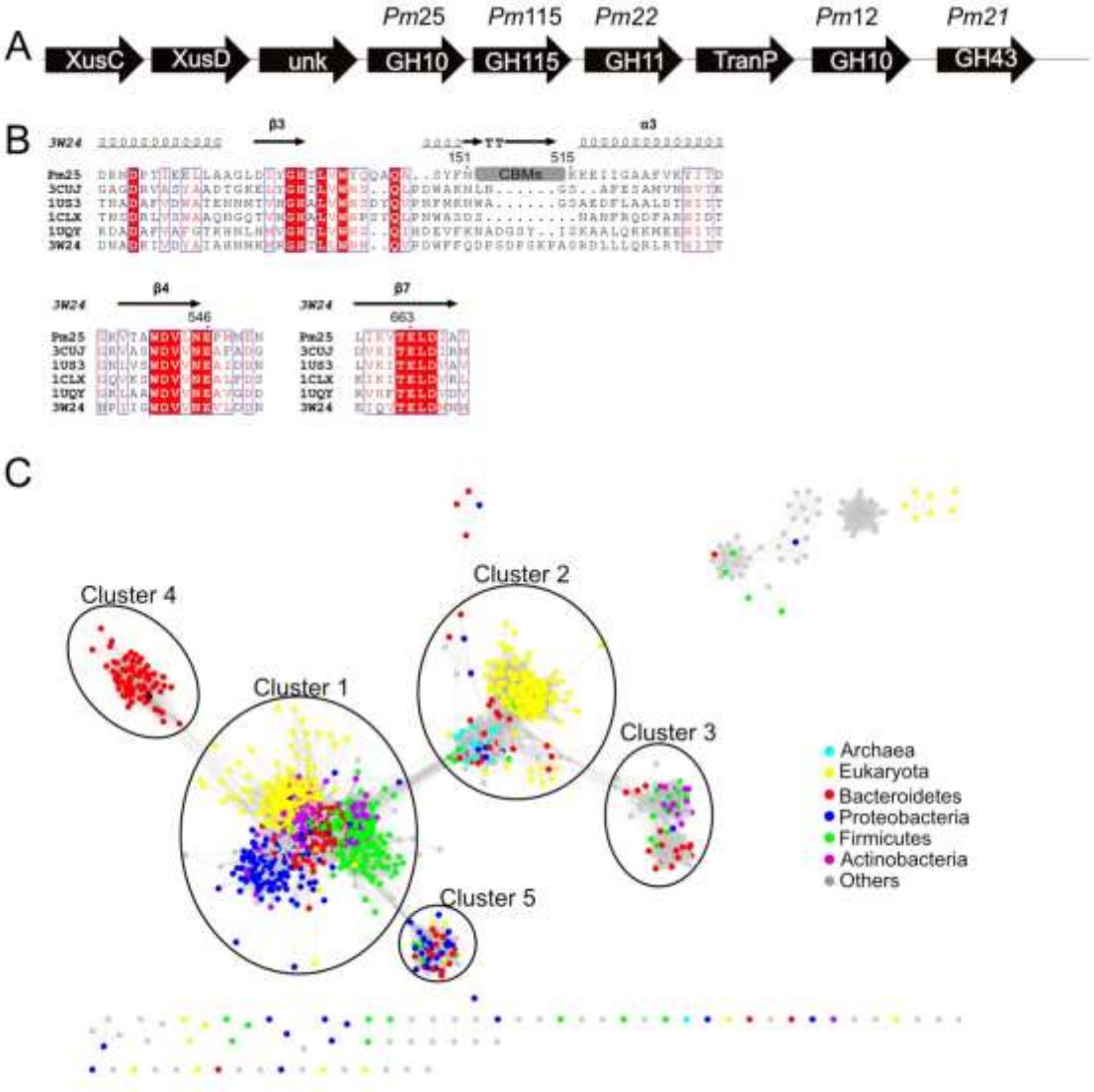


Fig I Présentation schématique et réseaux de séquences similaires (SSN).

(A) Représentation schématique du cluster de gène xylanolytique. (B) Alignement de la séquence de *Pm*25 avec les séquences des xylanases GH10 caractérisées et dont la structure est connue (identifiées par leur code PDB). (C) Réseau de séquences similaires construit avec 5068 séquences de GH10 non redondantes. Chaque nœud représente des séquences avec 60% d'homologie et les segments reliant deux nœuds représentent les séquences partageant au moins 35% d'homologie.

Part A- Caractérisation d'une xylanase multi modulaire de la famille GH10 issue d'une banque metagénomique du tube digestif de termite.

L'analyse *in silico* de la séquence multi modulaire de *Pm*25 a permis d'identifier les frontières des différents modules. La séquence est composée d'un peptide signal (residus 1-32), suivi d'une première partie du domaine GH10 (residus 66-151), de deux CBM4s (CBM4-1, résidus 161 to 321 and CBM4-2, résidus 324 to 486) et terminant par la deuxième partie du domaine catalytique GH10 (résidus 514-753). L'alignement de *Pm*25 avec les séquences de structure connues montre que l'insertion des CBM4 se situe entre le troisième feuillet beta et la troisième hélice alpha (Fig I-B), ce qui est en accord avec les homologues de *Pm*25, aussi dénommés XynC-like, décrits dans la littérature (Flint et al., 1997).

Dans un premier temps, une analyse des réseaux d'homologie de séquences, sequence similarity network, SSN, a été réalisé afin de savoir si les homologues de *Pm*25 parmi la famille des GH10 partageaient les mêmes propriétés. Cette approche permet de facilement représenter des milliers de séquences de façon bien plus lisible qu'un arbre phylogénétique (Gerlt et al., 2015). L'analyse SSN réalisée à partir des 5068 séquences de GH10 non redondantes a montré que toutes les séquences analogues à *Pm*25 se retrouvent dans le même groupe composé uniquement de séquences de *Bacteroidetes* (cluster 4, Fig I-C) et ceci est vérifié si les séquences des CBM4 sont retirés de la séquence totale des GH10 indiquant que ce regroupement ne vient pas uniquement de la présence des CBMs. Il est ressorti que toutes les protéines homologues à *Pm*25 sont codées après des paires *susC/susD* avec quelques exceptions (4 sur 79), ce qui est en accord avec son implication dans un locus d'utilisation des xylans chez les *Bacteroidetes* (Dodd et al., 2011), on parle alors de XUS.

Afin d'étudier la fonction des différents domaines de *Pm*25, différentes mutations ou troncatures ont été réalisées et les protéines exprimées (Fig II). La protéine *Pm*25 a été

caractérisée biochimiquement sur différents arabinoxylanes, et l'affinité pour le substrat grâce à un mutant inactif (M1) ou de la protéine sans CBM (M2) a permis de montrer que l'affinité est guidée par les CBM (M3). (Fig IIA, milieu). Des différences importantes ont été constatées entre la construction M3 (CBMs) et M4 / M5 (chaque CBM séparé) indiquant une coopérativité entre les deux CBM4s (Fig IIA, milieu). Enfin des tests d'affinité sur substrat insoluble (solid depletion assay) ont confirmé le rôle des CBMs pour l'affinité avec la biomasse complexe telle que le son de blé (Fig IIA, panneau droit).

Les constructions de *Pm*25 dont les acides aminés importants pour la fixation au ligand ont été mutés et ont montré les mêmes profils de produits que la construction sans CBM, de plus les quantités de xylooligosaccharides libérés ont été plus importantes pour *Pm*25 sauvage confirmant ainsi les rôles des CBM dans l'hydrolyse de la biomasse complexe (Fig IIB).

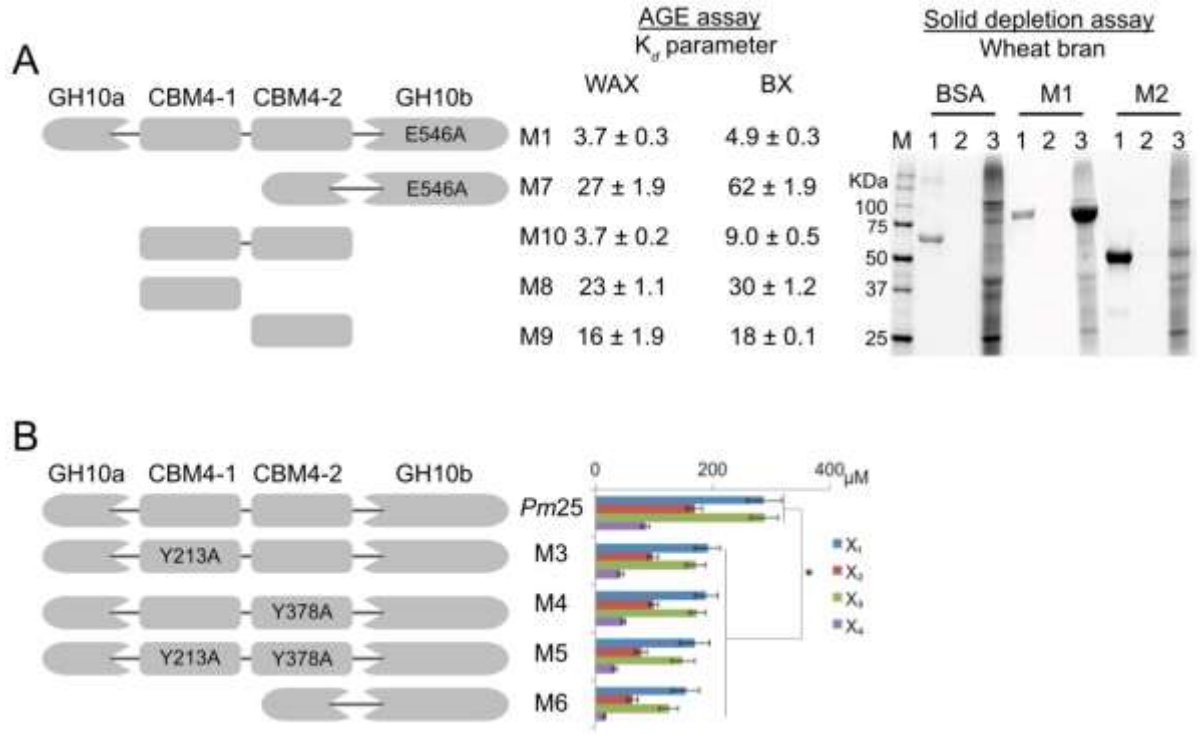


Fig II Comparaison des différents variants de *Pm*25.

(A) Gauche: représentation schématique des différentes constructions dérivées de Pm25 avec leurs constantes de dissociations (kd, mg/ml) sur l'arabinoxylane de blé (WAX) et le xylane de hêtre (BX) (panneau milieu). Trx, étiquette thioredoxine. Droite: Affinité sur substrat insoluble des constructions avec et sans CBM (M1/M2) sur son de blé. L'albumine de sérum bovin (BSA) a été utilisée comme control négatif. Fraction non-fixée (ligne 1), lavage (ligne 2) et fraction fixée (ligne 3). Ligne M, marqueur de masse moléculaire. (B) Profil des produits finaux après 14h d'hydrolyse du son de blé (à gauche). X1, xylose ; X2, xylobiose ; X3, xylotriose, X4, xylotetraose. *, $p \leq 0.01$.

Part B- Le role des CBMs de Pm25 en présence des enzymes du PUL.

Trois types de synergies peuvent être étudiés lorsque l'on s'intéresse à l'hydrolyse des polysaccharides: la synergie intermoléculaire entre des enzymes aux activités différentes (endo/exo-glucanases), ou entre des activités du meme types (exo/exo-glucanases ou endo/endo-glucanase), et la synergy intramoléculaire entre les différents domaines au sein d'une protéine multi modulaire tel que CBM/ domaine catalytique. Différentes synergies ont pu être observés pour différentes combinaisons ou orientations des CBM et de leur domaines catalytiques (Meng et al., 2015)(Tajwar et al., 2017). Les CBM sont retrouvés couplés à des GH, GT et des toxines et peuvent jouer différents rôles (Boraston et al., 2004; Gilbert et al., 2013; Guillén et al., 2010). La fonction la plus décrite des CBM est d'améliorer l'activité des enzymes auxquelles ils sont rattachés en permettant une plus grande proximité entre le l'enzyme et son substrat.

Cependant, dans certains cas, aucun effet du CBM sur l'activité du domaine catalytique n'a pu être observé (Ali et al., 2001; Black et al., 1996; Kellett et al., 1990; Valenzuela et al., 2012; Zhao et al., 2005). Certaines études sont basées sur des substrats modèles, de petite taille comme des oligosaccharides d'un degré de polymérisation, DP, inférieur à sept, ce qui n'est peut-être pas suffisant pour complètement invalidé la contribution du CBM. Au cours de ce deuxième chapitre l'activité des différentes constructions de *Pm*25 avec ou sans CBM ont été comparées en analysant les produits d'hydrolyse du xylane de hêtre. Les xylo-oligosaccharides XOS ont été analysés par HPLC pour les XOS DP1 à 6 et par spectrométrie de masse pour les XOS de DP 7 à 28. Afin de caractériser la synergie intermoléculaires entre les enzymes du PUL, les enzymes *Pm*12, *Pm*21, *Pm*24, *Pm*25 et *Pm*115 ont été produites et caractérisées. Le xylane

de hêtre a été hydrolysé avec différents cocktails des enzymes du PUL et les produits d'hydrolyse analysés.

La complémentarité des enzymes du PUL a pu être mise en évidence, néanmoins il n'y a pas de synergie notable. Les résultats montrent une faible contribution des CBM sur le substrat soluble suggérant que les CBM ont peut-être un rôle plus important sur une biomasse complexe.

Part C- Quantification des intéractions CBM-sucre par micro-thermophorèse.

Les interactions protéines – sucre peuvent être caractérisées par différentes techniques telles que les gels d'affinités par électrophorèse (AGE), la microcalorimétrie (ITC) qui renseigne sur les paramètres thermodynamiques de l'interaction ou la spectroscopie de fluorescence, basées sur des interactions avec des substrats solubles, (Abbott and Boraston, 2012). La micro-thermophorèse (MST) est une technologie émergeante pour étudier de nombreuses interactions moléculaires avec une grande sensibilité. Parmi les avantages on peut citer les volumes de réactions à l'échelle du microlitre ainsi que la rapidité du test de quelques minutes.

Cette technique a été utilisée pour quantifier l'affinité du CBM3a de *Clostridium Thermocellum* pour des nano-cristaux de cellulose et du CBM4 de Pm25 pour le xylohexaose. Apres le marquage et la purification des protéines l'expérience a été optimisée pour les paramètres de mesures et de traitement des données. Une constante de dissociation apparente a pu être déterminée pour le CBM3a et les nano-cristaux de cellulose (*0.24* ± 0.05 g/L). Cependant, du quenching de fluorescence a pu être observé pour le CBM4, ceci a pu être expliqué par la présence d'une lysine potentiellement marquée à proximité du site de fixation du ligand. Dans ce cas, la constante de dissociation du CBM4 pour le xylohexaose a pu être déterminée (2.5 ± 0.3 mM) par la mesure de la fluorescence sans utiliser la thermophorèse démontrant ainsi l'utilité de cette technique pour quantifier les intéractions CBM-sucres.

Référence:

Ali, M.K., Hayashi, H., Karita, S., Goto, M., Kimura, T., Sakka, K., and Ohmiya, K. (2001). Importance of the carbohydrate-binding module of Clostridium stercorarium Xyn10B to xylan hydrolysis. Biosci. Biotechnol. Biochem. *65*, 41–47.

Bastien, G.G., Arnal, G.G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., Henrissat, B., Lefèvre, F., Robe, P., Bouchez, O., et al. (2013). Mining for hemicellulases in the fungus-

growing termite Pseudacanthotermes militaris using functional metagenomics. Biotechnol. Biofuels *6*, 78.

Black, G.W., Rixon, J.E., Clarke, J.H., Hazlewood, G.P., Theodorou, M.K., Morris, P., and Gilbert, H.J. (1996). Evidence that linker sequences and cellulose-binding domains enhance the activity of hemicellulases against complex substrates. Biochem. J. *319*, 515–520.

Boraston, A.B., Bolam, D.N., Gilbert, H.J., Davies, G.J., Zverlov, V. V, Volkov, I.Y., Velikodvorskaya, G.A., and Schwarz, W.H. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. *382*, 769–781.

Dodd, D., Mackie, R.I., and Cann, I.K.O. (2011). Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. Mol. Microbiol. *79*, 292–304.

Fernández-Arrojo, L., Guazzaroni, M.E., López-Cortés, N., Beloqui, A., and Ferrer, M. (2010). Metagenomic era for biocatalyst identification. Curr. Opin. Biotechnol. *21*, 725–733.

Flint, H.J., Whitehead, T.R., Martin, J.C., and Gasparic, A. (1997). Interrupted catalytic domain structures in xylanases from two distantly elated strains of Prevotella ruminicola. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. *1337*, 161–165.

Gerlt, J.A., Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkhin, B., Slater, D.R., and Whalen, K.L. (2015). Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. Biochim. Biophys. Acta - Proteins Proteomics *1854*, 1019–1037.

Gilbert, H.J., Knox, J.P., and Boraston, A.B. (2013). Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. Curr. Opin. Struct. Biol. *23*, 669–677.

Guillén, D., Sánchez, S., Rodríguez-Sanoja, R., Guillen, D., Sanchez, S., and Rodriguez-Sanoja, R. (2010). Carbohydrate-binding domains: multiplicity of biological roles. Appl. Microbiol. Biotechnol. *85*, 1241–1249.

Handelsman, J. (2004). Metagenomics : Application of Genomics to Uncultured Microorganisms. Microbiol. Mol. Biol. Rev. *68*, 669–685.

Kellett, L.E., Poole, D.M., Ferreira, L.M., Durrant, a J., Hazlewood, G.P., and Gilbert, H.J. (1990). Xylanase B and an arabinofuranosidase from Pseudomonas fluorescens subsp. cellulosa contain identical cellulose-binding domains and are encoded by adjacent genes. Biochem. J. *272*, 369–376.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *42*, 490–495.

Madhavan, A., Sindhu, R., Parameswaran, B., Sukumaran, R.K., and Pandey, A. (2017).

Metagenome Analysis: a Powerful Tool for Enzyme Bioprospecting. Appl. Biochem. Biotechnol. 1–16.

Martens, E.C., Koropatkin, N.M., Smith, T.J., and Gordon, J.I. (2009). Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. J. Biol. Chem. *284*, 24673–24677.

Meng, D.-D., Ying, Y., Chen, X.-H., Lu, M., Ning, K., Wang, L.-S., and Li, F.-L. (2015). Distinct Roles for Carbohydrate-Binding Modules of Glycoside 10 (GH10) and GH11 Xylanases from Caldicellulosiruptor sp. Strain F32 in Thermostability and Catalytic Efficiency. Appl. Environ. Microbiol. *81*, 2006–2014.

Tajwar, R., Shahid, S., Zafar, R., and Akhtar, M.W. (2017). Impact of orientation of carbohydrate binding modules family 22 and 6 on the catalytic activity of Thermotoga maritima xylanase XynB. Enzyme Microb. Technol. *106*, 75–82.

Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Leclerc, M., et al. (2010). Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome Res. *20*, 1605–1612.

Valenzuela, S.V., Diaz, P., and Javier Pastor, F.I. (2012). Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module. Appl. Environ. Microbiol. *78*, 3923–3931.

Wu, H., Montanier, C.Y., and Dumon, C. (2017). Quantifying CBM Carbohydrate Interactions Using Microscale Thermophoresis. In Methods in Molecular Biology, (Humana Press, New York, NY), pp. 129–141.

Zhao, G., Ali, E., Araki, R., Sakka, M., Kimura, T., and Sakka, K. (2005). Function of the family-9 and family-22 carbohydrate-binding modules in a modular β-1,3-1,4-glucanase/xylanase derived from Clostridium stercorarium Xyn10B. Biosci. Biotechnol. Biochem. *69*, 1562–1567.

BIBLIOGRAPHY

PART A COMPOSITION AND STRUCTURE OF PLANT CELL WALLS (PCWS)

Unlike animal cells, plant cells are surrounded by a plant cell walls (PCWs) that are extracellular composite structures that contain polysaccharides as well as other chemical components. Primarily, PCWs serve as a physical barrier that defends the protoplast from mechanical injuries, osmotic shock and ultraviolet irradiation (Katō, 1981). Secondly, like scaffolding, PCWs buttress the cell and act as transmitters of signals following microbial assault (Sarkar et al., 2009; Somerville et al., 2004). Furthermore, PCWs are of significant economic importance, because it is widely recognized that they constitute the most abundant reservoir of organic carbon on the planet (Okkerse and van Bekkum, 1999). Finally, PCWs are also of nutritional importance, because they constitute a major source of dietary fiber (Selvendran, 1984).

While evolution has conferred many properties to PCWs that are beneficial to the plant, these properties are not always advantageous regarding the economic exploitation of PCWs. For example, in a biorefinery context, the natural resilience of PCWs towards abiotic and biotic aggressions means that a lot of energy and complex operations are required to extract renewable carbon from PCWs. The overall resistance of PCWs to processing is usually referred to as biomass "recalcitrance" (DeMartini et al., 2013a; McCann and Carpita, 2015). This is a composite property that arises from many factors including the chemical and structural complexity of PCWs.

A.1 Composition of plant cell walls

The chemical composition of PCW is heterogeneous, varying with the developmental stages, tissue types and plant species, which reflect the different needs of plants (Somerville et al., 2004). In general, in lignocellulosic biomass, cellulose, hemicellulose and lignin constitute 30-40%, 20-30%, and 10-20% of the dry weight of PCWs, respectively (Davison et al., 2013; Varner et al., 1989; White et al., 2014) (FigA-1and 2). Compositional variation is often linked to different species, for example with lignin being more abundant in softwood (Stephen, 1983).

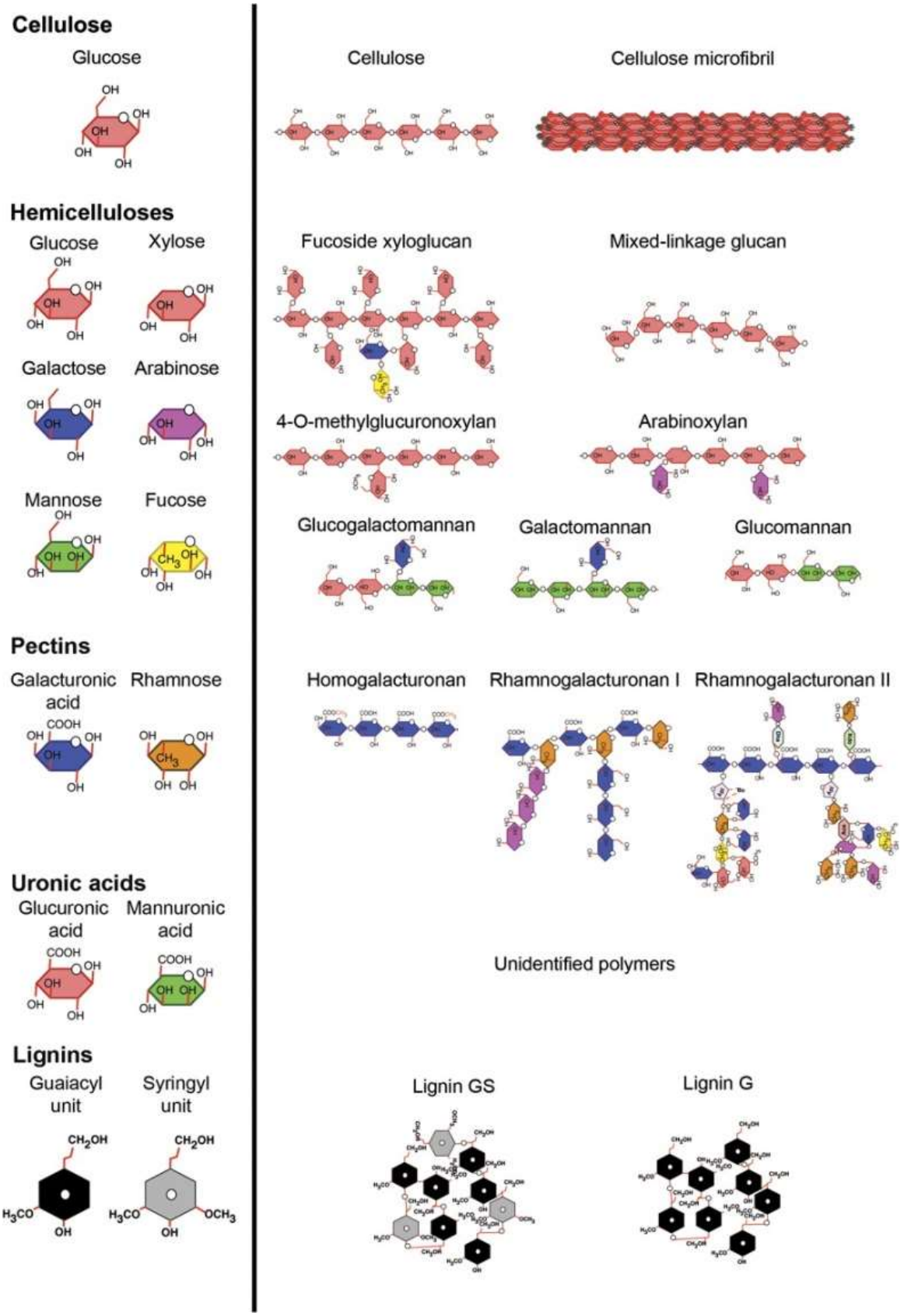

Cellulose
Glucose
Cellulose
Cellulose microfibril
Hemicelluloses
Glucose
Xylose
Galactose
Arabinose
Mannose
Fucose
Fucoside xyloglucan
Mixed-linkage glucan
4-O-methylglucuronoxylan
Arabinoxylan
Glucogalactomannan
Galactomannan
Glucomannan
Pectins
Galacturonic acid
Rhamnose
Homogalacturonan
Rhamnogalacturonan I
Rhamnogalacturonan II
Uronic acids
Glucuronic acid
Mannuronic acid
Unidentified polymers
Lignins
Guaiacyl unit
Syringyl unit
Lignin GS
Lignin G

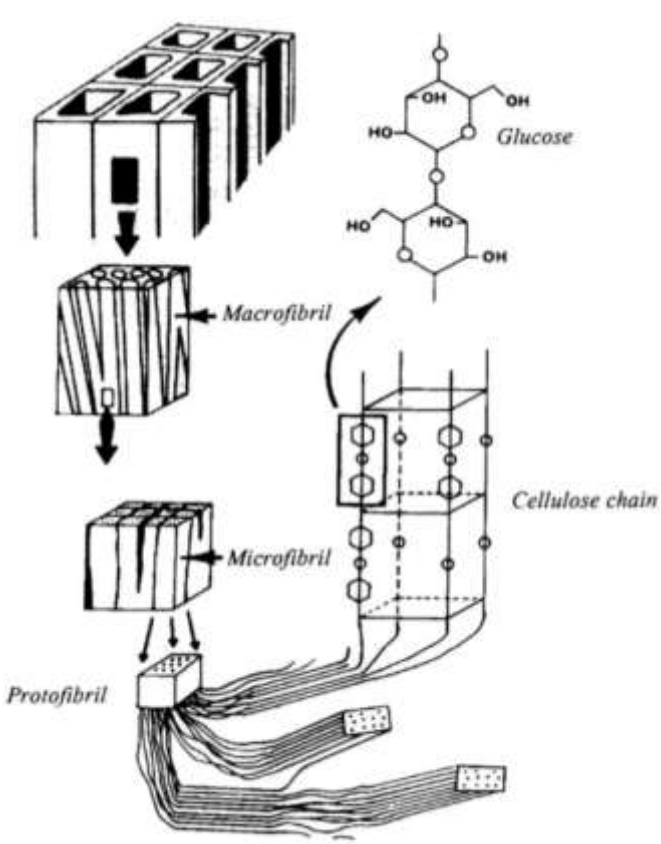
Fig A-1. Chemical structure of the predominant building blocks of plant cell walls (Sarkar et al., 2009)

A.1.1 Cellulose

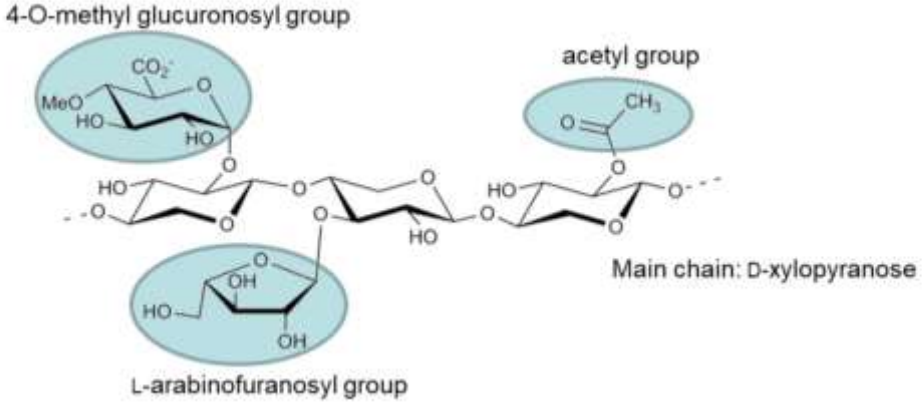
In 1838, Anselme Payen discovered and isolated cellulose from green plants (Payen, 1838). Cellulose is the main constituent of plant cell walls, but is also distributed in bacteria, fungi, algae and even in animals, although the amount of cellulose in the latter e.g. in tunicates is negligible when compared with plants (O'sullivan, 1997). Cellulose is used as an important raw material for paper products, fibers, and biofuels.

Cellulose is a high molecular polymer whose chain is constituted of D-glucose linked by β-1,4- glycosidic bonds (McKendry, 2002). Natural cellulose can have degrees of polymerization (DPs) in the range 5,000 to 15,000 glucose monomer units (O'sullivan, 1997). It is present in PCW as fibrils (about 500 nm in diameter, (Chen, 2014)), which are composed of very fine threads (or ribbons) called microfibrils containing approximately 10^3 chains molecular chains of (1→4)- β-D-glucan ranged in parallel, which may vary from 10-25 nm in breadth and can be more than 7 µm long (14,000 D-glulcose units) depending on the species (Gibson, 2012; Somerville et al., 2004; Tanner and Loewus, 1981)(Fig A-2A). Glucan chains in the microfibrils interact with each other via hydrogen bonds and can form amorphous (irregularly arranged) or crystalline regions (orderly arranged), which when observed by wide-angle X-ray scattering appear as dark and light areas respectively (O'sullivan, 1997; Tanner and Loewus, 1981). Overall, the properties of cellulose make it play a central structural role in plants (Lodish et al., 2000).

A


Glucose
Macrofibril
Cellulose chain
Microfibril
Protofibril

B


4-O-methyl glucuronosyl group
acetyl group
Main chain: D-xylopyranose
L-arabinofuranosyl group

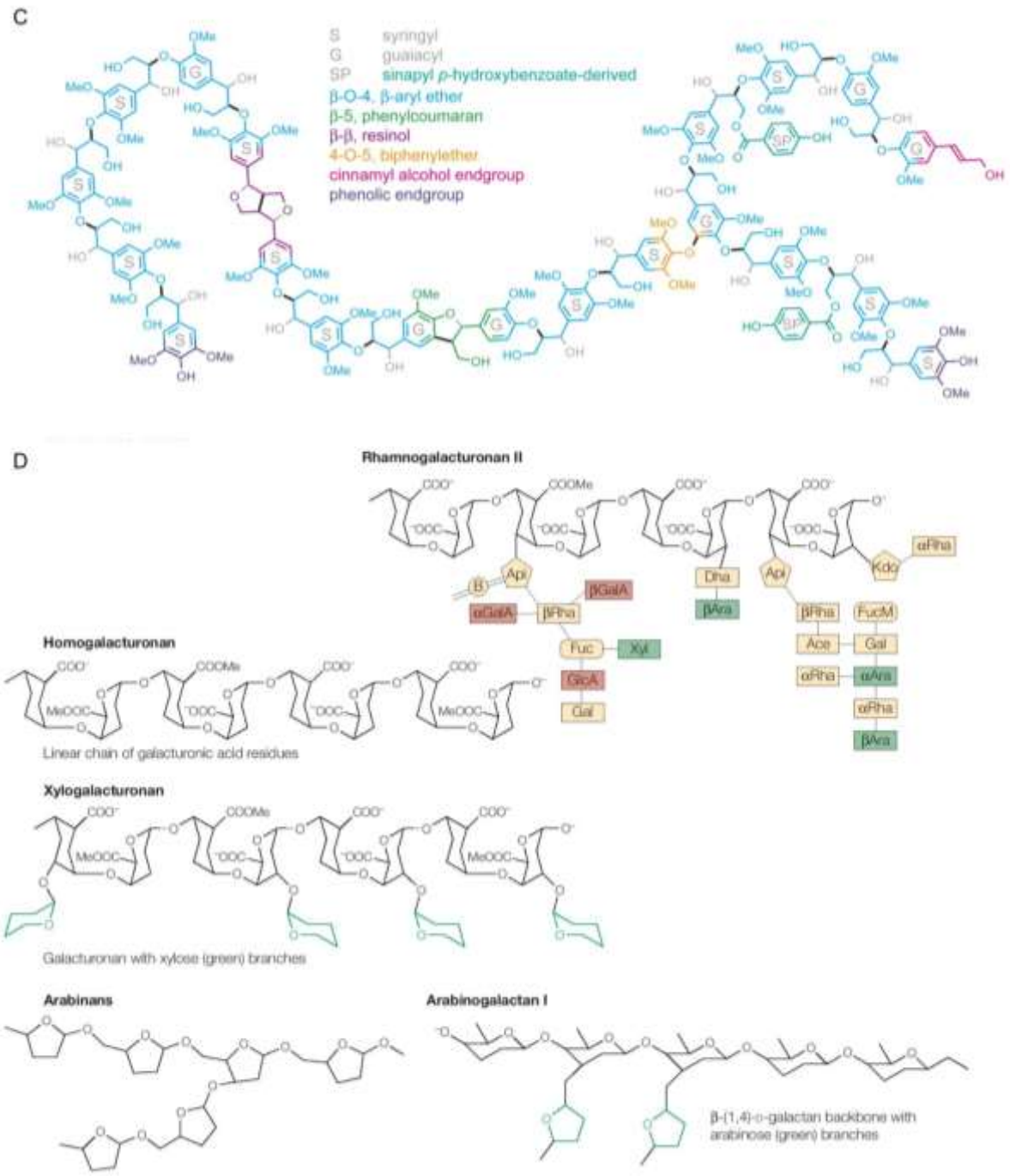


Fig A-2. Structural and chemical representation of cellulose (A)(Baucher et al., 2003), hemicellulose (B) , lignin (C) (Vanholme et al., 2010) and pectin (D) (Cosgrove, 2005).

A.1.2 Hemicelluloses (Xylan)

The term hemicelluloses was coined by E. Schulze in 1891. This term includes all PCW components that are fairly amenable to extraction and hydrolysis (Preece, 1931). Hemicelluloses cover a quite large, diverse family of (usually) heterogenous polymers that are nevertheless all based on a mainchain composed of β-linked D-sugars. In general,

hemicelluloses are amorphous, branched, single-chain polysaccharides displaying an average length of about 30 nm (Maureen et al., 1992). Hemicelluloses mainchains can contain pentose (D-xylose, L-arabinose), hexoses (D-mannose, D-glucose and D-galactose) and be decorated by a wide variety of sidechain substitutions, including uronic acids (α-D-glucuronic, α-D-4-O-methyl-galacturonic and α-D-galacturonic acids) and L-arabionfuranose (Fig A-2B). D-pentose sugars is more prevalent than D-hexose, while the latter is typically found in softwoods (Davison et al., 2013). According to the difference in the mainchain, hemicelluloses can be roughly divided into xylan, xyloglucan (XG), glucomannan, mannan and arabinogalactan *et al*. Different plants contain different main hemicelluloses with acetylated (galacto) glucomannan (as well as arabinoglucuronoxylan), glucuronoxylan, and glucuronoarabinoxylan are major hemicellulose in softwood, hardwood, and grassy species, respectively (*c.f.* Table A-1) (Scheller and Ulvskov, 2010). In addition to its structural role, hemicelluloses in PCWs can fulfill other roles, such as signaling molecules and seed storage carbohydrates (Scheller and Ulvskov, 2010).

Table A-1. Occurrence of hemicelluloses in primary and secondary cell walls of plants (Scheller and Ulvskov, 2010)

Amout of polysaccharide in wall (% w/w)[a]						
Polysaccharide	Dicot walls		Grass walls		Conifer walls	
	Primary	Secondary	Primary	Secondary	Primary	Secondary
Xyloglucan	20-25	Minor	2-5	Minor	10	-[b]
Glucuronoxylan	-	20-30	-	-	-	-
Glucuronoarabinoxylan	5	-	20-40	40-50	2	5-15
(Glucu)mannan	3-5	2-5	2	0-5	-	-
Galactoglucomannan	-	0-3	-	-	+[b]	10-30
β-(1→3, 1→4)-glucan	Absent	Absent	2-15	Minor	Absent	Absent

[a]Actual values vary between different species and tissue types. [b]-, absent or minor; +, present but data not available.

Xylan is widely distributed among cereal grains and hardwoods, and in lesser quantities in softwood (Stephen, 1983). It accounts for one third of renewable organic carbon available on Earth and constitutes the major component of hemicelluloses (Prade, 1996). The occurrence of homoxylan in higher plants is rather rare (Ebringerová and Heinze, 2000), with the majority of xylans being heteroxylans. Heteroxylans (hereafter referred to as xylans) of all higher plants are a diverse group of polysaccharides that shares as a common feature a backbone composed of β-1,4-linked D-xylosyl subunits that can be chemically modified (acetylation) and substituted at different position with various glycosyl residues (Aspinall,

1981). The decoration pattern correlates with taxonomy. Acetylation at C-3 and/or C-2 of xylosyl residues occurs in most xylans (especially in dicots), and the acetyl groups are numerous enough to avoid molecular aggregation (Katō, 1981). Glucuronoxylans (GX), substituted with α-1,2-linked 4-O-methyl glucuronosyl residues are often found in dicots, while glucuronoarabinoxylans (GAX), displaying decorations at O-2 and/or O-3 with L-arabinofuranose are mainly found in monocots (Dumon et al., 2012). In grass arabinoxylans, xylosyl residues are mostly substituted at O-3, while dicot arabinoxylans substitutions mostly occur at the O-2 position (Scheller and Ulvskov, 2010). Xylan in commelinid monocots is also characterized by the presence of ferulic acid esters, which are attached to the O-5 of some of the L-arabinosyl groups. This provides the basis for the presence of ferulate dehydro-dimers that help with intra- and intermolecular linkages (Grabber et al., 2000; Scheller and Ulvskov, 2010). Recent studies revealed several precise substitution pattern rules in xylan. For example, in gymnosperm xylan, every six xylosyl residues there is one methylated glucuronosyl and one arabinofuranosyl decoration, which are located precisely at the sixth and fourth xylosyl residue, respectively (Busse-Wicher et al., 2016).

A.1.3 Lignins

Lignins, the major non-polysaccharide polymeric components of PCWs, are generally thought to be amorphous and display high molecular weight. Lignins are a combination of three aromatic subunits, *p*-hydroxylphenyl (H), guaiacyl (G) and syringyl (S) respectively (Reale et al., 2004)(Fig A-2C) that are derived from *p*-coumaryl, coniferyl, and sinapyl alcohol respectively, which in turn are derived (in biosynthesis) from phenylalanine or tyrosine (Varner et al., 1989). Regarding the native *in muro* structure of lignin, this remains elusive, because all studies of lignins require their extraction and thus modification (Buranov and Mazza, 2008; Howard et al., 2003; You and Feng, 2016). However, a recent article has suggested that *in muro* lignins might not after all be high molecular weight polymers. Instead the authors propose that lignins are actually composed of linear oligomers of variable length that are covalently linked to cellulose and hemicellulose fibers (Banoub et al., 2015). Nevertheless, the major part of our current knowledge being based on the analysis of extracted lignins, the following information summarizes the state of the art.

The average degree of polymerization of lignins is around 13 to 20 (Vanholme et al., 2010). The composition of lignins varies among taxa. For example, softwood is predominantly made up of G units, whereas that of hardwood is abundant in both G and S units, which is also why

the former is more resistant to delignification than the latter (Baucher et al., 2003; Kishimoto et al., 2010). Various intermonomeric linkages have been found in lignins, such as ether and carbon-carbon bonds (Baucher et al., 2003). Regarding the role of lignins in PCWs, lignins confer stiffness, strength and hydrophobicity, the latter facilitating water and soluble nutrient transport as well as providing resistance to abiotic and biotic aggressions (Boerjan et al., 2003; Vanholme et al., 2010; Wyman and Kumar, 2017). Overall, it is widely considered that lignins constitute one of the major factors responsible for biomass recalcitrance and thus the economically-viable exploitation of biomass for industrial purposes.

A.1.4 Pectin

Pectin was originally identified in apples (in 1790) by the French chemist Louis Nicolas Vauquelin, although it was not until 1937 that Schneider and Bock established the basic formula of pectin (Mrak et al., 1962). Pectins are also complex polysaccharides, but compared to hemicelluloses they are more easily solubilized in hot water and dilute acid solutions, partly because their major components are 1,4 - linked α-D-galacturonic acid (Gal*p*A) residues (Cosgrove, 2005). The structural classes of the pectic polysaccharides include homogalacturonan (HG, ~65%), rhamnogalacturonan I (RG1, ~25%), rhamnogalacturonan II (RG2, ~10%), xylogalacturonan and others (Mohnen, 2008; Ridley et al., 2001) (Fig A-2D). Pectin is thought to be one of the most complex polysaccharide in biosphere (Cosgrove, 2005), which implied its multiplicity functions during plant growth, development, morphogenesis, defense and so on (Mohnen, 2008).

A.2 Structural organization of plant cell walls

PCWs are much thicker than the plasma membrane, ranging from 0.1 µm to several micrometers. The PCWs generally consist of three layers: the middle lamella (or more appropriately “intercellular substance” (Charles, 1901; Colvin, 1981)), the primary cell wall, and secondary cell wall that may not be necessary for some cells (Cosgrove, 2005) (Fig A-3A).

The middle lamella is the outermost layer and also the first layer laid down during cell division (Charles, 1901). It is initially rich in pectic polysaccharide and serves as an adhesive layer that glues cells together to prevent cell sliding under large stress (Cosgrove, 2005; Somerville et al., 2004). In the case of mature wood cells, the middle lamella is where most lignin deposited (Gibson, 2012; Mathews et al., 2015).

The primary cell wall is synthesized during cell expansion (Carpita and Gibeaut, 1993; Dhugga, 2007). It is composed of pectin compounds, hemicellulose, highly crystalline cellulose microfibrils that are oriented parallel to elongation direction (Caffall et al., 2009), and glycoproteins. Regarding the distribution of hemicelluloses, in dicots primary cell walls, xyloglucan accounts for 20 to 25% (w/w) (McCann and Carpita, 2015). In contrast, in grass primary cell walls, glucuronoarabinoxylan represent 20 to 40% (w/w) (Carpita and Gibeaut, 1993) (Table A-1). Primary cell walls provide mechanical and elastic strength, cell adhesion and defense response (Cosgrove, 2005). The composition of primary cell wall varies with time, and higher degree of lignification was found in mature cells than young ones (Gibson, 2012).

The secondary cell wall is a thick layer located between the plasma membrane and the primary cell wall (e.g. three sub layers in wood cell walls) that is deposited once cells have ceased to grow (Fig A-3A). It is composed of cellulose chains that display a higher molecular weight than those in primary cell walls, a fact that is linked to the structural (reinforcement) role of the secondary cell wall (Caffall et al., 2009; Cosgrove, 2005). Moreover, in secondary cell walls, the cellulose chains lie in antiparallel orientation with each other (Caffall et al., 2009) and hemicelluloses are more abundant than in primary cell walls (Table A-1). In dicots, glucuronoxylan constitutes 20-30% (w/w) of the secondary cell walls, while in grass glucuronoarabinoxylans comprise 40-50% (w/w) of the secondary cell walls. With age, the secondary cell wall becomes highly lignified fortifying the plant (Boerjan et al., 2003; Gibson, 2012).

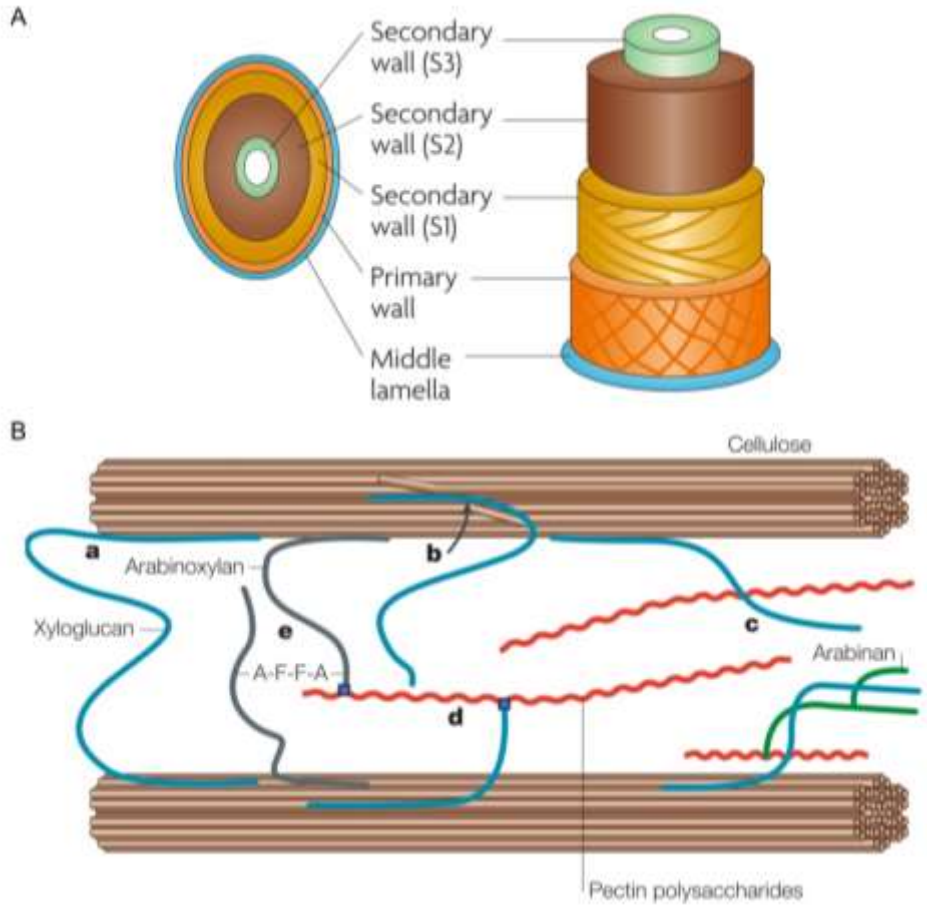


Fig A-3. The structure organization of plant cell wall.

(A) Typical structural organization in wood cell wall (Sticklen, 2008). (B) The interaction model in primary cell wall of dicots. a. hemicellulose bind (a) to the surfaces or entrapped into (Mccann et al., 1990)(b) of cellulose microfibrils and tighten two adjacent microfibrils by hemicellulose (a) or pectin polysaccharides (c). Macromolecules may be formed by hypothetical covalent bond (d) between xyloglucan and pectin polysaccharides and between arabinoxylan and ferulic acid esters (A-F-F-A)(e) (Cosgrove, 2005).

A.3 Interaction among the composition of plant cell wall

Various intra- and inter-molecular interactions between cell wall polymers help maintain fiber cohesion and plant cell wall integrity (Fig A-3B) (Collins et al., 2005). Hemicellulose chains can be hydrogen bonded to cellulose fibrils (Schädel et al., 2010)(Fig A-3B), a feature that is thought to provide robustness to the overall structure. Moreover, studies have also shown that covalent cross-linking helps to decrease stiffness, while increasing the extensibility (Chanliaud et al., 2004). In recent studies using molecular dynamics simulations, it has been demonstrated that α-1,2-linked decorations in xylan not only enhance the flexibility of the threefold screw conformation of xylan, but also help stabilize xylan binding to the cellulose surface (Berglund et al., 2017; Pereira et al., 2017). Ferulic acid ester substitutions on xylan can crosslink (forming dehydrodimers) that provide xylan-xylan and xylan-lignin interactions (Grabber, 2005; Grabber et al., 2000). Lignin is always linked to

other matrices, especially hemicellulose, via ester and ether bonds (Buranov and Mazza, 2008; Davison et al., 2013), forming an impenetrable barrier that prevents the penetration of external substances (Howard et al., 2003), which explains why the lignin is difficult to exract and isolate.

A.4 Comparison of different plant cell walls

Although many common biochemical and physical features are observed among different PCWs, it is clear that not all share exactly the same chemical composition and organizations (Table A-2). Therefore, understanding how these differences affect PCW structure and stability is a vital step towards the development of processing routes that can better surmount PCW recalcitrance (DeMartini et al., 2013b).

Primary and secondary cell walls are not clearly separated in Charophytes and Bryophytes (Sarkar et al., 2009). Charophytes are recognized as one of the earliest plant groups to appear on land. Their PCWs were designed to make them more adapted to terrestrial life, for example (1) the formation of microfibrils that can support their bigger bodies compared to those of more primitive algae; (2) the abundance in uronic acids confers Charophytes with the ability to defend against desiccation and microbial attacks. Bryophyte cell walls are evolved to be more rigid by introducing XG and pectin.

Following the emergence of lignified secondary cell walls in Leptosporangiates, Gymnosperm cell walls became reinforced in cellulose, fucoside XG, HG and RG. Together these confer plants with the ability to withstand the load of their tall woody bodies (Table A-2). However, the lignin in Leptosporangiates and Gymnosperms cell walls contain only guaiacyl units, while the lignin in angiosperms contain both guaiacyl and syringyl units, a composition that correlates with easier delignification (Kishimoto et al., 2010), although different opinions may exist on this point (DeMartini et al., 2013b). Grass cell walls are not particularly well adapted to support tall woody bodies, since they contain low levels of XG and pectins. Instead, the cell walls of grassy species contain higher amounts of mixed-linkage glucans (Sarkar et al., 2009).

Table A-2. Comparison of composition of plant cell wall in different groups of kingdom Plantae (Sarkar et al., 2009)

	hemicellulose	primary cell wall	secondary cell wall
Non-grass angiosperm	high XG, xylan and some mannan and structural proteins	high HG, RG I and II	high lignin with guaiacyl and syringyl units

Grass	low XG, high amount of mixed-linkage glucans, xylan	low HG, RG	high lignin with guaiacyl and syringyl units
Gymnosperm	high XG, xylan glucomannan and structural proteins	high HG, RG I and II	high lignin with guaiacyl units
Leptosporangiate	low XG, high xylan, mannan, uronic acids, 3-O-methylrhamnose	low HG, RG	high lignin with guaiacyl units
Eusporangiate and bryophyte	low XG, high mannan, uronic acids, 3-O-methylrhamnose	low HG, RG, phenolic compounds instead of lignin	-
Charophyte	Only mannan, glucuronic acid, mannuronic acid and 3-O-methyl rhamnose have been detected.	-	-

Table A-3. Major agricultural wastes and their worldwide availability, bioethanol production potential and compositions (Sarkar et al., 2012)

		Agrowaste			
		Rice straw	Wheat straw	Corn straw	Bagasse
Global production	Availability (million tons)	731.3	354.34	128.02	180.73
	Bioethanol potential (Gl)[a]	205	104	58.6	51.3
Polymer / dry matter	Cellulose (%)	32-47	35-45	42.6	65
	Hemicellulose (%)	19-27	20-30	21.3	
	Lignin (%)	5-24	8-15	8.2	18.4
monomer / dry matter	Glucose (%)	41-43.4	38.8 ± 0.5	39	38.1
	Xylose (%)	14.8-20.2	22.2 ± 0.3	14.8	23.3
	Mannose (%)	1.8	1.7 ± 0.2	0.3	-
	Galactose (%)	0.4	2.7 ± 0.1	0.8	1.1
	Arabinose (%)	2.7-4.5	4.7 ± 0.1	3.2	2.5

[a] potential annual bioethanol global production, Gl, giga liter.

A.5 Plant cell wall in the context of biofuels

The development of second generation biofuels is considered preferable to first generation fuels, because the former can be produced using non-edible, lignocellulosic biomass (Table A-3) (Rakshit, 2013; Wyman and Kumar, 2017). Lignocellulosic biomass can be sourced in many ways, either using so-called energy crops (e.g. switchgrass), wood or agricultural residues (e.g. wheat bran, wheat straw) (Saini et al., 2015). Many of these lignocellulosic biomass feedstocks arise from angiosperms, which means that their chemical composition and structural organization are complex and thus their conversion into products such as biofuels is difficult to achieve (Lynd et al., 2017) (Table A-3). As an alternative, it is possible to focus on

algal feedstocks, using the primitive cell walls of Charophytes as the source of carbon for 3rd generation biofuels, such as biodiesel (Menetrez, 2012).

A.6 Plant cell wall in the context of diet-microbiota interactions

PCWs from grass (family Poaceae) are the primary source of dietary fibers (i.e. non-starch polysaccharides consumed by human, cow and some insects). Dietary fibers are important, because it is known that they play an important role in the crosstalk between host and gut microbita, a current topic of great interest (Barratt et al., 2017; Cani and Everard, 2016; Koropatkin et al., 2012). Arabinoxylan, pectin, and XG are classified as soluble dietary fibers, whereas cellulose and lignin are grouped into insoluble dietary fibers (Williams et al., 2017). Cereal grains such as wheat and rye are rich in arabinoxylans for human nutrition and health (Andersson et al., 2009; Selvendran, 1984). A whole-grain diet is generally considered beneficial for human health (Gong et al., 2018; Topping, 2007), consistent with the fact that the amount of arabinoxylan in wheat bran is much more than in flour (20 and 2% of dry weight, respectively) (Kamp, 2010).

PART B ENZYMATIC DECONSTRUCTION OF LIGNOCELLULOSE

The degradation of lignocellulose embedded in the plant cell wall leads to the recycling of photosynthetically fixed carbon and is therefore of profound biological and industrial importance. Due to the broad range recalcitrance of lignocellulosic biomass, merely one step or one method cannot achieve full degradation (McCann and Carpita, 2015). Nevertheless, in this regard, the industrial exploitation of enzymes is advantageous, because they perform specific reactions with low energy costs (temperature) and generate hydrolysis products that are generally innocuous both with regard to fermentation microbes and the environment (Alvira et al., 2013).

B.1 Lignocellulose-degrading enzymes

B.1.1 Classification

Within the International Union of Biochemistry and Molecular Biology (IUB)'s Enzyme Nomenclature and Classification system, lignocellulose-degrading enzymes are classified according to the reaction they catalyze and their substrate specificities (http://www.chem.qmul.ac.uk/iubmb/enzyme/). Since the majority of lignocellulose-

degrading enzymes catalyze hydrolysis reactions, these are mostly classified in EC3 (Sweeney and Xu, 2012), although some lignocellulose-degrading enzymes are oxidases or lyases.

An alternative classification system has been proposed by Bernard Henrissat and his colleagues (Henrissat et al., 1991). This scheme, which is focused on carbohydrate-active hydrolases and their associated modules[1], organizes enzymes into families on the basis of amino acid sequence similarities. The data regarding these families is archived in the database named Carbohydrate-Active enZymes (CAZy). This database is an extremely valuable tool to advance understanding of structure/function relationships in carbohydrate-active enzymes ((Cantarel et al., 2009), http://www.cazy.org/). Presently, it includes glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), redox enzymes and carbohydrate-binding modules (CBMs). Members of any given CAZy family share a common fold, catalytic apparatus and mechanism, although substrate specificity is not necessarily conserved. A "clan" is a group of families that are thought to share common ancestry and are recognized by significant similarities in tertiary structure, together with conservation of the catalytic residues and catalytic mechanism (Henrissat and Bairoch, 1996). Glycosyl hydrolase families are grouped into 5 clans, GH-A, GH-B, GH-C, GH-D and GH-E. Enzymes that work on the same substrate are sometimes found in different clans. For example, although GH10 and GH11 both act on xylans, GH10 belongs to clan A while GH11 belongs to clan C. Conversely, enzymes belonging to the same family and/or the same clan can act on different substrates. This is exemplified by members of families GH10 and GH53 (both belong to clan A) that act on xylan and galactan respectively. The clan concept demonstrates a broader relationship between GHs, which suggests a more distant common evolutionary ancestor (Pollet et al., 2010). Recently, omics studies have greatly increased the number of CAZy families and enriched some families with a large number of new members. This is the case for GH 5, 13, 30 and 43 families that display a wide variety of substrate specificities, with several specificities being associated with single families. Consequently, in these large families it has been possible to divide them into subfamilies on the basis of substrate specificity (Aspeborg et al., 2012; Mewis et al., 2016; Stam et al., 2006).

[1] More recently, the CAZy database has included other enzymes that also contribute to the degradation of PCWs. These include polysaccharide lyases and redox enzymes (classed under the term 'Auxiliary Activities') that act in conjunction with CAZy hydrolases.

Taken together, the CAZy classification is more valuable for bioinformatics study than IUB system.

B.1.2 Catalytic mechanism

B.1.2.1 The retaining and reverting mechanism

GHs are enzymes that catalyze the hydrolysis of the glycosidic bonds of glycosides, glycans and glycoconjugates, producing a sugar hemiacetal or hemiketal and the corresponding free aglycon. Whether the anomeric configuration of the final product is identical to that of the reactant or not, provides the basis to classify GHs into two groups known as retaining or inverting GHs (Dies et al., 1995; Koshland, 1953)(*Fig B-1*). Catalysis in both cases is mostly due to the presence of two acidic (Asp or Glu) amino acids located in the active site. In retaining enzymes these act as a proton donor and a nucleophile respectively, whereas in inverting enzymes one acts as a general acid and the other as a general base. For any given enzyme, one can distinguish between the two mechanisms by close examination of structural and functional properties (Vuong and Wilson, 2010) and in either case mutation of the two catalytic residues to inert amino acids (Ala, Gly, *etc*.) essentially deactivates the enzyme (MacLeod et al., 1996).

In inverting GHs (*Fig B-1*), catalysis occurs via a two-pronged attack on the anomeric center. The catalytic acid residue donates a proton to the anomeric oxygen of the donor substrate, while the catalytic base concomitantly attacks the anomeric carbon from opposite side. This causes anomeric inversion, meaning that for example a β-glycoside donor yields an α-glycoside product.

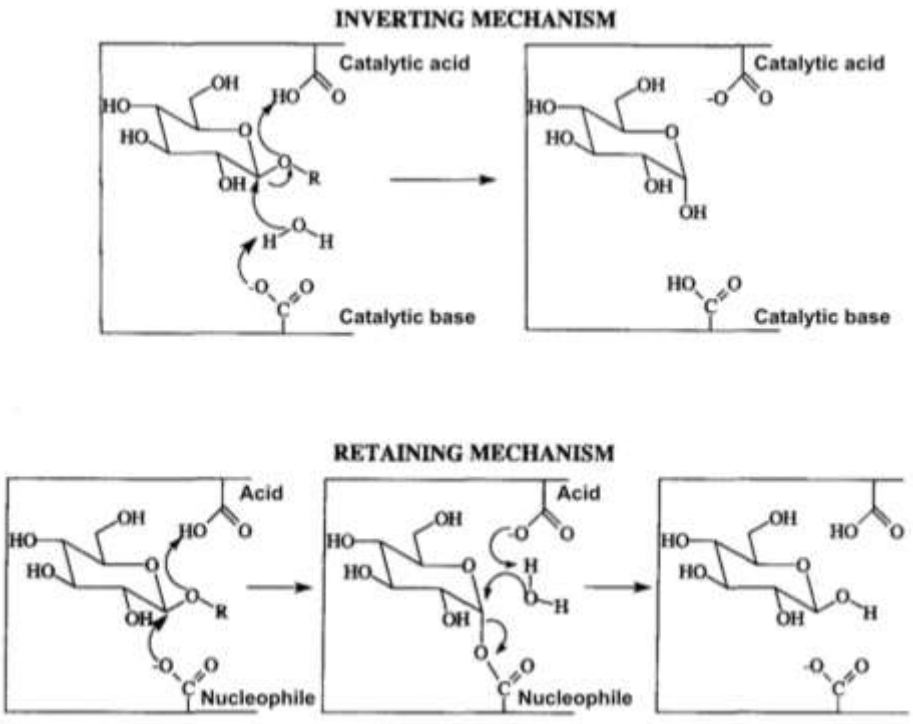


Fig B-1. The presumed inverting and retaining mechanism (Mccarter and Withers, 1994)

In retaining GHs (*Fig B-1*), the reaction occurs in two distinct steps. First, the proton donor (acid) transfers a proton to the donor substrate. Concomitantly, the catalytic nucleophile attacks the anomeric center. Together these bring about bond breakage, the liberation of a leaving group and the formation of a glycosyl-enzyme covalent intermediate that displays inverted anomeric configuration. In a second step, the deprotonated catalytic residue activates an incoming water molecule, which consequently adopts nucleophilic character and attacks the anomeric carbon of the glycosyl-enzyme. This leads to the breakdown of the covalent intermediate, a second inversion of the anomeric configuration and protonation of the acid/base catalyst. The dual role of acid/base residue is probably accomplished by so called "pKa cycling" (McIntosh et al., 1996) based on "upward"(base, deprotonated, low pKa) and "downward"(acid, protonated, high pKa) conformation switch of Glu (Wan et al., 2015). Overall, at the end of the reaction the anomeric configuration of the glycoside product is finally identical to that of the substrate. Therefore, careful analysis of the anomeric configuration of the reaction end products reveals whether any given GH operates via an inverting or a retaining mechanism.

From a structural standpoint, the nature of the retaining reaction means that the two catalytic residues are separated by a distance of ~5.5 Å (mean value calculated for many enzymes),

whereas in inverting GHs this distance is larger (~10 Å) (Dies et al., 1995). This is because inverting enzymes need to simultaneously accommodate the sugar donor and the incoming water molecule.

Despite the fact that the canonical double displacement mechanism has been experimentally validated (Davies et al., 1998; Suzuki et al., 2009), this mechanism has been extensively debated and other schemes, for example one involving an oxocarbenium ion intermediate, have been proposed (Phillips, 1967) and experimentally supported (Hövel et al., 2003; Tanakas et al., 1994). In fact the main difference between the two mechanistic schemes is the nature of the transition state: either it is a covalent glycosyl-enzyme intermediate (S_N2 reaction) or an oxocarbenium ion-like transition state without assistance of nucleophile (S_N1 reaction). Other complex transition states and intermediate structures are also mentioned in the literature (Sabini et al., 1999).

B.1.2.2 Sugar binding subsite topologies

The structural analysis of GHs generally reveals that the two catalytic residues are usually located somewhere within a pocket, which intimately accommodates the sugar donor substrate. This pocket or subsite is often numbered as the subsite -1, meaning that it is the subsite that accommodates the glycone group (sugar donor) that lies on the non-reducing side of the scissile bond. The subsite accommodating the sugar (or an aglycone group) on the other side of the bond is thus designated subsite +1. In enzymes possessing extended active sites, the other subsites are numbered accordingly with all subsites on the non-reducing side of the bond being negatively numbered (Davies et al., 1997; Pell et al., 2004a)(Fig B-2A). In GHs, subsites are generally structured to favor sugar binding via the formation of hydrogen bonds or hydrophobic stacking interactions. Accordingly, they play a crucial role in substrate recognition (Pollet et al., 2010).

Regarding the structural configuration of the active sites in GHs, it is convenient to use the term active site topology. This term refers to the specific 3D conformation and arrangement of the subsites that form an active site, and which specifically accommodate the glycoside substrate. Three common active-site topologies are described as pockets or craters, clefts or grooves and tunnels (Fig B-2).

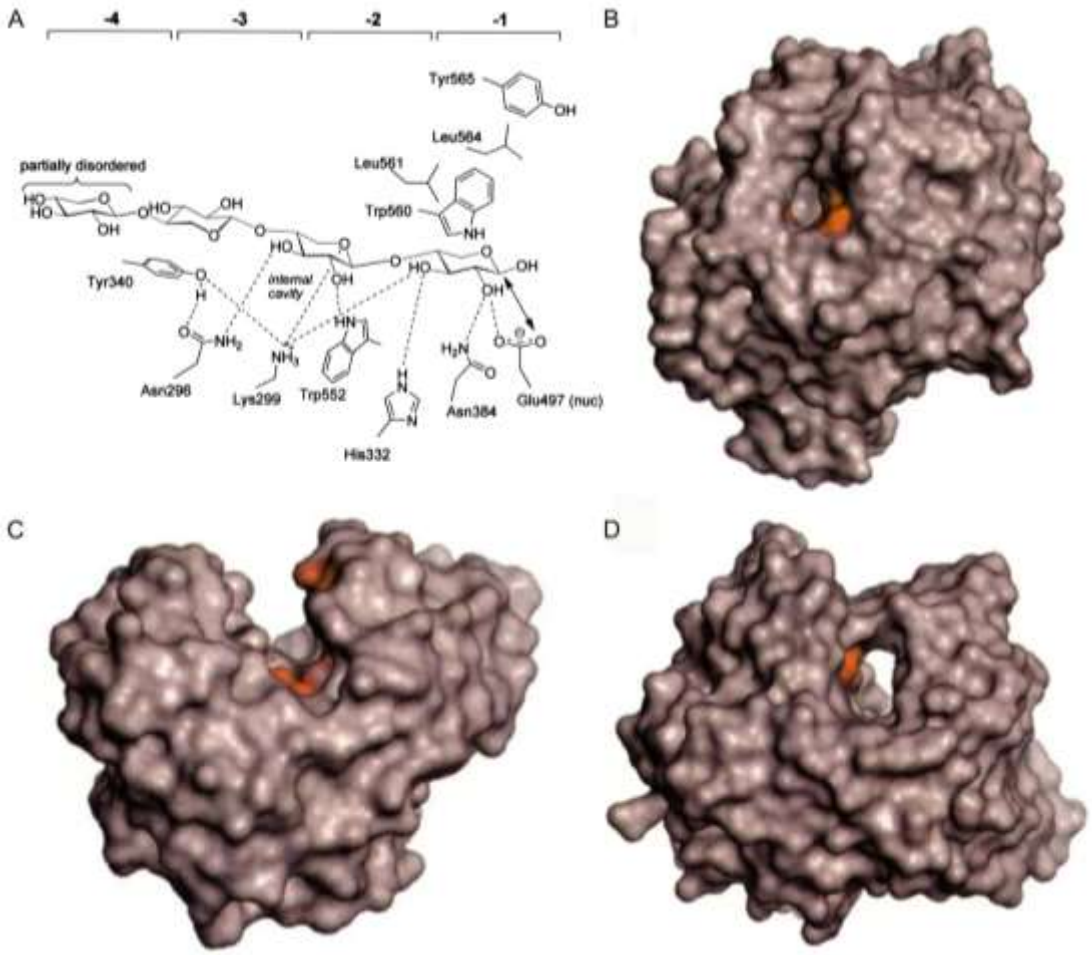


Fig B-2. Sugar binding subsite topologies in glycosyl hydrolases.

(A) Glycon subsite of *Cellvibrio japonicas* Xylanase 10C (Pell et al., 2004b). Three types of active site topologies (B) The pocket, PDB: 3GLY, glucoamylase from Aspergillus awamori; (C) The cleft, PDB: 1TML, endoglucanase E2 from Thermononospora fusca; (D) The tunnel, PDB: 3CBH, cellubiohydrolase II from Trichoderma reesei. The red zone is the proposed catalytic residues (Dies et al., 1995).

Enzymes with "pocket" shaped topology are usually exo-acting and act on non-reducing extremities of saccharides. Radial structural substrate is more preferable than fibrous ones. Examples of pocket enzymes are β-amylase, β-glucosidase. The enzymes possessing an open "cleft" are more adapted to cleave long polymers to short oligomers, such as endocellulases, endoxylanases. "Tunnel" topology is found in cellobiohydrolases and facilitate the release of product while cleavage, i.e. processivity (Dies et al., 1995).

B.1.3 Cellulases

Five types of cellulases are commonly identified. Most cellulases are endo-1,4- β -D-glucanases or 1,4- β- D -glucan-4-glucanohydrolases (EC 3.2.1.4), which have an open active site (Fig B-2B, e.g. PDB: 1TML)(Fischer et al., 2013) that binds a cellulose glucan chains at any accessible point along its length. Endo-1,4- β- D -glucanases then perform apparently

random cleavage and then dissociate from the chain. Archetypical endoglucanases (EG) are grouped in families GH 6, 9, 45 and 58, all of these being inverting enzymes. Other archetypical EGs are found within families GH 5, 7 and 12 which contain enzymes that operate via a retaining mechanism.

Unlike EGs, exocellulases are characterized by an active site that displays tunnel-like topology (Fig B-2C). Consequently, cellulose glucan chains bind at one extremity of the tunnel and are then fed through in a processive manner, with cleavage occurring at regular repeated intervals until substrate dissociation occurs. This leads to the release of short glucooligosaccharides (or cellodextrins), with cellubiose being a typical product (Kristensen, 2009). The action of exocellulase requires free glucan extremities and thus generally follows the action of EGs, which generate such extremities. There are two types of exocellulases, 1,4-β- D-glucan glucanohydrolases (also known as cellodextrinases, such as GH7, EC3.2.1.74) and 1,4- β- D-glucan cellobiohydrolases (cellobiohydrolases, such as GH6, EC 3.2.1.91). The former act on reducing ends of cellulose glucan chains, while the latter that attacks non-reducing ends. The co-occurrence and synergism between endo- and exo- cellulases are prerequisites for high efficient enzymatic systems for biorefineries (Sweeney and Xu, 2012).

A third class of cellulases is β-glucosidases or β-glucoside glucohydrolases (EC 3.2.1.21), which hydrolyze the exocellulase product (i.e. cellobiose) into individual monosaccharides. The representative families are GH1 and 3. Unlike the majority of lignocellulases, β-glucosidases act upon soluble rather than insoluble substrate. In biorefinery applications where high substrate concentrations are desirable (>20% dry matter content) to achieve sufficiently high end product titers, β-glucosidase (especially GH1) is particularly useful to mitigate the inhibition of endo- and exo-cellulases by cellodextins.

Oxidative cellulases, for instance cellobiose dehydrogenase (acceptor), depolymerize cellulose by radical reactions, Finally, cellulose phosphorylases depolymerize cellulose using phosphates instead of water (Lynd et al., 2002; Sweeney and Xu, 2012).

The GH9 family

GH9, formerly known as cellulase family E, is mainly composed of members that possess endoglucanase (EG) activity and operate by random cleavage of β-1,4 glycosidic bonds. Their role is to reduce the degree of polymerization of the glucan chains and provide substrates for the exo-cellulases. The catalytic core of GH9 EGs displays a (α/α6-fold) and is characterized by the presence of two sequence motifs Asp-Ala-Gly-Asp and Asn-Glu-Val-Ala, which

contain the catalytic nucleophile (an Asp residue) and the proton donor (Glu residue) respectively. Studies showed that the mutation on Gly147 to Arg in the GH9 EG from the termite *Reticulitemes speratus* led to complete loss of enzymatic activity, suggesting that Gly147 might play a significant role in maintaining enzyme activity (Ni et al., 2010). More generally, GH9 endo- β -1,4-glucanases are common among different invertebrates such as beetles, flies, cockroaches, and termites (Bujang et al., 2014).

B.1.4 Hemicellulases

The degradation of hemicelluloses has several effects on biomass. First, since hemicelluloses in plant cell walls are in intimate contact with cellulose microfibrils, the action of hemicellulases unmasks the microfibrils and opens up access for cellulases. Second, hemicellulases also complete hydrolyze hemicelluloses making the sugars available for metabolism by microorganisms. Since hemicelluloses are chemically and structurally complex, their degradation relies upon the interplay of a whole range of enzymes (Table B-1) that can be collectively designated using the term hemicelluases (Sweeney and Xu, 2012).

Table B-1 The major hemicellulases and their classification

Enzymes	Substrates	EC number	Family
Exo-β-1,4-xylosidase	β-1,4-Xylooligomers xylobiose	3.2.1.37	GH 3 39 43 52 54
Endo- β-1,4 xylanase	β-1,4-Xylan	3.2.1.8	GH 5 8 10 11 43
Exo- β-1,4-mannosidase	β-1,4-Mannooligomers mannobiose	3.2.1.25	GH 1 2 5
Endo- β-1,4-mannanase	β-1,4-Mannan	3.2.1.78	GH 5 26
Endo-α-1,5-arabinanase	α-1,5-Arabinan	3.2.1.99	GH 43
α – L-arabinofuranosidase	α-Arabinofuranosyl (1→ 2) or (1→ 3) xylooligomers α-1,5-arabinan	3.2.1.55	GH 3 43 51 54 62
α-Glucuronidase	4-O-Methyl- α-glucuronic acid (1→2) xylooligomers	3.2.1.139	GH 67
α-Galatosidase	α-Galactopyranose (1→6) mannooligomers	3.2.1.22	GH 4 27 36 57
Endo-galactanase	β-1,4-Galactan	3.2.1.89	GH 53
β-Glucosidase	β-Glucopyranose (1→4) mannopyranose	3.2.1.21	GH 1 3
Acetyl xylan esterases	2-or 3-O-Acetyl xylan	3.1.1.72	CE 1 2 3 4 5 6
Acetyl mannan esterase	2-or 3-O-Acetyl mannan	3.1.1.6	CE 1
Ferulic and p-cumaric acid esterases	2-or 3-O-Acetyl mannan	3.2.1.73	CE 1

B.1.4.1 Xylanases

Owing to the heterogeneous structure of xylans, their degradation requires the intervention of a large number of enzymes (Fig B-3). The key enzyme activity responsible for the hydrolysis of heteroxylans is endo-1,4-β-xylanase (EC 3.2.1.8) that are commonly referred to as xylanases. These are produced by a wide variety of organisms, incuding fungi, bacteria, yeast, marine algae, protozoa, snails, crustaceans, insect, sees, etc (Sunna and Antranikian, 1997). Xylanases act on the β-1,4 bonds that link internal xylosyl residues in xylan backbones. The action of xylanases is completed by that of 1,4-β-D-xylosidase (EC 3.2.1.37), which are commonly referred to as xylosidases. These preferentially hydrolyze xylooligosaccharides (XOS), which are the products of xylanase action. Although it is often assumed that xylosidases preferentially act upon xylobiose, this is not completely true since some xylosidases clearly display preference for longer XOS. However, unlike xylanases,

xylosidases are exo-acting enzymes that hydrolyse substrates by removing β- D-xylopryranosyl residues from the non-reducing termini of XOS.

Finally, to be exhaustive, it is noteworthy that some xylanases are endo-1,3- β-xylanases that catalyze the hydrolysis of β-1,3-xylan, which is typically found in marine algal species (Dumon et al., 2012).

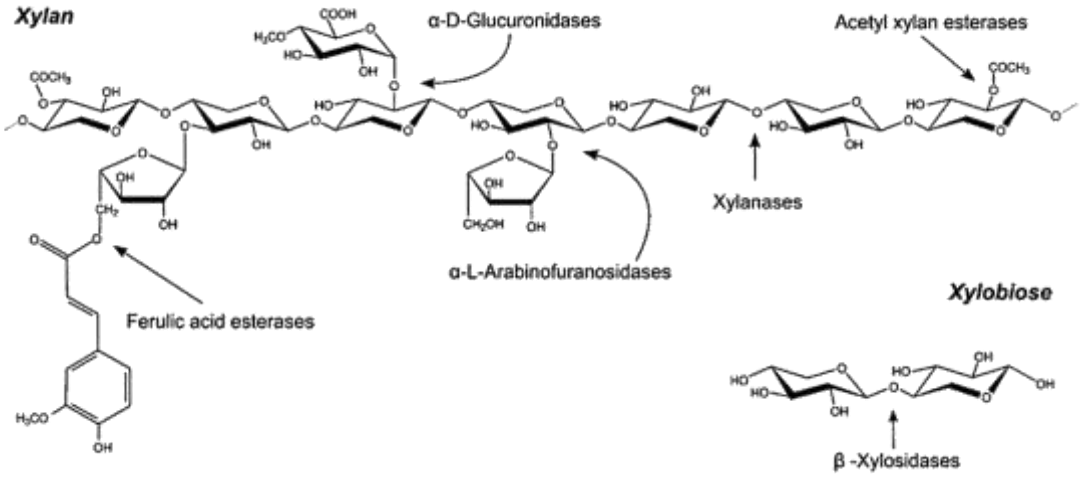


Fig B-3. Structure of xylan and the sites of attack by xylanolytic enzymes (Salis et al., 2007)

B.1.4.1.1 Endo-1,4- β-xylanase

Xylanases are key enzymes, because they depolymerize xylans and generate products that serve as substrates for most of the other hemicellulases (Collins et al., 2005). Most known endoxylanases belong to GH families 10 and 11 (formerly family F and G, respectively, (NR.Gilkes et al., 1991)) (over 4000 gene sequences are known), although xylanases are also found in families 5, 7, 8, 9, 12, 16, 26, 30, 43, 44, 51 and 62 (www.cazy.org). However, it is important to stress that xylanases belonging to families GH 16 and 62 are bifunctional enzymes that possess two distinct catalytic domains. From the perspective of evolution, it is also noteworthy that family 5 is related to the archetypal xylanase family GH10, while family GH12 is related to the equally archetypal xylanase family GH11. Indeed, in the case of the latter pair, it has been suggested that GH11 members possibly result from the evolutionary specialization of family GH12 (Collins et al., 2005). One manifestation of this specialization is the presence in GH11 enzymes of a large flexible loop that probably confers xylan specificity and catalytic potency, which is absent in GH10 enzymes (Paës et al., 2012; Polizeli et al., 2005).

Family GH10 is populated by enzymes of plant, fungal, and bacterial origin, whereas GH11 family arise from fungi and bacteria only (www.cazy.org). Moreover, the majority of bacterial xylanases are GH10 family members, whereas most fungal xylanases belong to GH11 family. Due to the general relationship between molecular weight (MW) and isoelectric point (pI)(Wong et al., 1988), most of the GH 10 xylanases display a MW ≥30 kDa, low pI, while GH11 members generally display a MW <30 kDa and high pI (Sunna and Antranikian, 1997). Moreover, GH11 enzymes are active across a wide pH range from from 2 to 11, with GH11 subgroups being classified as acidic and alkaline xylanases according to their pH optima (Joshi et al., 2000; Sapag et al., 2002). Like enzymes belonging to GH families 1, 2, and 5, the catalytic domain of GH10 endoxylanases is composed of a TIM-barrel fold, unlike that of GH11 counterparts that display a β-jelly roll. Nevertheless, despite these difference, both GH 10 and GH 11 xylanases perform catalysis via a retaining mechanism that requires a catalytic dyad composed of a catalytic nucleophile and a proton donor (Gebler et al., 1992; Withers et al., 1986).

Regarding their substrate specificity, GH10 xylanases are generally considered to be more catalytic versatile than GH11 xylanases, often displaying activity cellulose and aryl β-D-cellubiosides (Tull and Withers, 1994). When acting on branchy xylans, it has been shown that GH10 xylanases require the presence of one or two unsubstituted xylopyranosyl residues between decorated residues, corresponding to subsites -2 and -1 (Biely et al., 1997; Kolenová et al., 2006; Pell et al., 2004a). In the case of GH11 xylanases, it has been shown that these require three unsubstituted consecutive xylopyranosyl residues accommodating from subsites -2 to +1 (Biely et al., 1997; Katapodis et al., 2003). As a consequence, GH10 xylanase can produce smaller hydrolysis products from glucuronoxylan and arabinoxylan than GH11 xylanases (Pollet et al., 2010).

B.1.4.1.2 GH10 enzyme: Topology of catalytic site and its polymorphism

At the time of writing (December 2017), the structures of 29 GH10 xylanases have been resolved and publically released, thus providing precious structure-sequence information (Fig B-4).The active site of GH10 xylanases is commonly described by an open cleft in which the substrate can bind. The cleft is lined by a series of amino acid sidechains, with several being aromatic residues. The catalytic acid-base (proton donor) and nucleophile resides are located

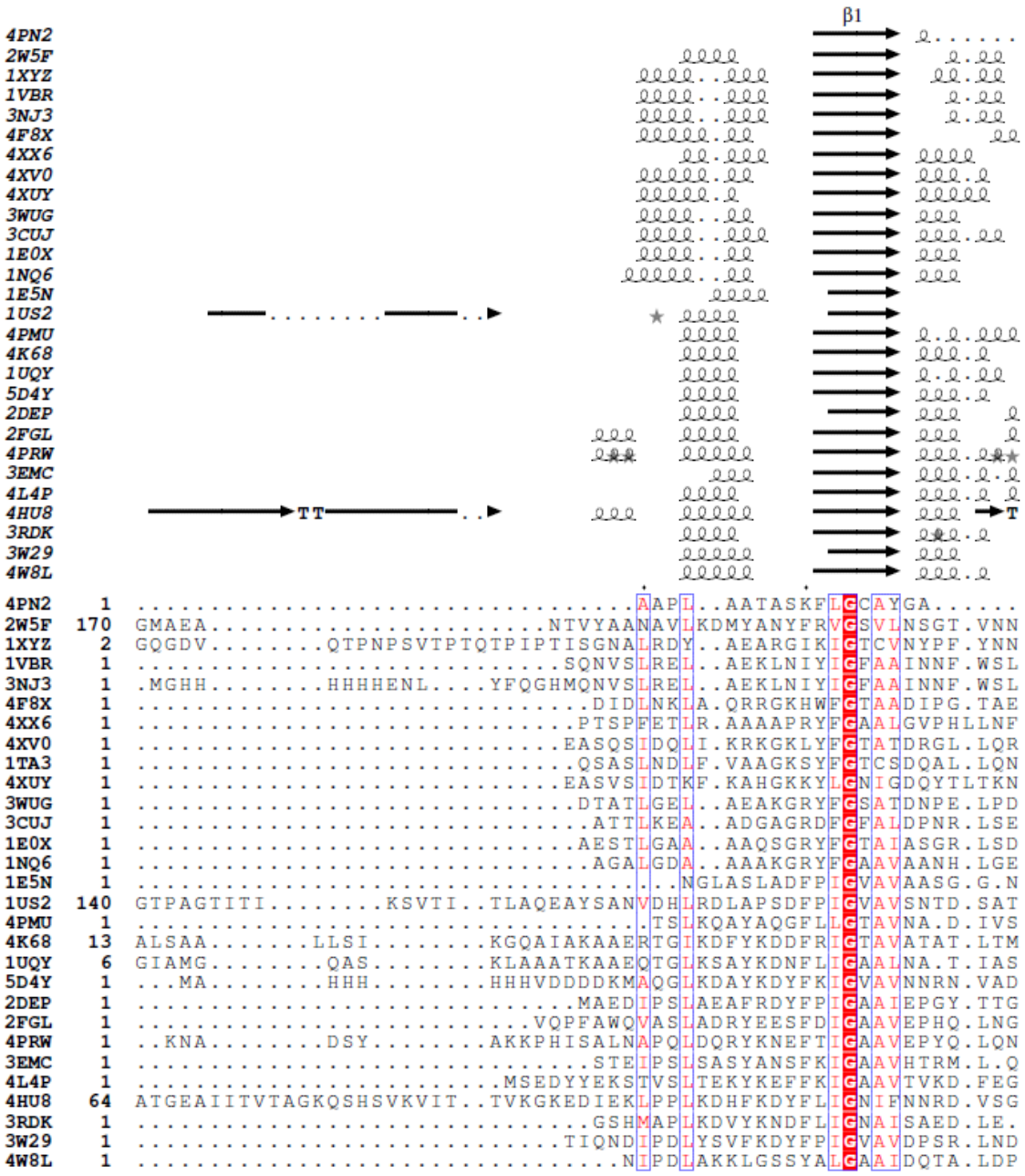
on loops at the end of β sheets 4 and 7, which explains why GH10 enzymes are assigned to the so-called 4/7 superfamily (Jenkins et al., 1995).

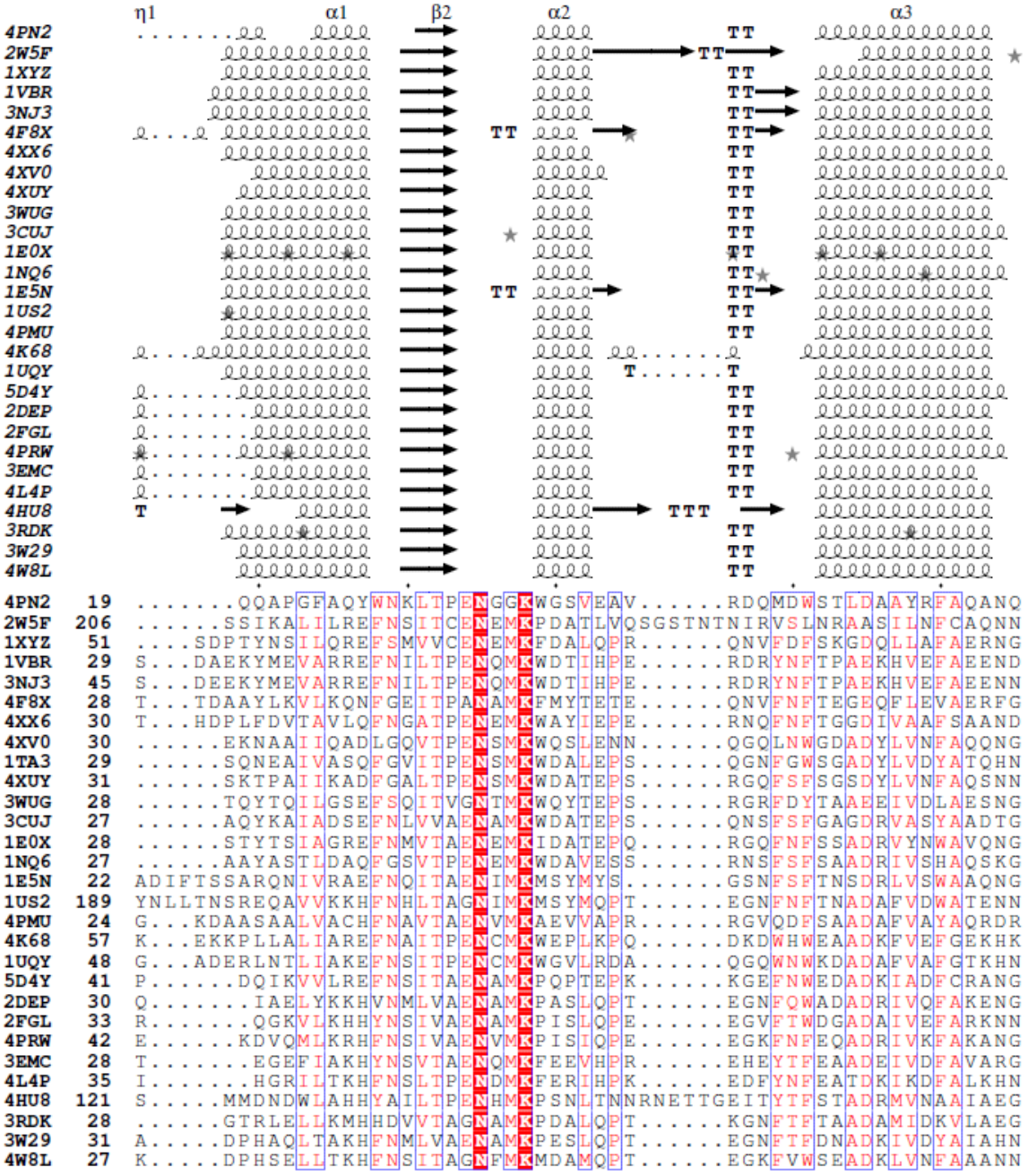
In the glycon region, xylose residues are bound through several hydrogen bonds. In contrast, ligand binding in the aglycon region of the substrate-binding cleft is dominated by hydrophobic interactions between xylopyranose rings and aromatic amino acids (Charnock et al., 1999; Zolotnitsky et al., 2004).

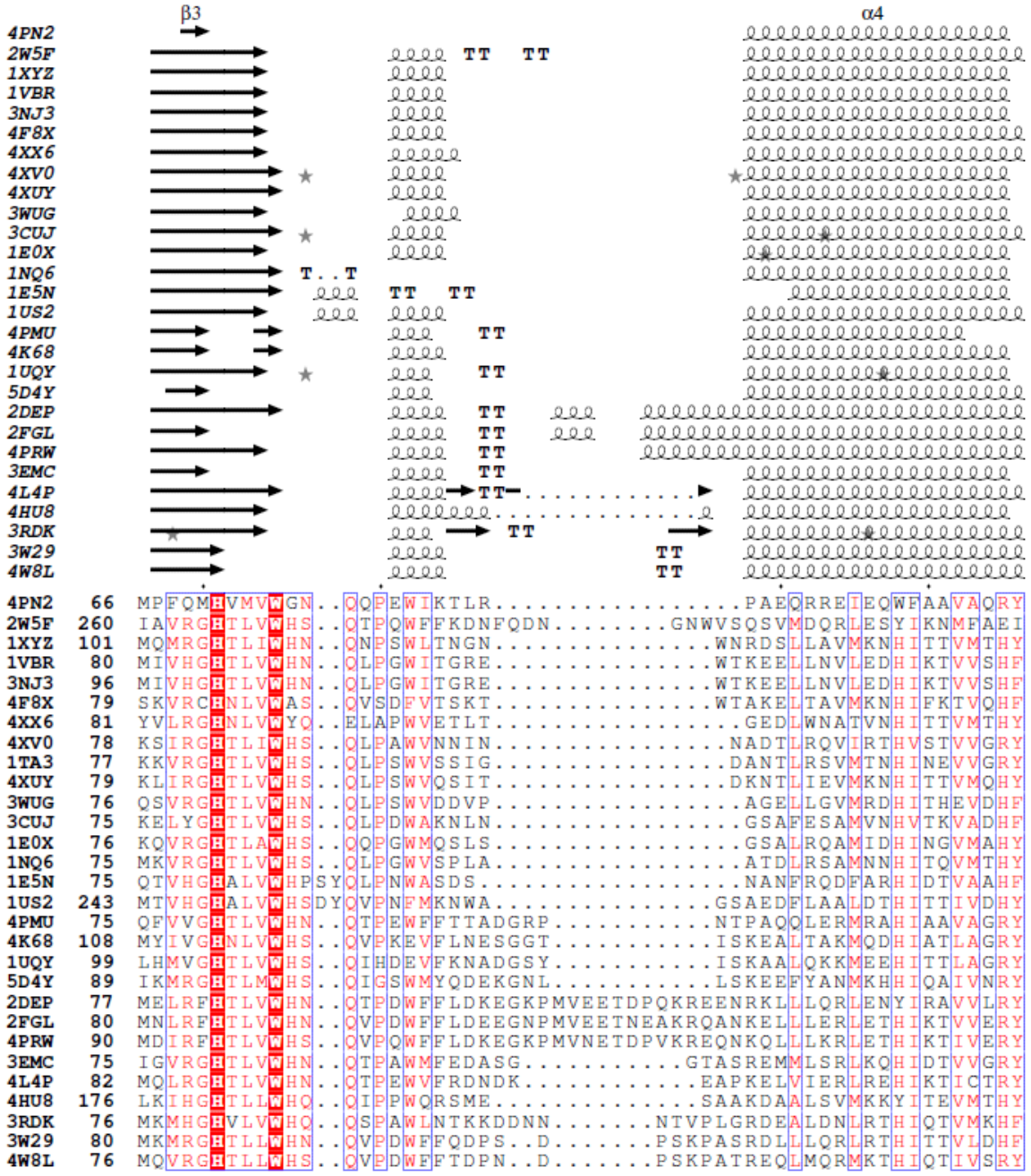
The crystal structures of GH10 xylanase in complex with oligosaccharides and inhibitors indicate that, in general, the -2, -1, and +1 subsites are highly conserved (Charnock et al., 2000; Notenboom et al., 1998; White et al., 1996). Amino acid residues in these subsites play a key role in subsite recognition and make several interactions between the enzyme and the xylose moieties.

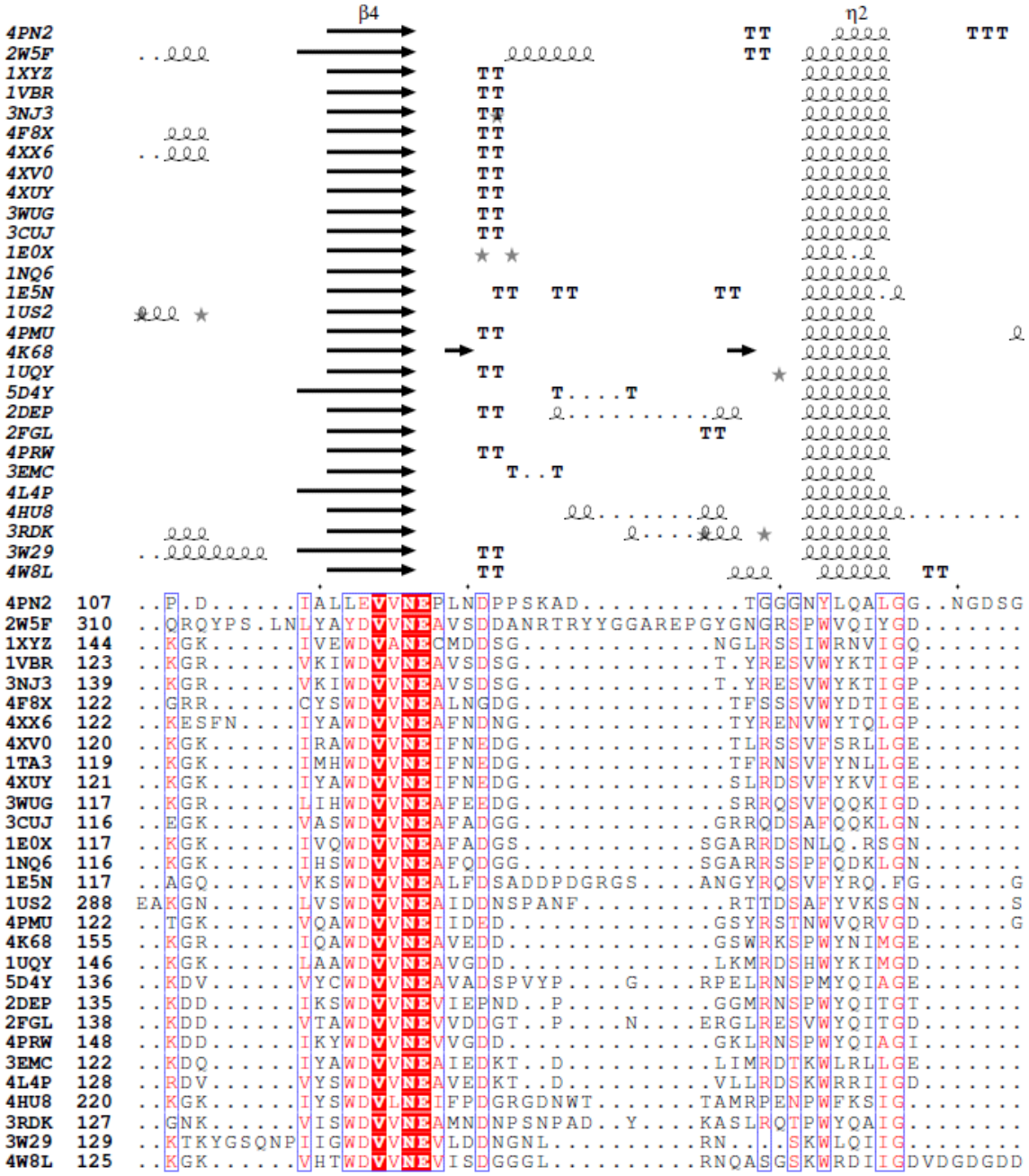
At the -2 subsite, the O-2 of the sugar hydrogen bonding to a tryptophan located in β8, the O-3 to an asparagine in the loop connecting β3 and α3, and the endocyclic oxygen to a lysine in the same loop, too. It has been shown that -2 subsite is prerequisite for cleavage branched xylan (Kolenová et al., 2006). At the -1 subsite the O-3 of xylose makes a hydrogen bond with a highly conserved histidine in β6, whereas O-2 interacts with an asparagine and a glutamate (which acts as the catalytic nucleophile) that located in the end of β4 (Charnock et al., 1999; Harris et al., 1994; Pell et al., 2004a).

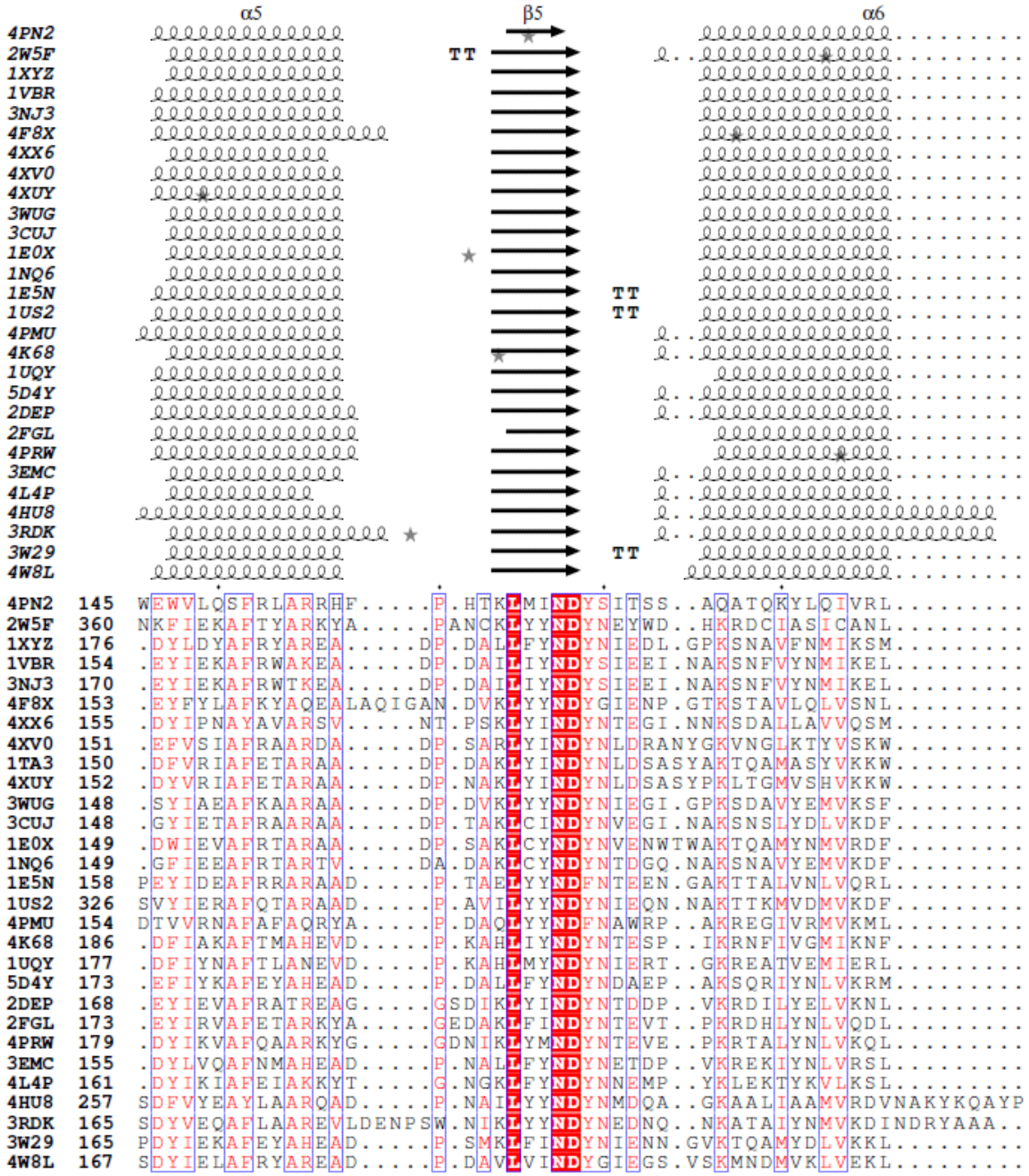
Generally, xylohexaose was mainly degraded into xylotriose by endoxylanase while xylopentaose was cleavage into xylobiose and xylotriose. In contrast, xylotetraose can be cleaved into primarily two xylobioses (Charnock et al., 1999; Hong et al., 2014; Millward-Sadler et al., 1995; Pell et al., 2004a; Sun et al., 2007), or xylose and xylotriose (Hong et al., 2014; Millward-Sadler et al., 1995; Pell et al., 2004b). The different product profile implied that the -3 subsite is not conserved in GH10 xylanases. At the -3 subsite, the aromatic ring of tyrosine stacks against the xylopyranose ring in *Cj*Xyn10C, *Cj*Xyn10A, *Cellvibrio mixtus* Xyn10A et al. The mutation of this tyrosine significantly reduced the activity against xylotetraose and xylan, consistent with the greater importance of the tyrosine at -3 subsite (Pell et al., 2004b). Furthermore, the product profile of xylotetraose hydrolysis was exclusively xylobiose when the tyrosine at 340 of *Cj*Xyn10C mutated into alanine. It was proposed that the presence of a tyrosine in GH10 enzymes in the equivalent position to Y340 in *Cj*Xyn10C is indicative of a high affinity -3 subsite (Pell et al., 2004b).


β1
4PN2
2W5F
1XYZ
1VBR
3NJ3
4F8X
4XX6
4XV0
4XUY
3WUG
3CUJ
1E0X
1NQ6
1E5N
1US2
4PMU
4K68
1UQY
5D4Y
2DEP
2FGL
4PRW
3EMC
4L4P
4HU8
3RDK
3W29
4W8L
4PN2 1 .....AAPL..AATASKFLGCAYGA.....
2W5F 170 GMAEA.....NTVYAANAVLKDMYANYFRVGSVLNSGT.VNN
1XYZ 2 GQGDV.....QTPNPSVTPTQTPIPTISGNALRDY..AEARGIKIGTCVNYPF.YNN
1VBR 1 .....SQNVSLREL..AEKLNIYIGFAAINNF.WSL
3NJ3 1 .MGHH.....HHHHENL....YFQGHMQNVSLREL..AEKLNIYIGFAAINNF.WSL
4F8X 1 .....DIDLNKLA.QRRGKHWFGTAADIPG.TAE
4XX6 1 .....PTSPFETLR.AAAAPRYFGAALGVPHLLNF
4XV0 1 .....EASQSIDQLI.KRKGKLYFGTATDRGL.LQR
1TA3 1 .....QSASLNDLF.VAAGKSYFGTCSDQAL.LQN
4XUY 1 .....EASVSIDTKF.KAHGKKYLGNIGDQYTLTKN
3WUG 1 .....DTATLGELL..AEAKGRYFGSATDNPE.LPD
3CUJ 1 .....ATTLKEAL..ADGAGRDFGFALDPNR.LSE
1E0X 1 .....AESTLGAAL..AAQSGRYFGTAIASGR.LSD
1NQ6 1 .....AGALGDAL..AAAKGRYFGAAVAANH.LGE
1E5N 1 .....NGLASLADFPIGVAVAASG.G.N
1US2 140 GTPAGTITI.....KSVTI..TLAQEAYSANVDHLRDLAPSDFPIGVAVSNTD.SAT
4PMU 1 .....TSLKQAYAQGFLLGTAVNA.D.IVS
4K68 13 ALSAA.....LLSI.....KGQAIAKAAERTGIKDFYKDDFRIGTAVATAT.LTM
1UQY 6 GIAMG.....QAS.....KLAAATKAAEQTGLKSAYKDNFLIGAALNA.T.IAS
5D4Y 1 ...MA.....HHH.....HHHVDDDDKMAQGLKDAYKDYFKIGVAVNNRN.VAD
2DEP 1 .....MAEDIPSLAEAFRDYFPIGAAIEPGY.TTG
2FGL 1 .....VQPFAWQVASLADRYEESFDIGAAVEPHQ.LNG
4PRW 1 ..KNA.....DSY.....AKKPHISALNAPQLDQRYKNEFTIGAAVEPYQ.LQN
3EMC 1 .....STEIPSLSASYANSFKIGAAVHTRM.L.Q
4L4P 1 .....MSEDYYEKSTVSLTEKYKEFFKIGAAVTVKD.FEG
4HU8 64 ATGEAIITVTAGKQSHSVKVIT..TVKGKEDIEKLPPLKDHFKDYFLIGNIFNNRD.VSG
3RDK 1 .....GSHMAPLKDVYKNDFLIGNAISAED.LE.
3W29 1 .....TIQNDIPDLYSVFKDYFPIGVAVDPSR.LND
4W8L 1 .....NIPDLAKKLGSSYALGAAIDQTA.LDP


η1 α1 β2 α2 α3
4PN2 19 .....QQAPGFAQYWNKLTPENGGKWGSVEAV.....RDQMDWSTLDAAYRFAQANQ
2W5F 206 .....SSIKALILREFNSITCENEMKPDATLVQSGSTNTNIRVSLNRAASILNFCAQNN
1XYZ 51 ....SDPTYNSILQREFSMVVCENEMKFDALQPR.....QNVFDFSKGDQLLAFAERNG
1VBR 29 S...DAEKYMEVARREFNILTPENQMKWDTIHPE.....RDRYNFTPAEKHVEFAEEND
3NJ3 45 S...DEEKYMEVARREFNILTPENQMKWDTIHPE.....RDRYNFTPAEKHVEFAEENN
4F8X 28 T...TDAAYLKVLKQNFGEITPANAMKFMYTETE.....QNVFNFTEGEQFLEVAERFG
4XX6 30 T...HDPLFDVTAVLQFNGATPENEMKWAYIEPE.....RNQFNFTGGDIVAAFSAAND
4XV0 30 .....EKNAAIIQADLGQVTPENSMKWQSLENN.....QGQLNWGDADYLVNFAQQNG
1TA3 29 .....SQNEAIVASQFGVITPENSMKWDALEPS.....QGNFGWSGADYLVDYATQHN
4XUY 31 .....SKTPAIIKADFGALTPENSMKWDATEPS.....RGQFSFSGSDYLVNFAQSNN
3WUG 28 .....TQYTQILGSEFSQITVGNTMKWQYTEPS.....RGRFDYTAAEEIVDLAESNG
3CUJ 27 .....AQYKAIADSEFNLVVAENAMKWDATEPS.....QNSFSFGAGDRVASYAADTG
1E0X 28 .....STYTSIAGREFNMVTAENEMKIDATEPQ.....RGQFNFSSADRVYNWAVQNG
1NQ6 27 .....AAYASTLDAQFGSVTPENEMKWDAVESS.....RNSFSFSAADRIVSHAQSKG
1E5N 22 ADIFTSSARQNIVRAEFNQITAENIMKMSYMYS.....GSNFSFTNSDRLVSWAAQNG
1US2 189 YNLLTNSREQAVVKKHFNHLTAGNIMKMSYMQPT.....EGNFNFTNADAFVDWATENN
4PMU 24 G...KDAASAALVACHFNAVTAENVMKAEVVAPR.....RGVQDFSAADAFVAYAQRDR
4K68 57 K...EKKPLLALIAREFNAITPENCMKWEPLKPQ.....DKDWHWEAADKFVEFGEKHK
1UQY 48 G...ADERLNTLIAKEFNSITPENCMKWGVLRDA.....QGQWNWKDADAFVAFGTKHN
5D4Y 41 P.....DQIKVVLREFNSITAENAMKPQPTEPK.....KGEFNWEDADKIADFCRANG
2DEP 30 Q.....IAELYKKHVNMLVAENAMKPASLQPT.....EGNFQWADADRIVQFAKENG
2FGL 33 R.....QGKVLKHHYNSIVAENAMKPISLQPE.....EGVFTWDGADAIVEFARKNN
4PRW 42 E.....KDVQMLKRHFNSIVAENVMKPISIQPE.....EGKFNFEQADRIVKFAKANG
3EMC 28 T.....EGEFIAKHYNSVTAENQMKFEEVHPR.....EHEYTFEAADEIVDFAVARG
4L4P 35 I.....HGRILTKHFNSLTPENDMKFERIHPK.....EDFYNFEATDKIKDFALKHN
4HU8 121 S.....MMDNDWLAHHYAILTPENHMKPSNLTNNRNETTGEITYTFSTADRMVNAAIAEG
3RDK 28 .....GTRLELLKMHHDVVTAGNAMKPDALQPT.....KGNFTFTAADAMIDKVLAEG
3W29 31 A.....DPHAQLTAKHFNMLVAENAMKPESLQPT.....EGNFTFDNADKIVDYAIAHN
4W8L 27 K.....DPHSELLTKHFNSITAGNFMKMDAMQPT.....EGKFVWSEADKLVNFAAANN


β3
α4
4PN2
2W5F
1XYZ
1VBR
3NJ3
4F8X
4XX6
4XV0
4XUY
3WUG
3CUJ
1E0X
1NQ6
1E5N
1US2
4PMU
4K68
1UQY
5D4Y
2DEP
2FGL
4PRW
3EMC
4L4P
4HU8
3RDK
3W29
4W8L
4PN2 66 MPFQMHVMVWGN..QQPEWIKTLR.....PAEQRREIEQWFAAVAQRY
2W5F 260 IAVRGHTLVWHS..QTPQWFFKDNFQDN.....GNWVSQSVMDQRLESYIKNMFAEI
1XYZ 101 MQMRGHTLIWHN..QNPSWLTNGN.....WNRDSLLAVMKNHITTVMTHY
1VBR 80 MIVHGHTLVWHN..QLPGWITGRE.....WTKEELLNVLEDHIKTVVSHF
3NJ3 96 MIVHGHTLVWHN..QLPGWITGRE.....WTKEELLNVLEDHIKTVVSHF
4F8X 79 SKVRCHNLVWAS..QVSDFVTSKT.....WTAKELTAVMKNHIFKTVQHF
4XX6 81 YVLRGHNLVWYQ..ELAPWVETLT.....GEDLWNATVNHITTVMTHY
4XV0 78 KSIRGHTLIWHS..QLPAWVNNIN.....NADTLRQVIRTHVSTVVGRY
1TA3 77 KKVRGHTLVWHS..QLPSWVSSIG.....DANTLRSVMTNHINEVVGRY
4XUY 79 KLIRGHTLVWHS..QLPSWVQSIT.....DKNTLIEVMKNHITTVMQHY
3WUG 76 QSVRGHTLVWHN..QLPSWVDDVP.....AGELLGVMRDHITHEVDHF
3CUJ 75 KELYGHTLVWHS..QLPDWAKNLN.....GSAFESAMVNHVTKVADHF
1E0X 76 KQVRGHTLAWHS..QQPGWMQSLS.....GSALRQAMIDHINGVMAHY
1NQ6 75 MKVRGHTLVWHS..QLPGWVSPLA.....ATDLRSAMNNHITQVMTHY
1E5N 75 QTVHGHALVWHPSYQLPNWASDS.....NANFRQDFARHIDTVAAHF
1US2 243 MTVHGHALVWHSDYQVPNFMKNWA.....GSAEDFLAALDTHITTIVDHY
4PMU 75 QFVVGHTLVWHN..QTPEWFFTTADGRP.....NTPAQQLERMRAHIAAVAGRY
4K68 108 MYIVGHNLVWHS..QVPKEVFLNESGGT.....ISKEALTAKMQDHIATLAGRY
1UQY 99 LHMVGHTLVWHS..QIHDEVFKNADGSY.....ISKAALQKKMEEHITTLAGRY
5D4Y 89 IKMRGHTLMWHS..QIGSWMYQDEKGNL.....LSKEEFYANMKHHIQAIVNRY
2DEP 77 MELRFHTLVWHN..QTPDWFFLDKEGKPMVEETDPQKREENRKLLLQRLENYIRAVVLRY
2FGL 80 MNLRFHTLVWHN..QVPDWFFLDEEGNPMVEETNEAKRQANKELLLERLETHIKTVVERY
4PRW 90 MDIRFHTLVWHS..QVPQWFFLDKEGKPMVNETDPVKREQNKQLLLKRLETHIKTIVERY
3EMC 75 IGVRGHTLVWHN..QTPAWMFEDASG.....GTASREMMLSRLKQHIDTVVGRY
4L4P 82 MQLRGHTLVWHN..QTPEWVFRDNDK.....EAPKELVIERLREHIKTICTRY
4HU8 176 LKIHGHTLLWHQ..QIPPWQRSME.....SAAKDAALSVMKKYITEVMTHY
3RDK 76 MKMHGHVLVWHQ..QSPAWLNTKKDDNN.....NTVPLGRDEALDNLRTHIQTVMKHF
3W29 80 MKMRGHTLLWHN..QVPDWFFQDPS..D.....PSKPASRDLLLQRLRTHITTVLDHF
4W8L 76 MQVRGHTLLWHS..QVPDWFFTDPN..D.....PSKPATREQLMQRMKTHIQTIVSRY


β4
η2
4PN2
2W5F
1XYZ
1VBR
3NJ3
4F8X
4XX6
4XV0
4XUY
3WUG
3CUJ
1E0X
1NQ6
1E5N
1US2
4PMU
4K68
1UQY
5D4Y
2DEP
2FGL
4PRW
3EMC
4L4P
4HU8
3RDK
3W29
4W8L
4PN2 107 ..P.D.....IALLEVVNEPLNDPPSKAD.....TGGGNYLQALGG..NGDSG
2W5F 310 ..QRQYPS.LNLYAYDVVNEAVSDDANRTRYYGGAREPGYGNGRSPWVQIYGD.....
1XYZ 144 ..KGK.....IVEWDVANECMDDSG.....NGLRSSIWRNVIGQ.....
1VBR 123 ..KGR.....VKIWDVVNEAVSDSG.....T.YRESVWYKTIGP.....
3NJ3 139 ..KGR.....VKIWDVVNEAVSDSG.....T.YRESVWYKTIGP.....
4F8X 122 ..GRR.....CYSWDVVNEALNGDG.....TFSSSVWYDTIGE.....
4XX6 122 ..KESFN....IYAWDVVNEAFNDNG.....TYRENVWYTQLGP.....
4XV0 120 ..KGK.....IRAWDVVNEIFNEDG.....TLRSSVFSRLLGE.....
1TA3 119 ..KGK.....IMHWDVVNEIFNEDG.....TFRNSVFYNLLGE.....
4XUY 121 ..KGK.....IYAWDVVNEIFNEDG.....SLRDSVFYKVIGE.....
3WUG 117 ..KGR.....LIHWDVVNEAFEEDG.....SRRQSVFQQKIGD.....
3CUJ 116 ..EGK.....VASWDVVNEAFADGG.....GRRQDSAFQQKLGN.....
1E0X 117 ..KGK.....IVQWDVVNEAFADGS.....SGARRDSNLQ.RSGN.....
1NQ6 116 ..KGK.....IHSWDVVNEAFQDGG.....SGARRSSPFQDKLGN.....
1E5N 117 ..AGQ.....VKSWDVVNEALFDSADDPDGRGS....ANGYRQSVFYRQ.FG.....G
1US2 288 EAKGN.....LVSWDVVNEAIDDNSPANF.....RTTDSAFYVKSGN.....S
4PMU 122 ..TGK.....VQAWDVVNEIIDED.....GSYRSTNWVQRVGD.....G
4K68 155 ..KGR.....IQAWDVVNEAVEDD.....GSWRKSPWYNIMGE.....
1UQY 146 ..KGK.....LAAWDVVNEAVGDD.....LKMRDSHWYKIMGD.....
5D4Y 136 ..KDV.....VYCWDVVNEAVADSPVYP....G....RPELRNSPMYQIAGE.....
2DEP 135 ..KDD.....IKSWDVVNEVIEPND..P.....GGMRNSPWYQITGT.....
2FGL 138 ..KDD.....VTAWDVVNEVVDDGT..P....N....ERGLRESVWYQITGD.....
4PRW 148 ..KDD.....IKYWDVVNEVVGDD.....GKLRNSPWYQIAGI.....
3EMC 122 ..KDQ.....IYAWDVVNEAIEDKT..D.....LIMRDTKWLRLLGE.....
4L4P 128 ..RDV.....VYSWDVVNEAVEDKT..D.....VLLRDSKWRRIIGD.....
4HU8 220 ..KGK.....IYSWDVLNEIFPDGRGDNWT.....TAMRPENPWFKSIG.....
3RDK 127 ..GNK.....VISWDVVNEAMNDNPSNPAD..Y....KASLRQTPWYQAIG.....
3W29 129 ..KTKYGSQNPIIGWDVVNEVLDDNGNL.....RN....SKWLQIIG.....
4W8L 125 ..KGK.....VHTWDVVNEVISDGGGL.....RNQASGSKWRDIIGDVDGDGDD


α5
β5
α6
4PN2
2W5F
1XYZ
1VBR
3NJ3
4F8X
4XX6
4XV0
4XUY
3WUG
3CUJ
1E0X
1NQ6
1E5N
1US2
4PMU
4K68
1UQY
5D4Y
2DEP
2FGL
4PRW
3EMC
4L4P
4HU8
3RDK
3W29
4W8L
4PN2 145 WEWVLQSFRLARRHF.....P.HTKLMINDYSITSS..AQATQKYLQIVRL.....
2W5F 360 NKFIEKAFTYARKYA.....PANCKLYYNDYNEYWD..HKRDCIASICANL.....
1XYZ 176 .DYLDYAFRYAREA.....DP.DALLFYNDYNIEDL.GPKSNAVFNMIKSM.....
1VBR 154 .EYIEKAFRWAKEA.....DP.DAILIYNDYSIEEI.NAKSNFVYNMIKEL.....
3NJ3 170 .EYIEKAFRWTKEA.....DP.DAILIYNDYSIEEI.NAKSNFVYNMIKEL.....
4F8X 153 .EYFYLAFKYAQEALAQIGAN.DVKLYYNDYGIENP.GTKSTAVLQLVSNL.....
4XX6 155 .DYIPNAYAVARSV.....NT.PSKLYINDYNTEGI.NNKSDALLAVVQSM.....
4XV0 151 .EFVSIAFRAARDA.....DP.SARLYINDYNLDRANYGKVNGLKTYVSKW.....
1TA3 150 .DFVRIAFETARAA.....DP.DAKLYINDYNLDSASYAKTQAMASYVKKW.....
4XUY 152 .DYVRIAFETARAA.....DP.NAKLYINDYNLDSASYPKLTGMVSHVKKW.....
3WUG 148 .SYIAEAFKAARAA.....DP.DVKLYYNDYNIEGI.GPKSDAVYEMVKSF.....
3CUJ 148 .GYIETAFRAARAA.....DP.TAKLCINDYNVEGI.NAKSNSLYDLVKDF.....
1E0X 149 .DWIEVAFRTARAA.....DP.SAKLCYNDYNVENWTWAKTQAMYNMVRDF.....
1NQ6 149 .GFIEEAFRTARTV.....DA.DAKLCYNDYNTDGQ.NAKSNAVYEMVKDF.....
1E5N 158 PEYIDEAFRRARAAD.....P.TAELYYNDFNTEEN.GAKTTALVNLVQRL.....
1US2 326 SVYIERAFQTARAAD.....P.AVILYYNDYNIEQN.NAKTTKMVDMVKDF.....
4PMU 154 DTVVRNAFAFAQRYA.....P.DAQLYYNDFNAWRP..AKREGIVRMVKML.....
4K68 186 .DFIAKAFTMAHEVD.....P.KAHLIYNDYNTESP..IKRNFIVGMIKNF.....
1UQY 177 .DFIYNAFTLANEVD.....P.KAHLMYNDYNIERT..GKREATVEMIERL.....
5D4Y 173 .EFIYKAFEYAHEAD.....P.DALLFYNDYNDAEP..AKSQRIYNLVKRM.....
2DEP 168 .EYIEVAFRATREAG.....GSDIKLYINDYNTDDP..VKRDILYELVKNL.....
2FGL 173 .EYIRVAFETARKYA.....GEDAKLFINDYNTEVT..PKRDHLYNLVQDL.....
4PRW 179 .DYIKVAFQAARKYG.....GDNIKLYMNDYNTEVE..PKRTALYNLVKQL.....
3EMC 155 .DYLVQAFNMAHEAD.....P.NALLFYNDYNETDP..VKREKIYNLVRSL.....
4L4P 161 .DYIKIAFEIAKKYT.....G.NGKLFYNDYNNEMP..YKLEKTYKVLKSL.....
4HU8 257 SDFVYEAYLAARQAD.....P.NAILYYNDYNMDQA..GKAALIAAMVRDVNAKYKQAYP
3RDK 165 SDYVEQAFLAAREVLDENPSW.NIKLYYNDYNEDNQ..NKATAIYNMVKDINDRYAAA..
3W29 165 PDYIEKAFEYAHEAD.....P.SMKLFINDYNIENN.GVKTQAMYDLVKKL.....
4W8L 167 SDYIELAFRYAREAD.....P.DAVLVINDYGIEGS.VSKMNDMVKLVEKL.....

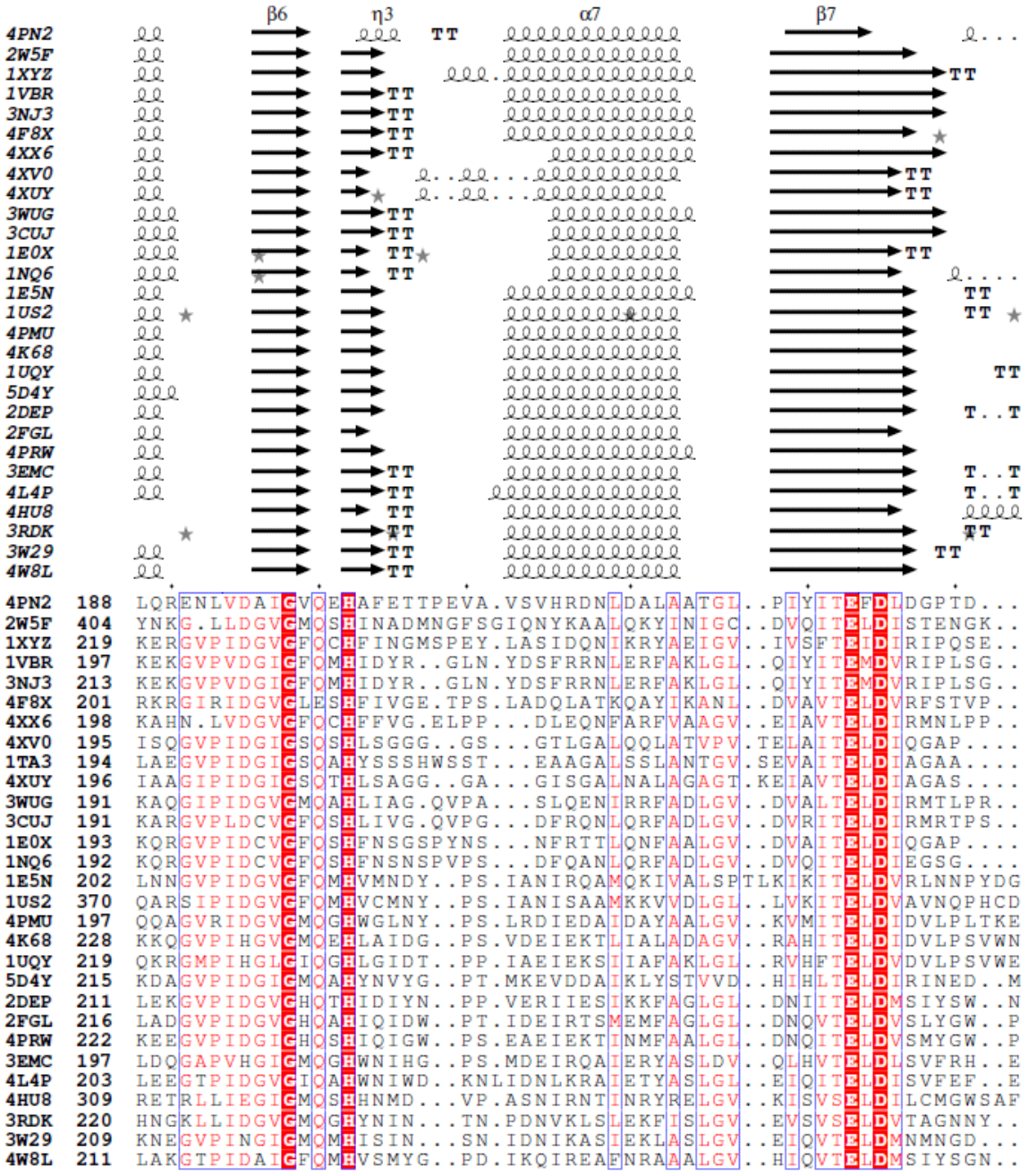


Fig B-4. Alignment of GH10 xylanases with characterized structures

The sequence of each protein was extracted from the PDB file, and the alignment analysis was done by ESPript 3 (Robert and Gouet, 2014). The labels α, β, η, T refer to α-helices, β-strands, 3_{10}-helices and π-helices respectively. Grey stars indicate alternative conformations.

B.1.4.1.3 GH10 enzyme: 3D structure-domain organization

The first determined structure of the catalytic domain of a GH10 xylanase was that from *Streptomyces lividans* xylanases A (Derewenda et al., 1994). Since then, numerous studies have shown that GH10 xylanases are characterized by a highly conserved $(\beta/\alpha)_8$-barrel fold, with only relatively small variations being observed especially in loops and helices

surrounding the inner β-strands (Schmidt et al., 1999; Solomon et al., 2007)(*Fig B-2*). Additionally, some variation of the shape of the barrel has been observed in GH10 xylanases, with this ranging from nearly circular to quite elliptical (Pollet et al., 2010).

$(\beta/\alpha)_8$-barrel fold

The canonical $(\beta/\alpha)_8$-barrel contains about 200 amino acids and is composed of eight units, each of which consists of a β-strand and an α-helix that are connected by a βα-loop. The eight individual units are linked by αβ-loops. The βα loops are important for function, because the actives sites of all TIM-barrel enzymes are located at the C-termini of β strands (Wierenga, 2001). On the other hand, αβ loops are believed to play important roles in maintaining stability of TIM-barrel proteins (Urfer and Kirschner, 1992). The βα loops display variable length, but tend to be longer than αβ loops. The word "barrel" refers to a curved central parallel β –sheet formed by eight β –strands, around the barrel are α –helices. $(\beta/\alpha)_8$-barrel protein fold is firstly found in the triosephosphate isomerase or TIM, and that's why it is also called TIM barrel fold (Spratt and Pardee, 1975). This fold is one of the most common protein folds known is most often associated with enzymes that are involved in molecular or energy metabolism within the cell (Copley and Bork, 2000; Nagano et al., 2002).

It remains an open question as to whether the TIM barrel fold is evolved via convergent evolution or divergent evolution. Understanding its evolutionary history is of significant importance for the prediction of β/α–barrel structures and for the design of enzymes with new functions. Comparing intron positions in pyruvate kinase and triosephosphate isomerase genes, Straus and Gilbert showed that these mapped onto the three-dimensional structures of the corresponding proteins (Straus and Gilbert, 1985), strongly favours the hypothesis of convergent evolution. Farber, Branden et al proposed the existence of a common ancestor, i.e. divergent evolution for TIM barrel fold based on structural similaries (Brändén, 1991; Farber and Petsko, 1990). This argument was further supported by more structural evidences (Nagano et al., 2002; Reardon and Farber, 1995) and phylogenic studies (Copley and Bork, 2000). Recent studies were also in support of its divergent origin and showed that many TIM barrel superfamilies probably originated from common ancestors that demonstrate a high degree of catalytic diversity thanks to evolutionarily malleable active site pockets (Goldman et al., 2016). Other voice concerning the evolution of TIM barrels is saying that the smallest evolutionary unit could be the half-barrel, $(\beta/\alpha)_4$. This is because HisA and HisF, two

enzymes involved in histidine biosynthesis, were shown to evolved by twofold gene duplication of a half-barrel ancestor, followed by gene fusion (Lang et al., 2000).

While the argument about its evolution origin is going on, enzyme engineers have achieved enzymatic activity migration (Saab-Rincón et al., 2012) by combining rational design and random mutagenesis and developed different tools for chimeric protein design (Wang et al., 2017). Research in the de novo design of a TIM barrel protein guided by geometrical and chemical principles suggests that the potential functional range of the TIM barrel is greater than represented naturally occurring proteins (Huang et al., 2016).

Insertional domain in $(\beta/\alpha)_8$-barrel fold

Formally, domains are defined as stable globular fragments of proteins that may refold autonomously and carry out specific functions (Khosla and Harbury, 2001). Proteins can comprise a single domain or a combination of domains. In multi-domain proteins, the amino acid sequence of domains almost always occur in tandem from N-terminal to C-terminal except 9% exceptions, where one domain (insert) inserted into another domain (parent) that is often in α/β fold (Aroul-Selvam et al., 2004).

TIM barrels can contain one or more insertional domains (Nagano et al., 2002), with these often occurring in the βα3 loop. In pyruvate kinase, the βα3 loop insertion is a ~100 residue-domain, known as the domain B, which barely interacts with the domain A of the protein (Larsen et al., 1994; Mattevi et al., 1995). Similarly, α-amylase also contains a so-called domain B in its βα3 loop. From one amylase to another, this insertional domain displays variable length and sequence (Janecek, 1997; Jespersen et al., 1993; Klein and Schulz, 1991; Qian et al., 1993), although researchers have reported the fact that there is a relatively conserved aspartate residue near the C-terminus of the domain B (D175 in Taka-amylase A), which is involved in the binding of calcium ions (Liu et al., 2017). It has been postulated by Buisson that Ca^{2+} ions preserves the structural integrity of the active site by linking the catalytic $(\beta/\alpha)_8$-barrel and the extruding domain B (Buisson et al., 1987). In α-amylase, it has been shown that domain B reinforces the conformation of the active site and enhances enzyme thermostability (Janecek, 1997; Liu et al., 2017). In the case of susG protein, an α-amylase belonging to the starch utilization system (SUS), the domain B is itself a target for insertion, containing a CBM58 (Koropatkin and Smith, 2010). This further modification of the TIM-barrel, plausibly related to the glycan foraging function of members of

Bacteroidetes, reveals the tolerance of the TIM barrel architecture to modification and provides a paradigm for the creation of chimeric enzymes. In this respect, further evidence of the relative 'plasticity' of the TIM-barrel is provided by the existence of circular permutations, which provide family GH70 α-glucan acting enzymes (MacGregor et al., 1996).

Finally, beyond the α amylase group, recent studies revealed that a GH10 xylanase also contains inserted CBMs in its βα3 loop (Flint et al., 1997; Zhang et al., 2014). Strikingly, this enzyme is part of polysaccharide utilization loci (PUL, see section B3.2) analogous to SUS and therein might play a role that is comparable to that of SusG.

B.1.4.2 1,4-β-D-Xylosidase

1,4-β-D-Xylosidases (β-D-xyloside xylohydrolases; EC 3.2.1.37) are exoglycosidases, hereafter referred to as xylosidases, that act on the nonreducing ends of short xylooligosaccharides (*FigB-3*). Xylosidases operate through either a retaining (e.g. family GH 39) or inverting (e.g. family GH43) mechanism and liberate D-xylose. It has been observed that many retaining xylosidases display substantial transferase activity especially at high substrate concentrations (Reilly, 1981; Wong et al., 1988), and this activity results in products of higher molecular weight than that of the substrate and the formation of both β-1,3 and β-1,4 bonds (Claeyssen, 1966; Sunna and Antranikian, 1997). Xylosidase activity is inversely related to the chain length of xylooligosaccharides, with xylobiose being the preferred substrate (Takenishi et al., 1973) except in some cases (Saha, 2001). Some xylosidases are strongly inhibited by D-xylose and its homologues (Claeyssens and De Bruyne, 1965; Saha, 2001), while others are not (Yan et al., 2008; Zanoelo et al., 2004).

Xylosidases are found in animal, plant and microbial kingdoms, with many of the latter being of fungal origin (e.g. from *Aspergillus*, *Botryodiploidia*, *Penicilliu*, *Pestalotia* and *Trichoderma*). Regarding bacterial xylosidases, those of *Bacillus* are numerous, while others are produced by plant pathogens, such as *Agrobacterium*, *Corynebacterium* and *Erwinia* (Reilly, 1981). A striking feature of xylosidases is their multiplicity and the fact that they are distributed throughout several GH families, including families GH 30 (contains 8 subfamilies), 39, 43 (contains 37 subfamilies) 51, 52, 116 and 120. Consistent with this, different xylosidases display quite different specificities, being active on from aryl-β-D-xylopyanosides, aryl-α-L-arabinopyranosides, aryl-β-D-glucopyranosides, aryl-β-D-quinopyranosides, xylobiitol, xylotriitol, and L-serine xylopyranoside (Reilly, 1981; Sunna and Antranikian, 1997).

B.1.4.3 Accessory xylanolytic enzyme

Owing to the heterogeneity of xylan structure, a mere combination of endoxylanase and xylosidase is unlikely to achieve complete degradation. For this to occur, other so-called accessory (or auxiliary) enzymes are required, notably to release mainchain decorations. For example, α-L-arabinofuranosidases can release α-L-arabinofuranosides from heteroxylan, while various xylan esterases will release (for example) acetic and ferulic acid (*Fig B-3*). Accessory enzymes are produced by many microorganisms, such as *Penicillium capsulatum* and *Talaromyces emersonii*, which possess complete xylan-degrading enzyme systems (Saha, 2003).

B.1.4.3.1 α-L-Arabinofuranosidase

α-L-Arabinofuranosidases (α-L-arabinofuranoside arabinofuranohydrolase, EC 3.2.1.55), hereafter referred to as arabinofuranosidases of Abfs, release arabinose from arabinose-containing polysaccharides, such as arabinan, arabinoxylan, gum arabic and arabinogalactan (Saha, 2000). Abfs are exo-type enzymes that hydrolyze terminal nonreducing residues linked by 1,3 or 1,5-α-arabinosyl linkages. Beldman summarized the classification of this enzyme group based on substrate specificities: class 1 are not active towards polymers; class 2 are active towards polymers; class 3 are specific for arabinoxylans (Beldman et al., 1997). The CAZy database classifies Abfs within 6 GH families: GH2, 3, 43, 51, 54 and 62 on the basis of amino acid sequence similarities. There are only two characterized arabinofuranosidases from GH2 and 3, and these are bifunctional containing another activity such as β-galactosidase and β-xylosidase. All members of GH51 and 54 catalyze hydrolysis with net retention of anomeric configuration, while GH43 members operate via an inverting mechanism. The mechanism of GH62 is not elucidated, although it shares the same tertiary fold as GH43, with both families beloning to clan GH-F (Wilkens et al., 2017). Only GH62 contains exclusively α-L-arabinofuranosidases and these are of fungal and bacterial origin.

B.1.4.3.2 Hemicellulolytic esterase

Xylan esterases hydrolyze the ester bonds that either link acetyl groups to mainchain xylosyl residues acid (acetylxylan esterase, EC 3.1.1.72) or that link cinnamoyl (e.g. ferulolyl) groups to the O-5 position of arabinosyl side chain decorations in arabinoxylans (ferulic acid esterase, EC 3.1.1.73 and *p*-coumaric acid esterase, EC 3.1.1-). Acetylxylan esterases are

grouped within 7 of the 16 carbohydrate esterase familes in CAZy (Ryabova et al., 2009). Their role in heteroxylan hydrolysis is vital, since the presence of acetyl groups is known to be a major hindrance to xylanases (Salis et al., 2007). The majority of ferulic acid esterases are classified in CE1. The presence of hydroxycinnamic acids in heteroxylans not only hinders the action of enzymes such as Abfs, but also provides the basis for intermolecular crosslinking between xylan molecules and between xylan and lignin. Ferulyol and *p*-coumaric acid esterases are therefore precious, because they can disrupt crosslinking and help to solubilize xylan (Shallom and Shoham, 2003).

B.1.4.3.3 α-D-Glucuronidase

Only two CAZy families, GH67 and GH115, harbor α-glucuronidases, implying that this type of enzyme is subject to much less variability when compared to other accessory. One GH4 enzyme has also been described as an α-D-glucuronidase. However, it does not recognize 4-O-methyl-D-glucuronoxylan or derivatives thereof (Suresh et al., 2003). Members of both GH67 and 115 operate via an inverting catalytic mechanism. The first GH67 glucuronidase was described in 1983 (Dekker, 1983). This type of enzyme releases 4-O-methyl-D-glucuronic acid (MeGlcA) or GlcA from oligosaccharide motifs that are derived from glucuronoxylan (aldouronic acids), in which the uronic acid is linked to the non-reducing terminal xylopyranosyl residue. However, GH67 enzymes do not attack aryl α-glycosides of GlcA or MeGlcA (Ryabova et al., 2009). Nurizzo et al studied the 3D structure and biochemical properties of a GH67 enzyme from *P.cellulosa*. This revealed that its active site adopts the form of a deep pocket that recognizes both the 4-O-methyl and the carboxylate components (Nurizzo et al., 2002). The more recent discovery of family GH115 glucuronidases revealed that these enzymes hydrolyze linkages that bind MeGlcA moities to both terminal and internal xylopyranosyl residues of glucuronoxylan and aldouronic acids derivatives generated by the action of xylanases (Ryabova et al., 2009; Tenkanen and Siika-aho, 2000). The detailed characterization of a multidomain GH115 from *Bacteroides ovatus* showed that it is a dimeric enzyme that contains a flexible active site cleft housed in the TIM barrel domain (Rogowski et al., 2014).

B.1.4.4 Synergistic interaction between xylanases and accessory hemicellulases

The chemical complexity of heteroxylans is clearly a major obstacle for enzyme degradation (Kormelink and Voragen, 1993; Lee et al., 1987). In degrading arabinoxylans, the introduction of Abfs into xylanase can enhance the performance of the latter (Suh et al., 1996; de Vries et al., 2000) and sometimes is a pre-request when degrading agricultural residues (Tuohy et al., 1993), probably due to the involvement of arabinose in the cross-linking between lignin and hemicellulose (Buchmann and McCarthy, 1991; Chesson et al., 1983; Smith and Hartley, 1983).

The mix of acetyl esterase with xylanase is advantageous for the action of both enzymes (Biely et al., 1986; Blum et al., 1999; Suh and Choi, 1996). The release of ferulic acid by ferulic acid esterases rely more on the action of xylanase especially when complex substrate is utilized (Faulds and Williamson, 1991; Mackenzie and Bilous, 1988; de Vries et al., 2000; Yu et al., 2002), and different xylanases used may result in different product form (Faulds et al., 2006).

It was reported that the function of α-glucuronidase requires a low-molecular-weight glucuronoxylan substrate (Smith and Forsberg, 1991). Xylanases can improve the performance of α-glucuronidase by providing methylglucuronoxyl XOS (Shao et al., 1995; de Vries et al., 2000). The synergistic interaction between α-glucuronidase and β-xylosidase was also proven (Shao et al., 1995).

B.2 CBMs

Glycoside hydrolases are often inefficient for the degradation of insoluble polysaccharides. This is because it is difficult for GHs to access the substrate at the right position for catalysis (Bornscheuer et al., 2014; Salis et al., 2007). Therefore, a distinct structural feature of many lignocellulose-degrading enzymes is modularity, with the catalytic core being covalently linked to one or more non-catalytic protein domains. Majoritarily, these domains are carbohydrate-binding modules (CBM), although others are fibronectin 3-like modules, dockerins, immunoglobulin-like domains, or functionally unknown “X” domains. Regarding CBMs, these are found as discretely folded units within the multi-modular structures of many polysaccharide-targeting enzymes, such as cellulases and hemicellulases.

B.2.1 Classification

Boraston et.al. manually classified the structures of CBMs from 22 different CBM families into seven 'fold families' based on the conservation of the protein fold. These seven families possess β-sandwich, β-trefoil, cysteine knot, unique, OB, hevein and hevein-like folds respectively (Boraston et al., 2004). However, this system of classification is not useful to predict function. Therefore, an alternative system that considers the topology of the CBM binding site has been proposed. Herein, type A CBMs possess flat or platform-like hydrophobic surfaces composed of aromatic residues that are well-adapted for the binding of crystalline polysaccharides, such as cellulose or chitin. The planar architecture of the binding site is thought to be complementary to the flat surfaces presented by cellulose or chitin crystals (Hashimoto, 2006). Type B CBMs displays a cleft/groove arrangement in which aromatic residues are oriented in such a way that forms twisted or sandwich platforms (Guillén et al., 2010), and interact with internal glycan chains (endo type) including amorphous cellulose or xylan, and also recognize β-1,3-glucans, β-1,3-1,4-mixed linkage glycans, β-1,4-mannan, glucomannan, and galactomannan. Type C CBMs or lectin-like CBMs , lack the extended binding-site grooves of type B CBMs and bind to the termini of glycans (exo type), so it has limited ligands such as mono-, di-, or trisaccharides (Boraston et al., 2004; Bornscheuer et al., 2014). Finally, like the carbohydrate-active enzymes, CBMs are also present in the CAZy classification and are currently grouped into 81 families based on amino acid sequence similarity (see http://www.cazy.org/Carbohydrate-Binding-Modules.html). Together, the 81 families represent a wide range of ligand specificities. Considering the different classifications it is possible to note that type A CBMs are found CAZy CBM families 1, 2, 3, 5, and 10 and all of these bind to insoluble ligands. Type B CBMs are found in CAZY CBM families 4, 6, 15, 17, 20, 22, 27, 28, 29, 34 and 36, while type C CBMs are grouped in CAZY CBM families 9, 13, 32, 47, 66 and 67 (https://www.cazypedia.org/index.php/Carbohydrate-binding_modules).

B.2.2 The mechanism of binding

Lehito et al. has used transmission electron microscopy to probe the location of the CBM-binding site on *Valonia* crystalline cellulose. They confirmed that both CBM1 and CBM3a bind to the hydrophobic face (Lehtiö et al., 2003). The study of its 3D crystal structure indicates that the major interaction relies on a planar strip of aromatic residues. Additionally, polar amino acid residues are proposed to anchor the CBM molecule to two other adjacent

chains of crystalline cellulose (Tormo et al., 1996). Aromatic residues in three types of CBMs mentioned above play a pivotal role in ligand binding and for type B and C CBMs, the direct hydrogen bonds also play a key role in the affinity and ligand specificity.

The most important conformational element of the majority of CBMs is the beta-sheet, and β-sandwich is the major fold in CBM structures (Hashimoto, 2006). β-sandwich CBMs often bound metal ions that mainly contribute to structural stability (Ghosh et al., 2013; Tormo et al., 1996). Furthermore, the calcium ion in CBM35 (Bolam et al., 2004), CBM36 (Jamal-Talabani et al., 2004) was shown to involve in ligand binding. The residues involved in the binding sites have been identified using NMR, site-directed mutagenesis and structural investigations (Itoh et al., 2006; Newstead et al., 2005; Pantoom et al., 2008; Rodríguez-Sanoja et al., 2009).

In general, the importance of aromatic amino acids, especially tryptophan, but also the highly conserved tyrosine residues, is well known. They form stacking interactions with the sugar rings resulting in strong van der Waals interactions, thus stabilizing the complex. In addition, the side chain of other polar amino acid residues may form hydrogen bonds with the sugar ligand, helping to stabilize the interaction. Several studies suggest that orientation of aromatic residues is responsible for the different ligand specificities of the CBM families (Miyanaga et al., 2006; Najmudin et al., 2006; Simpson et al., 2000).

B.2.3 The function of CBM in glycoside hydrolases

The relationship between a CBM and a catalytic domain within a single polypeptide is variable depending on the protein under consideration. In some cases, the functional properties of the domains are apparently independent of one another. This has been illustrated by removing CBMs from enzymes and demonstrating that the catalytic domain remains active (Gilbert et al., 1990; Gilkes et al., 1988) and that the CBM retains its substrate-binding ability (Hachem et al., 2000; Zverlov et al., 2001). However, there are also numerous examples in the literature that show the contrary.

When catalytic and CBM domains display functional interdependency, this can be expressed to different degrees and be rather substrate dependent. For example, the separation of CBMs from catalytic modules in some enzymes has resulted in significant decreases in the activity of the enzymes on insoluble substrates (Ali et al., 2001; Black et al., 1996; Gilkes et al., 1988; Puranen et al., 2014; Sakka et al., 2011) and soluble polysaccharides (Kittur et al., 2003; Zverlov et al., 2001). In other cases removal of CBMs has been shown to have more globally

drastic effects on enzyme activity (Arai et al., 2003; Burstein et al., 2009; Gilad et al., 2003). This is well-exemplified by the removal of a CBM3 from a GH9 cellulase, which resulted in almost complete loss of activity, even when soluble substrates were tested (Gal et al., 1997). Likewise, the direct participation of a CBM32 in substrate recognition required for catalytic action was also suggested (Mizutani et al., 2014). In the case of the *Clostridium stercorarium* Xyn10B, an enzyme that is highly active toward barley β-1,3-1,4-glucan as well as β-1,4-xylan, excision of the CBM22s led to loss of the β-1,3-1,4-glucanase activity, but not xylanase activity (Araki et al., 2004; Zhao et al., 2005), which highlights the important role for the CBMs in the determining substrate specificity of cognate enzymes.

Studies have revealed that at high substrate concentration and thus lower water activity, the importance of CBMs for substrate recognition and catalysis diminishes. This is illustrated by a study that showed that at 20% (w/w) substrate concentration, the hydrolytic performance of cellulases devoid of CBMs was similar to that of cellulases with CBMs (Várnai et al., 2013). This phenomenon might explain why in some studies no synergistic effects between catalytic modules and CBMs were observed in assays performed on insoluble substrates, such as certain xylans (Hachem et al., 2000).

Lytic polysaccharide monooxygenases (LPMOs) also figure among the biomass-degrading enzymes and are now thought to be important components of enzyme arsenals (Eijsink et al., 2008). It is interesting to note that some studies have shown that the creation of chimeric LPMOs containing a CBM causes deleterious effects on catalytic activity (Crouch et al., 2016). These observations suggest that CBMs possess a variety of functions, which are different from one CBM to the next (Cuskin et al., 2012; Hervé et al., 2010; Montanier et al., 2009). This further implies that it is not straightforward to add a CBM from one enzyme to another and generate beneficial effects.

Finally, it is noteworthy that CBMs, often borne on multimodular (non-enzymatic) proteins, have been identified in several species. Study of these proteins has revealed a wide variety of functions for the CBMs, including toxins, virulence factors or pathogenesis associated proteins. In this case, CBMs are involved in the promotion of tissue destruction and the enhancement of either bacterial spread or pathogenesis in plants and animals (Guillén et al., 2010).

B.2.4 The mechanism of CBM in potentiating catalysis

The mechanism by which CBMs potentiate catalysis remains unclear. Three general mechanisms have been proposed: CBMs (i): create substrate disruption, (ii) produce a proximity effect, (iii) target catalytic modules to substrates.

Substrate disruption has been proposed as an important function for cellulose-specific CBMs. It has been hypothesized that these disrupt the ordered hydrogen-bonding network in crystalline cellulose, making the surface chains accessible to the appended cellulase (Jonathan Knowles, 1987; Teeri, 1997). This has been supported by biochemical, biophysical and microscopic data that indicate that CBMs mediate changes to the surface structure of cellulose (Din et al., 1994; Simpson et al., 2000; Sorimachi et al., 1997; Southall, 1999; Viegas et al., 2008; Wang et al., 2008).

Regarding the proximity effect, studies have revealed that some CBMs enhance the efficiency of appended catalytic domains by mediating prolonged and intimate contact between the respective catalytic domain and its target substrate, thereby improving substrate accessibility (Bolam et al., 1998; Gilbert, 2010).

Finally, work for example on agarases bearing CBM6 domains has shown that these latter bind specifically to the non-reducing end of agarose chains, recognizing only the first disaccharide repeat. This specific targeting towards the agarose chains apparently directs the appended catalytic domain to the area of the plant cell wall where the polysaccharide is likely to be most amenable to enzymatic hydrolysis (Henshaw et al., 2006). In the context of intact primary and secondary cell wall deconstruction, it has also been shown that CBMs can potentiate the action of a cognate catalytic domain through the recognition of non-substrate polysaccharides. In other words, the targeting function of CBMs can be used in the context of complex, heterogeneous matrices to create strong proximity effects. This explains why cellulose-specific CBMs are appended to many PCW-active glycoside hydrolases that do not possess cellulase activity (Hervé et al., 2010).

B.2.5 The distribution of CBM in lignocellulose-degrading enzymes

Based on sequence data, more than 60% of cellulases and xylanases identified lack CBMs. As a general rule, enzymes containing CBMs can bind to polymeric substrates. On the contrary, enzymes lacking CBMs are often specific for soluble molecules.

It is interesting to note that a number of glycoside hydrolases contain multiple CBMs, either arranged in tandem (Tomme et al., 1996) or at opposing N and C terminal ends (Sermsathanaswadi et al., 2017) of the protein sequence, or both (Guillén et al., 2010; Sainz-Polo et al., 2015). In the case of multiple CBMs, the question of function is particularly interesting. One might suppose that the CBMs target the same carbohydrate ligand, recognizing different motifs therein. On the other hand, if the enzyme acts on a complex substrate such as PCWs, one might imagine that the CBMs recognize different substrates within a wider polysaccharide amalgam. Interestingly, the study of enzymes that bear multiple CBMs reveals that these latter can belong to the same CBM family (homogeneous multimodularity) or different families (heterogeneous multimodularity) (Sainz-Polo et al., 2015). In the case of homogeneous multimodularity, this has been linked to multivalency and avidity effects (Bolam et al., 2001; Boraston et al., 2002, 2006; Freelove et al., 2001). In the case of heterogeneous multimodularity, it has been proposed that this might provide distinct ligand-binding specificities, which is consistent with the case of *Clostridium stercorarium* Xyn10B mentioned in B.2.3.

B.3 Microbial carbohydrate-degrading systems

Carbohydrate utilization is inextricably linked to the ability of microbes to persist in environments as diverse as fresh and salt water, soil, and animal gastrointestinal tracts. In these environments competition between organisms for a common and potentially temporary and limited pool of nutrients is fierce (Grondin et al., 2017). In this respect, the strategy adopted by many bacterial and fungal species is to produce a multitude of enzymes that enable them to access and metabolize complex substrates such as heteroxylans.

Improved culturing techniques and next-generation sequencing have recently provided access to enormous amounts of genetic information and thus are gradually revealing the secrets of various microbial ecosystems or microbiomes (Flint et al., 2008; Human Microbiome Project Consortium et al., 2012; Mayorga et al., 2016; Rossmassler et al., 2015). However, in the current post-genomic era, the urgent need is now related to functional data. This is required to elucidate how the microorganisms within a microbiome operate in order to best use available carbohydrates within a competitive environment that is certainly characterized by competition and cooperation.

Conventionally, the PCW-degrading systems of different microbial species are differentiated on the basis of whether these are produced by aerobic (mostly fungi) and or anaerobic

microbes (mostly bacteria) (Várnai et al., 2013). The majority of PCW-active aerobic organisms, such as *Trichoderma sp.* and *Aspergillus sp.*, secrete a whole host of carbohydrate-active enzymes into their surrounding environments. These enzymes, often operating with considerable synergy, bring about PCW hydrolysis and facilitate the importation of monosaccharides by the microorganism (Tomme et al., 1995; Warren, 1996). In contrast, most of anaerobic Gram positive microorganisms produce an array of hydrolytic enzymes that are secreted to and bound at the cell surface within complex enzyme machineries known as cellulosomes, which also contain CBMs (Gilbert, 2007). On the other hand, some anaerobic Gram negative microbes employ glycan utilization systems (Martens et al., 2009), which are known as Sus-like systems, because the paradigm for these is the well-characterized Starch Utilisation System (Sus).

In general terms, to date the most extensively studied plant fiber degrading enzyme systems are those produced by members of the Firmicutes and Bacteroidetes phyla that are present in gut microbiomes (White et al., 2014).

B.3.1 Cellulosome

The initial definition of the cellulosome was " a discrete, cellulose-binding, multienzyme complex for the degradation of cellulosic substrates" (Bayer et al., 1985, 2004). The principle component of the cellulosome is a scaffoldin subunit, which is composed of contains cohesion modules (usually in multiple copies) for incorporation of different enzymes and other cellulosomal components. Enzymes containing the dockerin modules are loaded onto the cellulosome by tenacious cohesion-dockerin interaction. In some cases, two or more types of scaffoldin may be involved in anchoring the cellulosome to the cell surface. In addition, CBM is commonly found in scaffoldin and plays important roles in substrate binding (*Fig B-5*).

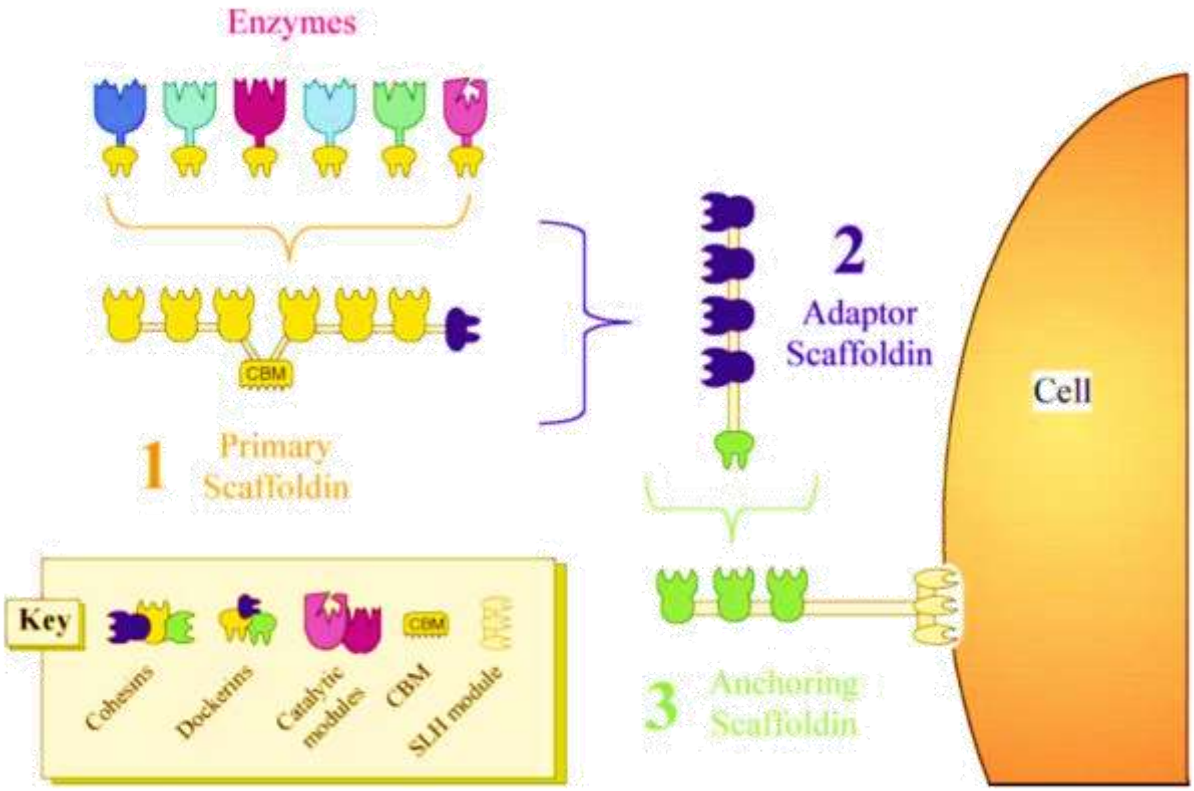


Fig B-5. Schematic representation of the proposed cellulosome architecture in A.cellulolyticus (Bayer et al., 2004).

In class-I (dockerin-bearing) primary scaffoldin, enzymes were incorporated by dockerin-cohesin interaction. An adaptor scaffoldin connects the dockerin of the primary scaffoldin and the cohesins of an anchoring scaffoldin.

The discovery of glycan-specific transporter systems has revealed another paradigm for glycan degradation and uptake. Generally, the examination of Firmicutes genomes reveals that these encode fewer carbohydrate-degrading enzymes than members of Bacteroidetes, but possess more ABC transporters (ATP-binding cassette transporters) and PTS (phosphoenolpyruvate-dependent phosphotransfer systems) that transport carbohydrates (Mahowald et al., 2009). ABC transporter are frequently encoded adjacent to genes encoding glycoside hydrolases, suggesting that these two functions are co-regulated and function together (Rey et al., 2010).

B.3.2 Sus-like systems

The Sus-like systems are only found in Gram negative bacteria of phylum Bacteroidetes. The difference between Gram positive and negative bacteria lies in the composition of their cell envelope. Gram-negative bacteria display two cell membranes, the inner and the outer membranes, that are separated by the periplasm. The outer membrane consists of

phospholipids, lipopolysaccharides, integral membrane proteins, and lipoproteins (Bos et al., 2007). Bacteroidetes genomes typically contain 200-350 GH and polysaccharide lyases genes, compared with less than 200 in those of Firmicutes (White et al., 2014).

Regarding glucose-based substrates, it has been shown that maltodextrins longer than maltotetraose cannot be used by the maltotriose/maltose uptake systems (Martens et al., 2009). Therefore, one might assume that other mechanisms exist to ensure the uptake of more complex glycans for catabolism. Seminal work by members of Prof. Abigail Salyers's laboratory (Reeves et al., 1997) revealed a cell envelope-associated multiprotein system, which was designated by the acronym Sus. This system enables *Bacteroides thetaiotaomicron* to bind and degrade starch (*Fig B-6*). Subsequent microbioal genome sequencing projects revealed that derivatives of this paradigm system (Sus-like systems) are highly represented in many other saccharolytic Bacteroidetes species. A key characteristic of these Sus-like systems is the coordinated action of eight gene products, encoded by the adjacent genes *susRABCDEFG* (*Fig B-6*), which perform functions such as substrate binding proteins and hydrolysis.

The concerted and sequential action of Sus proteins results in a series of glycan recognition, acquisition and degradation operations. Initially, the extracellular starch transits through a polysaccharide capsule to the cell surface (*Fig B-6*, *upper right inset*). SusC is a TonB-dependent receptor, membrane-spanning proteins transporting solutes and macromolecules via energy derived from the proton-motive force and the TonB-ExbBD complex (Ferguson and Deisenhofer, 2002). It is suggested that SusD ligand recognition is driven by recognition of the starch helical backbone rather than by the stereochemistry of the individual glucosyl units (Martens et al., 2009) and that the combined effect of SusE, SusC and SusD accounts for over 80% of the affinity for starch. The remaining affinity can be attributed to SusF (Shipman et al., 2000). The postulated role for SusG, is that of a typical GH13 α-amylase that cleaves internal α-1,4 bonds between glucosyl units thus depolymerizing starch and producing maltodextrins (larger than maltotriose). Then according to current knowledge, the latter are transported into the periplasmic compartment by SusC/SusD (Martens et al., 2009). The localization of SusCDEFG on the outer membrane has been confirmed by immunohistochemical studies (Shipman et al., 2000). In the periplasm, the oligosaccharides produced by SusG are further degraded by SusA and SusB prior to final transport to the cytosol. SusA possesses neopullulanase activity, while SusB is an α-glucosidase. SusR, spans

both the periplasm and cytoplasm and is constitutively expressed and responsible for transcriptional activation of seven sus genes (D'Elia and Salyers, 1996).

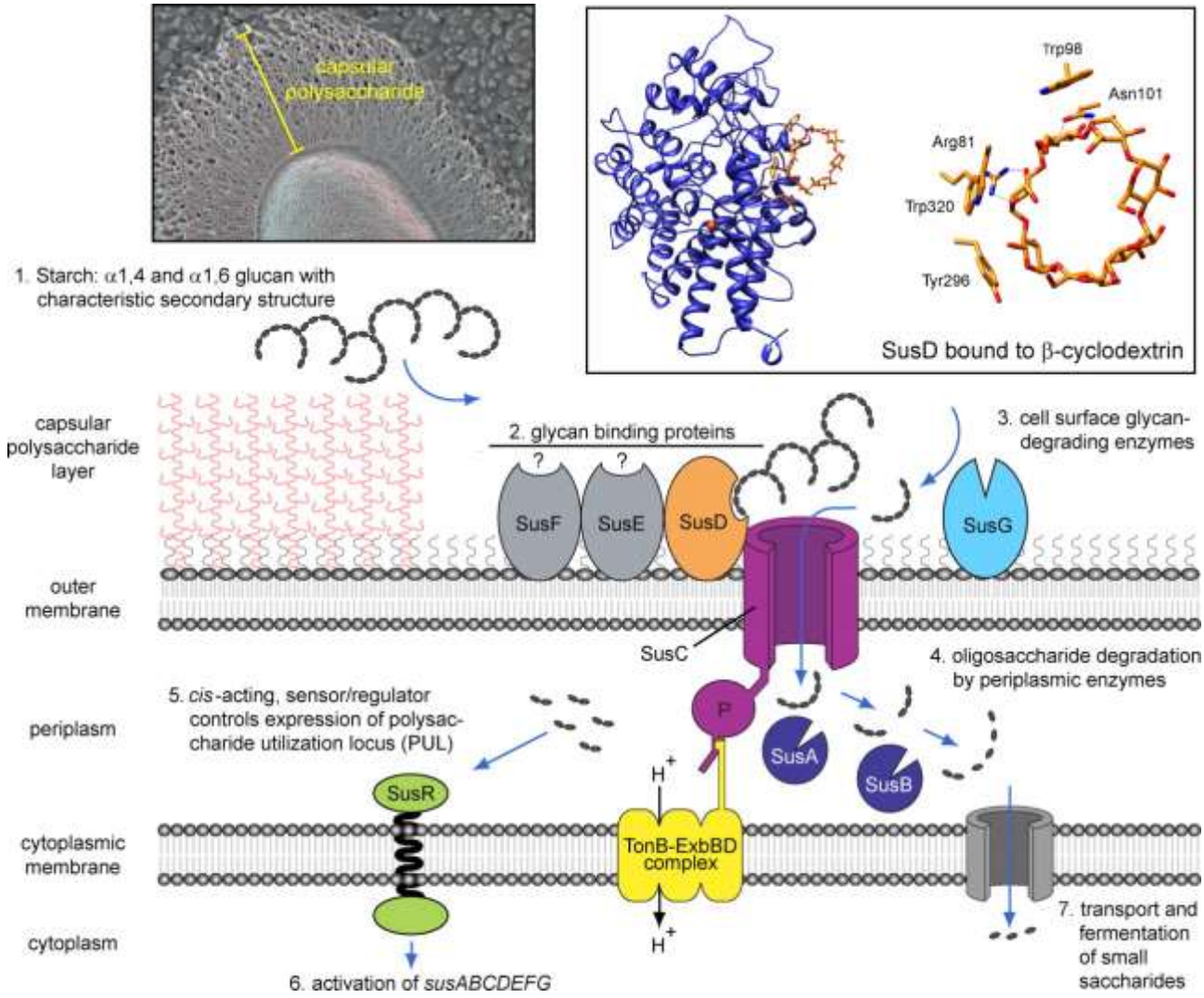


Fig B-6. Functional model of glycan processing based on eight-gene B.thetaiotaomicron starch utilization system (Sus) (Martens et al., 2009).

The *B. thetaiotaomicron* genome contains 101 homologs of *sus*C / *sus*D pairs, suggesting that the similar glycan catabolism strategy may expand to other nutrients (Xu et al., 2007). A PUL (polysaccharide utilization loci), first described in (Bjursell et al., 2006), is a set of physcially linked genes organized around a *sus*CD gene pair (Terrapon et al., 2015). Except *sus*CD, a PUL includes a group of enzymes related to one specific substrate degradation (Bjursell et al., 2006; Martens et al., 2008; Xu, 2003). More generally, Sus-like complexes appear as a paradigm for glycan uptake in bacteria belonging to the phylum Bacteroidetes. The characterized /predicted PULs have been summarized in CAZy database (http://www.cazy.org/PULDB/).

Of the xylan-degrading Bacteroidetes described to date, the xylanolytic system of the rumen bacterium *Prevotella bryantii* B_14 is the best studied (Dodd et al., 2011; Gasparic et al., 1995). This bacterium has been shown to efficiently utilize water-soluble, but not water-insoluble xylans (Miyazaki et al., 1997). The xylan-degradation genes in the same cluster were shown to be induced at the transcriptional level to medium to large-sized xylan fragments via the HTCS regulator, XynR (Miyazaki et al., 2003). This HTCS regulator is also conserved in the genomes of all other xylanolytic Bacteroidetes that contain Xus clusters. Subsequent transcriptomic analysis has shown that this cluster represents part of a large PUL that includes 12 genes that are highly expressed (up to 400-fold induction) during growth on wheat arabinoxylan (Dodd et al., 2010, 2011). These include two pairs of *sus*CD homologues and also an unusual GH10 enzyme (Xyn10C) in which the catalytic module is interrupted by an internal CBM module (Flint et al., 1997). The function of this enzyme has not been established, although it has been suggested that it plays a role similar to SusG in performing limited cell surface cleavage of polysaccharides to produce fragments that can enter the periplasm (Dodd et al., 2010). In vitro biochemical characterization of Xyn10C-like protein from *B.ovatus* was performed and showed moderate activity towards xylan (Wang et al., 2016; Zhang et al., 2014). Additional enzymes like arabinofuranosidase, α-glucuronidase, acetyl xylan esterase encoded either by the large PUL or by additional genes encoded elsewhere on the genome that are also highly induced on xylan (Dodd et al., 2010).

More recently, a more complex model of xylan degradation by *Bacteroide ovatus* was proposed (Rogowski et al., 2015). Different xylan degradation strategies are employed according to the complexity of xylan. It showed that the xylan utilization loci where there is *xyn10C* are dedicated to breakdown relateively simple xylans, such as wheat arabinoxylan instead of corn glucuronoarabinoxylan. What is more, this paper supported the Xyn10C is a surface enzyme and suggested that it may contain an extended substrate bind cleft due to the insertion of two CBMs.

PART C BIOMASS UTILIZATION SYSTEM OF TERMITE

C.1 Introduction to termite

Termites are superabundant in the tropic terrestrial environment and are widely spread from humid forests to the savanna and even arid areas. Termites are notorious for the wood damage they cause, therefore are menaces for crops and houses. On the other hand, their excellence in lignocellulose decomposition can be applied into biomass utilization.

C1.1 Classification of termite

Termites are members of the Animalia kingdom and belong to the of Arthropoda phylum. They form part of the Insecta class the Blattodea order and the Isoptera infraorder. Termites are generally classified into seven families (Table C-1). There is agreement among systematists that Mastotermitidae, a single extant species restricted to northern Australia, is the basal lineage among extant termites (Eggleton, 2011). Termitidae is the largest family, consisting of at least seven sub-families, namely Macrotermitinae, Sphaerotermitinae, Foraminitermitinae, Apicotermitinae, Termitinae, Syntermitinae and Nasutitermitinae (Lo and Eggleton, 2011). In term of species, the number is over three thousand combing the described and under described (Eggleton, 2011).

Termites can be grouped into lower and higher termites based on the presence or absence of flagellated protistan symbionts in the hindgut. Devoid of symbiotic protozoa, Termitidae are higher termites while the remaining six families are referred to as “lower termites”. Higher termites represent over 80% of all termite species (Yoon et al., 2015). Compared to lower termite, higher termite have diversified habitats and diets (Brune, 2014).

Termites of the subfamily Macrotermitinae, which appear to have originated in Africa (Aanen and Eggleton, 2005), are broadly known as termites that live in a complex symbiotic relationship with fungi of the genus *Termitomyces* (Hinze et al., 2002). They are abundant in savannah ecosystems and directly mineralize up to 20% of carbon (Ohkuma, 2003). One of the specie, *P. militaris*, causes significant damage to sugar cane crops in Kenya, Takasdal, and Uganda (Valentin Dibangou, 2012).

Table C-1 Extant families of termites (Lo and Eggleton, 2011)

Family	Food	Habitat	Regions	Number of spp
Mastotermitidae	w	Dry forest/savanna	N Australia	1
Hodotermitidae	g	Arid grasslands	Africa, Palaearctic	15
Termopsidae	ww	Temperate rain forest	Bipolar	
Kalotermitidae	dw	Tropical tree canopy, oceanic islands, deserts, coasts	Cosmopolitan	~400
Rhinotermitidae	w	Rain forest edges and gaps, temperate woodland, oceanic islands, mangroves, deserts	Pantropical, also in sub-tropics, warm temperate	~300
Serritermitidae	nests ?	Cerrado	S America	~2
Termitidae	diverse	Tropical rain forests, savannas, semi-deserts	Pantropical	~ 2,000

w, wood; g, grass; ww, wet wood; dw, dry wood.

Termites are eusocial insects and an individual separated from the colony will die (Eggleton, 2011). Based on cast structure, termites can be classified into alates, queens, kings (reproduction), workers (construction), soldiers (defence), and immature forms (fortification). This obligate association with the colony is mainly because each caste lacks some elements that are present in solitary insects, which lead to the proposition that a termite colony is a superorganism (Eggleton, 2011).

C.1.2 Anatomy

A normal termite has a head, thorax and abdomen. The termite gut generally consists of the foregut (stomatodeum), midgut (mesenteron), and hindgut (proctodeum). Both the foregut and hindgut are ectodermal and covered by cuticle whereas the midgut is of endodermal origin and its epithelial surface can be accessed by microorganisms (Brune and Dietrich, 2015). The ingested lignocellulose underwent mechanical grinding in the foregut, which is composed of an enlarged anterior segment ("the crop") and a posterior segment ("the gizzard"). The endogenous lignocellulases were mostly secreted into the columnar midgut (Eggleton, 2011). The hindgut absorbing gut fluid is more developed and compartmentalized in higher termites than in lower termites (Hongoh, 2011)(FigC-1cd). The hindgut from higher termite is further subdivided into P1 (ileum), P2 (enteric valve), P3 (paunch), P4 (colon) and P5 (rectum) segments (Watanabe and Tokuda, 2010)(FigC-1d). The elongated P3 compartment, is the major site of symbiosis and where oxygen content is steeply shifted from anaerobic (P1-3) to aerobic (P4-5) (Brune et al., 1995). The hindguts in lower termites are relatively short, so as

to that of Macrotermitinae that possesses “fungal gardens” (C1.4). The pH of foregut can be slightly acidic while that of midgut is circumneutral (Brune and Kühl, 1996). Regarding the hindgut, the P1, P2 and P3 are more alkaline than P4 and P5, with the highest pH over 12 (Bignell, 2010).

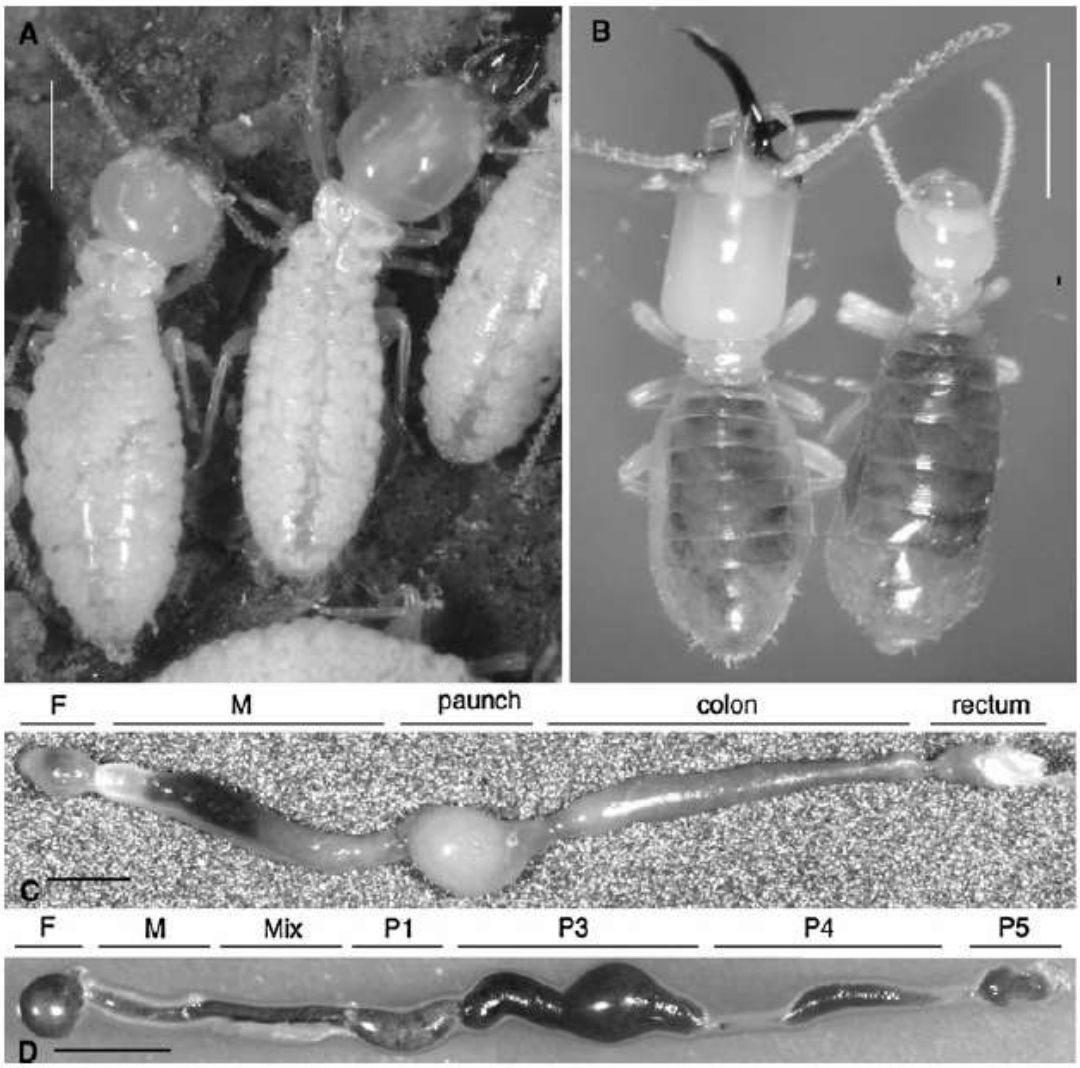


Fig C-1: Termites and their intestinal tract (Hongoh, 2011).

A the lower termites *Coptotermes formosanus*. **B** the interface (wood and humus)-feeding higher termite *Termes comis*. **C** the gut of *C.formosanus*. **D** the gut of *T.comis*. *F* forgut, *M* midgut. *Bars*=1mm.

C.1.4 Digestive strategies

Termite is one of the most efficient lignocellulose-degrading systems (Sun et al., 2014). In terms of hemicellulose, up to 87% of it in the components of lignocellulose ingested can be dissimilated, which are more efficient than in human (up to 72%) (Ohkuma, 2003; Slavin et al., 1981). It has dual enzyme sources, one is from endogenous enzymes, the other is from the microbial inhabitants of the gut compartments (Hongoh, 2011; Nakashima et al., 2002; Watanabe and G. Tokuda, 2001; Watanabe and Tokuda, 2010), which plays a major role in lignocellulose degradation (Breznak and Brune, 1994; Veivers et al., 2006).

In spite of the utilization of enzyme systems, the digestive strategies of termite also include the delicate procedures for degradation. The fungus-growing termites employed a series of procedures to achieve the pretreatment and almost complete degradation of lignocellulose (Hinze et al., 2002; Ohkuma, 2003). The plant litters collected by old workers are firstly pretreated by young workers via mandibles and gizzards and turned to be small fragments (10-20 μm in diameter) (Brune, 2014). The resulting fecal pellets form combs allowing the growth of the symbiotic fungi, which decreases the lignin content therein (Hyodo et al., 2000). The mature fungus nodules are then dissimilated by old workers to a great extent as final feces are rarely found (Liu et al., 2013). In all other higher termites, symbiotic digestion is independent of fungal symbionts. Instead, they rely on different community of bacteria that adapted with the host phylogeny and diet (Brune, 2014).

In lower termites, over 90% of lignocellulose can be assimilated benefiting from employing dual enzyme systems (Ohkuma, 2003). The biomass are partially degraded by endogenous cellulases in salivary gland, foregut and midgut, and then be endocytosed and fermented by the symbiotic systems, i.e., protists and bacteria in hindgut. The symbiotic bacteria play major roles in producing acetate from precursors (Slaytor et al., 1988). Without the symbiotic system, the termites can't grow on lignocellulose (Todaka et al., 2010). In addition, the combined host/symbiont effect is synergistic in glucose liberation and apparently additive in pentose release (Scharf et al., 2011).

C.2 Termite microbial enzymes

The termite gut is a rich reservoir of diverse microorganisms/mutualistic symbionts. The gut microbial community contains all three domains of organisms: Archaea, Bacteria, and Eukaryotes (protists) (Hongoh et al., 2008a). Relatively few microbes are found in the foregut and midgut, whereas the hindgut microbiota is densely populated (Hongoh, 2011; Köhler et al., 2012). In some cases no 16SrRNA genes were found in P3 lumen of *Nasutitermes* species

(Warnecke et al., 2007). A better understanding of termite microbiota is facilitated by "omic" research, which has been reviewed by Scharf (Scharf, 2015a).

C.2.1 Omics studies in termite gut

Using classical microbiology it is impossible to characterize the genomes of most bacterial species, because 99.9% of microorganisms in environmental niches cannot be cultivated using standard laboratory techniques (Madhavan et al., 2017). However, the advent of metagenomics has furnished a way to surmount this problem thus provides a way to probe the genetic resources of complex microbial communities (Ferrer et al., 2009).

As the term metagenomics suggests, the approach involves the isolation and characterization of the genomic patrimony of whole microbial communities. To exploit the information contained in so-called metagenomic libraries, which contain the fragmented metagenome of a given microbiota, one of several strategies can be deployed. First, to establish a repertory of the species present, it is possible to sequence 16s RNA genes using universal primer sets. Second, one can go beyond this using high throughput sequencing to sequence all DNA. This allows large genome fragments and even whole genomes to be reconstructed and protein-encoding genes to be identified. Finally, to target specific functions, one can deploy a functional screening strategy. This involves the confrontation of individual metagenomic library clones with a functional screen (e.g. an enzyme assay to detect cellulase activities). The advantage of this strategy is that positive clones can be fully sequenced and the protein responsible for the detected activity identified, thus establishing a link between sequence and function (Tasse et al., 2010).

C.2.2 Functional screening of metagenomic libraries

To perform functional metagenomics, DNA is isolated directly from a target sample, such as termite gut and then fragmented. The metagenomic libraries are constructed by the cloning of fragmented genomic DNA in various lengths using appropriate expression vectors, such as plasmids (<10 kb), cosmids (25-35 kb), fosmids (25-40 kb), or bacterial artificial chromosomes (100-200 kb) (Uchiyama and Miyazaki, 2009) and then are transformed to a surrogate host, such as *Escherichia.coli*, for expression of the desired genes. Instead of directly sequencing the libraries, they are firstly screened against different polysaccharides (usually chromogenic substrates). Positive clones then undergo another screening to pinpoint the glycosidic linkage specificity within clones that were able to target the same

polysaccharide in the primary screens prior to in-depth pyrosequencing for putative enzymes annotation (Tasse et al., 2010). The main advantage of functional metagenomics is that it avoids unnecessary sequence analysis, especially when clones do not contain genes of interest (e.g. lignocellulase-encoding genes) (Hongoh et al., 2008a, 2008b). However, the drawbacks of functional screening is low success (hit) rate (Uchiyama and Miyazaki, 2009). This is linked to the fact that some genes may not be expressed in the host cell (perhaps because of poor promoter recognition, translational inefficiency, misfolding of proteins, defective post-translational modification of desired proteins, etc), or that they might not be detected in the screen due to low activity or an inappropriate choice of substrate (Lam et al., 2015; Madhavan et al., 2017).

Regarding the use of metagenomics for bioprospection in termites, many omic studies have been performed (*c.f.* Michael E. Scharf 2015). An article published in 2007 reported the results of a metagenomics study conducted on the gut microbiota of *Nasutitermes corniger* and revealed detailed taxonomic and functional data (Warnecke et al., 2007). This watershed study by Warnecke demonstrated that metagenomics is a powerful approach to probe the functionalities of termite microbiota. Importantly, this study established the fact that the gut microbiota of *N. corniger* is a rich reservoir of carbohydrate-active enzymes, since this study alone revealed 700 genes, or gene modules, that display homology to glycoside hydrolases (GHs), with more than 100 of these being putative cellulose or hemicellulose-active enzymes. Since then, the termite gut has been a popular resource for probing new enzymes (Bastien et al., 2013; Brennan et al., 2004; Liu et al., 2011, 2013). A huge quantity of enzymes aside, omics approacesh have also revelaled a variety of enzymes, such as cellulase, hemicellulase, ligninase (Geib et al., 2008) and laccase (Beloqui et al., 2006).

C.2.3 Composition of bacterial groups

In higher termites, the microbial population housed in the hindgut is almost exclusively responsible for enzymatic degradation of lignocellulose. Regarding bacterial diversity in termite guts, for *Reticulitermes speratus* it was estimated that up to 740 phylotypes are present (Hongoh et al., 2003), which indicates that termite hindguts are extremely rich from a biodiversity standpoint. However, the community structure is typically uneven, with just a few numerically predominant phylotypes, associated with a large number of rarer phylotypes. The dominant phyla in termite gut is quite dependent upon the host species, an observation that has led scientists to suggest that termite gut bacteria have coevolved with termites

(Hongoh et al., 2005). To illustrate this point, it is noteworthy that the gut of *Amitermes wheeleri* (Termitidae family, subfamily Amitermitinae), a termite that forages on herbivore dung containing undigested fibrous plant biomass, is characterized by a predominance of Firmicutes and Spirochaetes. In contrast, in *N. corniger* (Termitidae family, subfamily Nasutitermitinae), Spirochaetes and Fibrobacteres dominate (He et al., 2013). Likewise, in the termites *Trinervitermes trinervoides* (Termitidae family, Nasutitermitinae subfamily) and *Pseudacanthotermes militaris* (Termitidae family, subfamily Macrotermitinae) the dominant phyla are Firmicutes and Bacteroidetes (Bastien et al., 2013; Sanyika et al., 2012).

From a general standpoint, six major bacterial phyla are represented across higher and lower termites. These are Bacteroidetes, Firmicutes, Spirochaetes, Proteobacteria, Fibrobacteres, and Elusimicrobia (Brune, 2014). Among them, Bacteroidetes are highly abundant in fungus-growing termites, with clade Bacteroidale cluster V exclusively found in termites and cockroaches (Brune and Dietrich, 2015; Noda et al., 2009; Schauer et al., 2012).

C.2.4 CAZyme of bacterial enzymes

The expression and distribution of lignocellulases in termite gut vary with the termite castes and developmental stages (Fujita et al., 2008; Ni and Tokuda, 2013). Among 45 GH families in the hindgut microbial communities of *Nasustitermes* sp., 7 GH families (5,8,9, 44,45,51 and 74) are putative EGs, whereas GH1 and GH3 might be acting as BGs (Warnecke et al., 2007). CAZyme bioprospecting in the gut of *Coptotermes gestroi* generated more than 12,000 ORFs with predicted functions related to carbohydrate metabolism, 587 of them encoding lignocellulases distributed in 18 GH families (Do et al., 2014). Potential CBHs or components of cellulosomes has not been identified from metagenomic studies (Ni and Tokuda, 2013). It is worth noting that horizontal gene transfer is an important process in gut bacterial evolution and adaptation, and key enzymes probably evolved from a common ancestor and were conserved among the evolution notably concerning CAZymes production (Ricard et al., 2006). So far, the production of recombinant CAZymes of termite gut microbiotal origin has been reported in the literature. These are from 1 GH1, 27 GH5s, 12 GH9, 1 GH45, 1 GH10, 1 GH11, 1GH8 (Scharf, 2015b).

C.3 Endogenous CAZymes of termite

Initially, insects were thought to fully depend on the lignocellulases from symbiotic microbes until endogenous cellulases were found from orders Dictyoptera, Orthoptera, and

Coleoptera (Oppert et al., 2010). The first case was discovered in 1998 in the termite *Reticulitermes speratus*, which retained its ability to feed on wood even when its gut fauna were ablated (Watanabe et al., 1998). The endogenous EGs of termites showed high homology, implying that the endogenous genes were present before the establishment of host/symbionts effect (Nakashima et al., 2002). Compared to lower termites, the endogenous lignocellulases are more important for higher termite, in which case, the endogeneous cellulolytic activity were demonstrated to meet its metabolic requirements (Zhou et al., 2007)(Ni and Tokuda, 2013).

Compared to vast genomic data of termite gut bacteria, only two termite genome sequences are currently available (Scharf, 2015a). Mining endogenous termite enzymes relies mostly on separation from the host and metagenomic studies (Husseneder et al., 2012). Compared to the CAZymes of gut bacteria, the endogenous termite enzymes are less diverse. All endogenous EGs are affiliated with the family GH9, while almost all endogenous BGs belongs to GH1 (Ni and Tokuda, 2013; Watanabe and Tokuda, 2010).

C.4 The potential application in biomass degrading

Wood-chewing insects are miniature factories converting lignocellulose into sugars, which were used to sustain life rather than produce target products such as biofuels. The potential application of the digestive processes/strategies of termite in biomass degradation is striking as up to 99% of cellulose and 87% of hemicellulose can be saccharified (Fischer et al., 2013; Ohkuma, 2003).

Being fascinated by the humble termite's ability to turn wood into energy for life, scientists are focusing on how to advance biochemical biomass processing in industry. Ke et al studied the chewing stage of termite, and proved that the termite mechanical chewing process combines with specific biological pretreatment on biomass, resulting *in vitro* in increased enzymatic cellulose digestibility (Ke et al., 2012). The mimicking of this process could make the pretreatment of biomass more efficient in industrial biomass conversion.

PART D INDUSTRIAL APPLICATIONS OF LIGNOCELLULOSIC ENZYMES

D.1 Biomass and biofuels

D.1.1 Defining the term Biomass in the context of the bioeconomy

World energy needs are currently met mainly by the use of fossil resources, in the form of coal, gas and oil. These resources result from processes that took place over millions of years and in specific circumstances, and which led to the fossilization of plant-based organic matter. Fossil resources constitute massive reservoirs of carbon, which when used by human for energy purposes leads to the release of CO_2 into the atmosphere. For well-known reasons the continued use of fossil resources is considered to be unsustainable and ultimately dangerous for the geoclimatic equilibrium on Earth.

The term biomass is widely used to describe organic matter produced by plants via photosynthesis. Moreover, with increased interest in the use of plant-based resources for industrial processes, the term biomass is often used to specifically describe lignocellulosic matter that is composed of cellulose, lignin and hemicelluloses (Howard et al., 2003)(Table D1). This biomass is the major component of tress, herbaceous plants/grasses, aquatic plants and organic waste, such as manure. This biomass has always been a major source of energy for mankind and is presently estimated to contribute approximately 10-14% of the world's energy supply (McKendry, 2002).

Increasing interest in the use of biomass as a substitute raw material for oil, gas and coal has sparked interest in plants such as switchgrass and poplar trees, non-comestible species that are renowned for fast growth, high biomass yields, their capability to grow on marginal land and their relatively low use of water and nutrients (Sarkar et al., 2009). These and other abundant biomass resources are now widely considered to be important raw materials for a future fossil-free economy, which is now designated by the term bioeconomy.

Table D-1: A wide variety of biomass resources for conversion into bioproducts (Howard et al., 2003).

Lignocellulosic material	Residues	Competing use
Grain harvesting Wheat, rice, oats barley and corn	Straw, cobs, stalks, husks,	Animal feed, burnt as fuel, compost, soil conditioner
Processed grains Corn, wheat, rice, soybean	Waste water, bran,	Animal feed
Fruit and vegetable harvesting	Seeds, peels, husks, stones, rejected whole fruit and juice	Animal and fish feed, some seeds for oil extraction
Fruit and vegetable processing	Seeds, peels, waste water, husks, shells, stones, rejected whole fruit and juice	Animal and fish feed, some seeds for oil extraction
Sugar cane other sugar products	Bagasse	Burnt as fuel
Oils and oilseed plants Nuts, cotton seeds, olives, soybean etc.	Shells, husks, lint, fibre, sludge, presscake, wastewater	Animal feed, fertiliser, burnt fuel
Animal waste	Manure, other waste	Soil conditioners
Forestry-paper and pulp Harvesting of logs	Wood residuals, barks, leaves etc.	Soil conditioners, burnt
Saw-and plywood waste	Woodchips, wood shavings, saw dust	Pulp and paper industries, chip and fibre board
Pulp & paper mills	Fibre waste, sulphite liquor	Reused in pulp and board industry as fuel
Lignocellulose waste from communities	Old newspapers, paper, cardboard, old boards, disused furniture	Small percentage recycled, others burnt
Grass	Unutilised grass	Burnt

D.1.2 Defining the term Biofuels

Using suitable chemical and/or biological operations, biomass can be converted into a whole range of chemicals including energy vectors such as methane, ethanol, butanol and fatty acid methyl esters (FAMEs). During the latter part of the 20th century, much focus was put on the production of FAMEs and ethanol from biomass, with major R&D programs being financed in both the USA and Europe.

Ethanol and FAMEs are of particular interest to industry, because these energy vectors are liquid at ambient temperature and can be used in standard (albeit with minor adjustments) combustion engines. So-called 1st generation biofuels are either ethanol or FAME-based liquid fuels that are currently used in vehicle fleets in Europe, the USA and Brazil. They are referred to as 1st generation, because they were the first biofuels to be commercially produced using technologies that convert either vegetable oils (e.g. sunflower or rapeseed oil) into

FAMEs, or sucrose (sugar cane or sugar beet) and starch-based (Maize and wheat grain) feedstocks into ethanol. Although the production of 1st generation biofuels is commercially viable, concerns related to competition with the food chain have led to a recent decrease (at least in Europe) in their desirability and an increased ambition to transit to 2nd generation biofuels.

Second generation biofuels (especially ethanol) are ones that are produced from non-comestible biomass, such as lignocellulosic biomass, which is highly abundant. Ethanol (10%, v/v) has been used in gasohol or oxygenated fuels since the 1980s in the USA (Howard et al., 2003). Several studies have shown that 2nd generation biofuels could be more sustainable (i.e. better carbon life cycle performance and less competition with food) and that they can be made from a wide variety of biomass feedstocks, including coproducts from agriculture, forestry and industry and dedicated lignocellulosic crops. Nevertheless, as mentioned previously, lignocellulosic biomass is recalcitrant and thus less readily convertible into products of interest. Therefore, despite extensive research during the 20th century, the commercial production of 2nd generation biofuels remains to be achieved on a wide scale, even though some industrial-scale units operate in the USA.

D.2 Plant biomass biorefinery

D.2.1 The concept of biorefinery

Biorefining is an integrated facility that sustainably converts biomass into a range of valuable chemicals and power (Cherubini, 2010).

D.2.2 Biorefinery processes

Ethanol as a transportation fuel is currently being produced over 100 billion liters per year form the readily extractable sugars of corn and sugar cane (Lopes et al., 2016; Sarkar et al., 2009). The production of bioethanol from lignocellulosic biomass requires the implementation of several steps: (1) production and pre-harvest of feedstock; (2) harvest and post-harvest logistics, including storage, pre-processing, handling and transportation to the bioethanol plant gate; (3) conversion to bioethanol (Gnansounou, 2010; Lei et al., 2017). Two main production routes have been identified to procure second generation bioethanol. The first using thermochemical operations, while the second relies heavily on biochemical processes. Much R&D focus has been put on the biochemical route because it is thought that progress in biotechnology will drive strong innovation and offer greater opportunities to

achieve overall sustainability. Although the biochemical route is actually a general term that can be applied to a wide variety of processes, most of these are characterized by some key steps.

The biochemical conversion of biomass usually starts with some form of mechanical process (e.g. chopping, milling etc) that aims to reduce particle size and increase surface area and accessibility (Merino and Cherry, 2007). This is followed by what is known as pretreatment. This is a critical step whose aim is to partially dismantle the macroscopic structure of biomass, increasing accessibility to enzymes and removing or modifying components, such as lignin and perhaps some hemicelluloses (Öhgren et al., 2007; Wyman, 1999). Usually, the product of pretreatment is a cellulose-enriched biomass pulp. Pretreatment is followed by enzymatic deconstruction (Alvira et al., 2013). This step employs enzyme cocktails that breakdown down the polysaccharides components into sugars that are then suitable for fermentation by ethanologenic strains (most frequently *Saccharomyces cerevisiae*)(Van Zyl et al., 2007). The final step is distillation, which furnishes the final product (Lopes et al., 2016).

D.2.3 Potential bioproducts and their applications

In the current petro-economy, over 75% of organic chemicals are produced from five primary oil-based chemicals (petro-platform molecules), which are ethylene, propylene, benzene, toluene and xylene. These are used to synthesize all other chemicals that are required by industry (Howard et al., 2003). Except biofuels (D1.2), plant biomass biorefinery could also bring chemical and material products such as vanillin (4-hydroxy-3-methoxybenzaldehyde), one of the most widely used and important flavouring materials (Zhang et al., 2016) and bio-platform molecules, the carbon-intermediate chemicals (building blocks) in producing value-added chemicals (Cherubini, 2010). In comparison to petro-platform molecules, the bio-platform molecules are more variable, the top 12 of them are four carbon 1,4-diacids (succinic), 2,5-furan dicarboxylic aicd, 3-Hydroxy propionic acid, aspartic acid, glucaric acid, glutamic acid, itaconic acid, levulinic acid, 3-hydroxyburtyrolactone, glycerol, sorbitol, xylitol/arabinitol (Werpy et al., 2004). The added value products derived from succinate include 1,4-butanediol, tetrahydrofuran, γ-butyrolactone, adipic acid, n-methypyrrolidone and linear aliphatic esters (Zeikus et al., 1999).

D.2.4 Challenges in biorefinery

The recalcitrance of biomass, enzyme cost, low yields of fermentable sugars and undesired by-products are challenging issues in plant biomass biorefinery (McCann and Carpita, 2015; Merino and Cherry, 2007; Sarkar et al., 2009).

Recent research addressing these challenges found that the introduction of hemicellulases can enhance ethanol production yields by decreasing competitive inhibition of cellulases by XOS and by providing a better accessibility of the cellulases to their substrates (Zhang and Viikari, 2012; Zhang et al., 2013). To make enzyme deconstruction more economically-viable, the following approaches have been considered: (1) novel enzymes bioprospecting; (2) strain and enzyme engineering; (3) optimizing production and operations related factors such as substrate, culturing conditions, enzymes recycling (Howard et al., 2003; Knauf and Moniruzzaman, 2004). A new route to reduce this cost is consolidated bioprocessing (CBP) - featuring cellulase production, cellulose hydrolysis and fermentation in one step using cellulolytic bacteria (Lynd et al., 2005). In a recent work, *Yarrowia lipolytica*, a yeast strain that can produce biodiesel precursors, was engineered to metabolize cellulose, providing a new host platform for CBP (Guo et al., 2017).

Reference

Aanen, D.D.K., and Eggleton, P. (2005). Fungus-growing termites originated in African rain forest. Curr. Biol. 15, 851–855.

Ali, M.K., Hayashi, H., Karita, S., Goto, M., Kimura, T., Sakka, K., and Ohmiya, K. (2001). Importance of the carbohydrate-binding module of Clostridium stercorarium Xyn10B to xylan hydrolysis. Biosci. Biotechnol. Biochem. 65, 41–47.

Alvira, P., Ballesteros, M., and Negro, M.J. (2013). Progress on enzymatic saccharification technologies for biofuels production. In Biofuel Technologies: Recent Developments, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 145–169.

Andersson, R., Fransson, G., Tietjen, M., and Åman, P. (2009). Content and molecular-weight distribution of dietary fiber components in whole-grain rye flour and bread. J. Agric. Food Chem. 57, 2004–2008.

Arai, T., Araki, R., Tanaka, A., Karita, S., Kimura, T., Sakka, K., and Ohmiya, K. (2003). Characterization of a cellulase containing a family 30 carbohydrate-binding module (CBM) derived from Clostridium thermocellum CelJ: Importance of the CBM to cellulose hydrolysis. J. Bacteriol. 185, 504–512.

Araki, R., Ali, M.K., Sakka, M., Kimura, T., Sakka, K., and Ohmiya, K. (2004). Essential role of the family-22 carbohydrate-binding modules for β-1,3-1,4-glucanase activity of Clostridium stercorarium Xyn10B. FEBS Lett. 561, 155–158.

Aroul-Selvam, R., Hubbard, T., Sasidharan, R., Aroul-Selvam, R., Hubbard, T., and Sasidharan, R. (2004). Domain insertions in protein structures. J. Mol. Biol. 338, 633–641.

Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H., and Henrissat, B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol. Biol. 12, 186.

Aspinall, G.O. (1981). Constitution of Plant Cell Wall Polysaccharides. In Plant Carbohydrates II, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 3–8.

Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., and Wilson, I.A. (1975). structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution. Nature 255, 516–517.

Banoub, J., Delmas, G.H., Joly, N., Mackenzie, G., Cachet, N., Benjelloun-Mlayah, B., and Delmas, M. (2015). A critique on the structural analysis of lignins and application of novel tandem mass spectrometric strategies to determine lignin sequencing. J. Mass Spectrom. 50, 5–48.

Barratt, M.J., Lebrilla, C., Shapiro, H.-Y., and Gordon, J.I. (2017). The Gut Microbiota, Food Science, and Human Nutrition: A Timely Marriage. Cell Host Microbe 22, 134–141.

Bastien, G.G., Arnal, G.G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., Henrissat, B., Lefèvre, F., Robe, P., Bouchez, O., et al. (2013). Mining for hemicellulases in the fungus-growing termite Pseudacanthotermes militaris using functional metagenomics. Biotechnol. Biofuels 6, 78.

Baucher, M., Halpin, C., Petit-Conil, M., and Boerjan, W. (2003). Lignin: genetic engineering and impact on pulping.

Bayer, E.A., Setter, E., and Lamed, R. (1985). Organization and distribution of the cellulosome in Clostridium thermocellum. J. Bacteriol. 163, 552–559.

Bayer, E.A., Belaich, J.P., Shoham, Y., and Lamed, R. (2004). The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu. Rev. Microbiol. 58, 521–554.

Beldman, G., Schols, H.A., Pitson, S.M., Searle-van Leeuwen, M.F., and Voragen, A.G.J. (1997). Arabinans and arabinan degrading enzymes. In Advances in Macromolecular Carbohydrate Research, pp. 1–64.

Beloqui, A., Pita, M., Polaina, J., Martínez-Arias, A., Golyshina, O. V., Zumárraga, M., Yakimov, M.M., García-Arellano, H., Alcalde, M., Fernández, V.M., et al. (2006). Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen: Biochemical properties, structural analysis, and phylogenetic relationships. J. Biol. Chem. 281, 22933–22942.

Berglund, J., Toriz, G., Gatenholm, P., Wohlert, J., and Vilaplana, F. (2017). Regular Motifs in Xylan Modulate Molecular Flexibility and Interactions with Cellulose Surfaces. Plant Physiol. 175, 1579–1592.

Biely, P., MacKenzie, C.R., Puls, J., and Schneider, H. (1986). Cooperativity of Esterases and Xylanases in the Enzymatic Degradation of Acetyl Xylan. Bio/Technology 4, 731–733.

Biely, P., Vršanská, M., Tenkanen, M., and Kluepfel, D. (1997). Endo-β-1,4-xylanase families: Differences in catalytic properties. J. Biotechnol. 57, 151–166.

Bignell, D.E. (2010). Morphology, Physiology, Biochemistry and Functional Design of the Termite Gut: An Evolutionary Wonderland. In Biology of Termites: A Modern Synthesis, (Dordrecht: Springer Netherlands), pp. 375–412.

Bjursell, M.K., Martens, E.C., and Gordon, J.I. (2006). Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. J. Biol. Chem. 281, 36269–36279.

Black, G.W., Rixon, J.E., Clarke, J.H., Hazlewood, G.P., Theodorou, M.K., Morris, P., and Gilbert, H.J. (1996). Evidence that linker sequences and cellulose-binding domains enhance the activity of hemicellulases against complex substrates. Biochem. J. 319, 515–520.

Blum, D.L., Li, X.L., Chen, H., and Ljungdahl, L.G. (1999). Characterization of an acetyl xylan esterase from the anaerobic fungus Orpinomyces sp. strain PC-2. Appl. Environ. Microbiol. 65, 3990–3995.

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. Annu. Rev. Plant Biol. 54, 519–546.

Bolam, D.N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M.P., Rixon, J.E., Boraston, A., Hazlewood, G.P., and Gilbert, H.J. (1998). Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem. J. 331, 775–781.

Bolam, D.N., Xie, H., White, P., Simpson, P.J., Hancock, S.M., Williamson, M.P., and Gilbert, H.J. (2001). Evidence for synergy between family 2b carbohydrate binding modules in Cellulomonas fimi Xylanase 11A. Biochemistry 40, 2468–2477.

Bolam, D.N., Xie, H., Pell, G., Hogg, D., Galbraith, G., Henrissat, B., and Gilbert, H.J. (2004). X4 modules represent a new family of carbohydrate-binding modules that display novel properties. J. Biol. Chem. 279, 22953–22963.

Boraston, A.B., Mclean, B.W., Chen, G., Li, A., Warren, R.A.J., and Kilburn, D.G. (2002). Co-operative binding of triplicate carbohydrate-binding modules from a thermophilic xylanase. Mol. Microbiol. 43, 187–194.

Boraston, A.B., Bolam, D.N., Gilbert, H.J., Davies, G.J., Zverlov, V. V, Volkov, I.Y., Velikodvorskaya, G.A., and Schwarz, W.H. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. 382, 769–781.

Boraston, A.B., Healey, M., Klassen, J., Ficko-Blean, E., Van Bueren, A.L., and Law, V. (2006). A structural and functional analysis of α-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. J. Biol. Chem. 281, 587–598.

Bornscheuer, U., Buchholz, K., and Seibel, J. (2014). Enzymatic degradation of (ligno)cellulose. Angew. Chemie - Int. Ed. 53, 10876–10893.

Bos, M.P., Robert, V., and Tommassen, J. (2007). Biogenesis of the gram-negative bacterial outer membrane. Annu. Rev. Microbiol. 61, 191–214.

Brändén, C.-I. (1991). The TIM barrel—the most frequently occurring folding motif in proteins. Curr. Opin. Struct. Biol. 1, 978–983.

Brennan, Y., Callen, W.N., Christoffersen, L., Dupree, P., Goubet, F., Healey, S., Hernández, M., Keller, M., Li, K., Palackal, N., et al. (2004). Unusual microbial xylanases from insect guts. Appl. Environ. Microbiol. 70, 3609–3617.

Breznak, J.A., and Brune, A. (1994). Role of microorganisms in the digestion of lignocellulose by termites. Annu. Rev. Entomol. 39.

Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. Nat. Rev. Microbiol. 12, 168–180.

Brune, A., and Dietrich, C. (2015). The Gut Microbiota of Termites: Digesting the Diversity in the Light of Ecology and Evolution. Annu. Rev. Microbiol. 69, 145–166.

Brune, A., and Kühl, M. (1996). pH profiles of the extremely alkaline hindguts of soil-feeding termites (Isoptera: Termitidae) determined with microelectrodes. J. Insect Physiol. 42, 1121–1127.

Brune, A., Emerson, D., and Breznak, J.A. (1995). The termite gut microflora as an oxygen sink: Microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. Appl. Environ. Microbiol. 61, 2681–2687.

Buchmann, S.L., and McCarthy, A.J. (1991). Purification and cooperative activity of enzymes constituting the xylan-degrading system of Thermomosospora fusca. Appl Env. Microbiol 57, 2121–2130.

Buisson, G., Duée, E., Haser, R., and Payan, F. (1987). Three dimensional structure of porcine pancreatic alpha-amylase at 2.9 A resolution. Role of calcium in structure and activity. EMBO J. 6, 3909–3916.

Bujang, N.S., Harrison, N. a., and Su, N.-Y.Y. (2014). A phylogenetic study of endo-beta-1,4-glucanase in higher termites. Insectes Soc. 61, 29–40.

Buranov, A.U., and Mazza, G. (2008). Lignin in straw of herbaceous crops. Ind. Crops Prod. 8, 237–259.

Burstein, T., Shulman, M., Jindou, S., Petkun, S., Frolow, F., Shoham, Y., Bayer, E.A., and Lamed, R. (2009). Physical association of the catalytic and helper modules of a family-9 glycoside hydrolase is essential for activity. FEBS Lett. 583, 879–884.

Busse-Wicher, M., Li, A., Silveira, R.L., Pereira, C.S., Tryfona, T., Gomes, T.C.F., Skaf, M.S., and Dupree, P. (2016). Evolution of xylan substitution patterns in gymnosperms and angiosperms: implications for xylan interaction with cellulose. Plant Physiol. 171, 2418–2431.

Caffall, K.H., Mohnen, D., Hosmer, K., Mohnen, D., Caffall, K.H., Mohnen, D., Hosmer, K., and Mohnen, D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydr. Res. 344, 1879–1900.

Cani, P.D., and Everard, A. (2016). Talking microbes: When gut bacteria interact with diet and host organs. Mol. Nutr. Food Res. 60, 58–66.

Cantarel, B.I., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. 37, 233–238.

Carpita, N.C., and Gibeaut, D.M. (1993). Structural models of primary cell walls in flowering plants : consistency of molecular structure with the physical properties of the walls during growth. Plant J. 3, 1–30.

Chanliaud, E., Silva, J. De, Strongitharm, B., Jeronimidis, G., and Gidley, M.J. (2004). Mechanical effects of plant cell wall enzymes on cellulose / xyloglucan composites. Plant J. 38, 27–37.

Charles, E.A. (1901). On the origin and nature of the middle lamella. Bot. Gaz. 32, 1–34.

Charnock, S.J., Spurway, T.D., Xie, H., Beylott, M.H., Virden, R., Warren, R. a J., Hazlewood, G.P., and Gilbert, H.J. (1999). The topology of the substrate binding clefts of glycosyl hydrolase family 10 xylanases are not conserved. J. Biol. Chem. 273, 32187–32199.

Charnock, S.J., Derewenda, U., Derewenda, Z.S., Dauter, Z., Dupont, C., Morosoli, R., Kluepfel, D., and Davies, G.J. (2000). Substrate Specificity in Glycoside Hydrolase Family 10. J. Biol. Chem. 275, 23020–23026.

Chen, H. (2014). Chemical Composition and Structure of Natural Lignocellulose. In Biotechnology of Lignocellulose, (Dordrecht: Springer Netherlands), pp. 25–71.

Cherubini, F. (2010). The biorefinery concept : Using biomass instead of oil for producing energy and chemicals. Energy Convers. Manag. 51, 1412–1421.

Chesson, A., Gordon, A.H., and Lomax, J.A. (1983). Substituent groups linked by alkali - labile bonds to arabinose and xylose residues of legume, grass and cereal straw cell walls and their fate during digestion by rumen microorganisms. J. Sci. Food Agric. 34, 1330–1340.

Claeyssen, M. (1966). Transfer reactions catalysed by a fungal beta-d-xylosidase:enzymic synthesis of phenl beta-d-xylobioside. Carbohydr. Res. 3, 32–37.

Claeyssens, M., and De Bruyne, C.K. (1965). d-Xylose-derivatives with sulfur or nitrogen in the ring: Powerful inhibitors of glycosidase-activities. Naturwissenschaften 52, 515.

Collins, T., Gerday, C., and Feller, G. (2005). Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiol. Rev. 29, 3–23.

Colvin, J.R. (1981). Ultrastructure of the Plant Cell Wall: Biophysical Viewpoint. In Plant Carbohydrates II, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 9–24.

Copley, R.R., and Bork, P. (2000). Homology among (βα)8 barrels: implications for the evolution of metabolic pathways. J. Mol. Biol. 303, 627–641.

Cosgrove, D.J. (2005). Growth of the plant cell wall. Nat. Rev. Mol. Cell Biol. 6, 850–861.

Crouch, L.I., Labourel, A., Walton, P.H., Davies, G.J., and Gilbert, H.J. (2016). The contribution of non-catalytic carbohydrate binding modules to the activity lytic polysaccharide monooxygenases. J. Biol. Chem. 291, 7439–7449.

Cuskin, F., Flint, J.E., Gloster, T.M., Morland, C., Baslé, A., Henrissat, B., Coutinho, P.M., Strazzulli, A., Solovyova, A.S., Davies, G.J., et al. (2012). How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. Proc. Natl. Acad. Sci. U. S. A. 109, 20889–20894.

D'Elia, J.N., and Salyers, A.A. (1996). Effect of regulatory protein levels on utilization of starch by Bacteroides thetaiotaomicron. J. Bacteriol. 178, 7180–7186.

Davies, G.J., Wilson, K.S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem. J. 321 (Pt 2, 557–559.

Davies, G.J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, a. M., Schülein, M., and Withers, S.G. (1998). Snapshots along an Enzymatic Reaction Coordinate: Analysis of a Retaining β-Glycoside Hydrolase † , ‡. Biochemistry 37, 11707–11713.

Davison, B.H., Parks, J., Davis, M.F., and Donohoe, B.S. (2013). Plant Cell Walls: Basics of Structure, Chemistry, Accessibility and the Influence on Conversion.

Dekker, R.F.H. (1983). Bioconversion of hemicellulose: Aspects of hemicellulase production by Trichoderma reesei QM 9414 and enzymic saccharification of hemicellulose. Biotechnol. Bioeng. 25, 1127–1146.

DeMartini, J.D., Pattathil, S., Miller, J.S., Li, H., Hahn, M.G., and Wyman, C.E. (2013). Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. Energy Environ. Sci. 6, 898.

Derewenda, U., Swenson, L., Green, R., Wei, Y., Morosoli, R., Shareck, F., Kluepfel, D., and Derewenda, Z.S. (1994). Crystal structure, at 2.6-Å resolution, of the Streptomyces lividans xylanase A, a member of the F family of beta-1,4-D-glycanases. J. Biol. Chem. 269, 20811–20814.

Dhugga, K.S. (2007). Maize biomass yield and composition for biofuels. Crop Sci. 47, 2211–2227.

Dies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure 3, 853–859.

Din, N., Damude, H.G., Gilkes, N.R., Miller, R.C., Warren, R. a, and Kilburn, D.G. (1994). C1-Cx revisited: intramolecular synergism in a cellulase. Proc. Natl. Acad. Sci. U. S. A. 91, 11383–11387.

Do, T.H., Nguyen, T.T., Nguyen, T.N., Le, Q.G., Nguyen, C., Kimura, K., and Truong, N.H. (2014). Mining biomass-degrading genes through Illumina-based de novo sequencing and metagenomic analysis of free-living bacteria in the gut of the lower termite Coptotermes gestroi harvested in Vietnam. J. Biosci. Bioeng. 118, 665–671.

Dodd, D., Moon, Y.-H.H., Swaminathan, K., Mackie, R.I., and Cann, I.K.O. (2010). Transcriptomic analyses of xylan degradation by Prevotella bryantii and insights into energy acquisition by xylanolytic bacteroidetes. J. Biol. Chem. 285, 30261–30273.

Dodd, D., Mackie, R.I., and Cann, I.K.O. (2011). Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. Mol. Microbiol. 79, 292–304.

Dumon, C., Song, L., Bozonnet, S., Fauré, R., O'Donohue, M.J., Faure, R., and O'Donohue, M.J. (2012). Progress and future prospects for pentose-specific biocatalysts in biorefining. Process Biochem. 47, 346–357.

Ebringerová, A., and Heinze, T. (2000). Xylan and xylan derivatives - Biopolymers with valuable properties, 1: Naturally occurring xylans structures, isolation procedures and properties. Macromol. Rapid Commun. 21, 542–556.

Eggleton, P. (2011). An introduction to termites: Biology, taxonomy and functional morphology. In Biology of Termites: A Modern Synthesis, (Dordrecht: Springer Netherlands), pp. 1–26.

Eijsink, V.G.H., Vaaje-Kolstad, G., Vårum, K.M., and Horn, S.J. (2008). Towards new enzymes for biofuels: lessons from chitinase research. Trends Biotechnol. 26, 228–235.

Farber, G.K., and Petsko, G.A. (1990). The evolution of α/β barrel enzymes. Trends Biochem. Sci. 15, 228–234.

Faulds, C.B., and Williamson, G. (1991). The purification and characterization of 4-hydroxy-3-methoxycinnamic (ferulic) acid esterase from Streptomyces olivochromogenes. J. Gen. Microbiol. 137, 2339–2345.

Faulds, C.B., Mandalari, G., Lo Curto, R.B., Bisignano, G., Christakopoulos, P., and Waldron, K.W. (2006). Synergy between xylanases from glycoside hydrolase family 10 and family 11 and a feruloyl esterase in the release of phenolic acids from cereal arabinoxylan. Appl. Microbiol. Biotechnol. 71, 622–629.

Ferguson, A.D., and Deisenhofer, J. (2002). TonB-dependent receptors-structural perspectives. Biochim. Biophys. Acta 1565, 318–332.

Ferrer, M., Beloqui, A., Timmis, K.N., and Golyshin, N. (2009). Metagenomics for Mining New Genetic Resources of Microbial Communities. J Mol Microbiol Biotechnol 16, 109–123.

Fischer, R., Ostafe, R., and Twyman, R.M. (2013). Cellulases from insects. In Advances in Biochemical Engineering/Biotechnology, pp. 51–64.

Flint, H.J., Whitehead, T.R., Martin, J.C., and Gasparic, A. (1997). Interrupted catalytic domain structures in xylanases from two distantly elated strains of Prevotella ruminicola. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1337, 161–165.

Flint, H.J., Bayer, E. a, Rincon, M.T., Lamed, R., and White, B. a (2008). Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat. Rev. Microbiol. 6, 121–131.

Freelove, A.C.J., Bolam, D.N., White, P., Hazlewood, G.P., and Gilbert, H.J. (2001). A Novel Carbohydrate-binding Protein Is a Component of the Plant Cell Wall-degrading Complex of Piromyces equi. J. Biol. Chem. 276, 43010–43017.

Fujita, A., Miura, T., and Matsumoto, T. (2008). Differences in cellulose digestive systems among castes in two termite lineages. Physiol. Entomol. 33, 73–82.

Gal, L., Gaudin, C., Belaich, A., Pages, S., Tardif, C., and Belaich, J.P. (1997). CelG from Clostridium cellulolyticum: A multidomain endoglucanase acting efficiently on crystalline cellulose. J. Bacteriol. 179, 6595–6601.

Gasparic, A., Martin, J., Daniel, A.S., Flint, H.J., Gasparic, A., Martin, J., and Daniel, A.S. (1995). A xylan hydrolase gene cluster in Prevotella ruminicola B (1) 4 : sequence relationships , synergistic interactions , and oxygen sensitivity of a novel enzyme with exoxylanase and beta- (1 , 4) -xylosidase activities . A Xylan Hydrolase Gene Cluster i. Microbiology 61, 2958–2964.

Gebler, J., Gilkes, N.R., Claeyssens, M., Wilson, D.B., Beguin, P., Wakarchuk, W.W., Kilburn, D.G., Miller, R.C., Warren, R.A.J., and Withers, S.G. (1992). Stereoselective hydrolysis catalyzed by related β-1,4-glucanases and β- 1,4-xylanases. J. Biol. Chem. 267, 12559–12561.

Geib, S.M., Filley, T.R., Hatcher, P.G., Hoover, K., Carlson, J.E., Jimenez-Gasco, M.D.M., Nakagawa-Izumi, A., Sleighter, R.L., and Tien, M. (2008). Lignin degradation in wood-feeding insects. Proc. Natl. Acad. Sci. U. S. A. 105, 12932–12937.

Ghosh, A., Luis, A.S., Bras, J.L.A., Pathaw, N., Chrungoo, N.K., Fontes, C.M.G.A., and Goyal, A. (2013). Deciphering ligand specificity of a Clostridium thermocellum family 35

carbohydrate binding module (CtCBM35) for gluco- and galacto- Substituted mannans and its calcium induced stability. PLoS One 8.

Gibson, L.J. (2012). The hierarchical structure and mechanics of plant materials. J. R. Soc. Interface 9, 2749–2766.

Gilad, R., Rabinovich, L., Yaron, S., Bayer, E.A., Lamed, R., Gilbert, H.J., and Shoham, Y. (2003). Cell, a noncellulosomal family 9 enzyme from Clostridium thermocellum, is a processive enduglucanase that degrades crystalline cellulose. J. Bacteriol. 185, 391–398.

Gilbert, H.J. (2007). Cellulosomes: Microbial nanomachines that display plasticity in quaternary structure. Mol. Microbiol. 63, 1568–1576.

Gilbert, H.J. (2010). The biochemistry and structural biology of plant cell wall deconstruction. Plant Physiol. 153, 444–455.

Gilbert, H.J., Hall, J., Hazlewood, G.P., and Ferreira, L.M. (1990). The N-terminal region of an endoglucanase from Pseudomonas fluorescens subspecies cellulosa constitutes a cellulose-binding domain that is distinct from the catalytic centre. Mol. Microbiol. 4, 759–767.

Gilkes, N.R., Warren, R.A., Miller Jr., R.C., Kilburn, D.G., Miller, R.C., and Kilburn, D.G. (1988). Precise excision of the cellulose binding domains from two Cellulomonas fimi cellulases by a homologous protease and the effect on catalysis. J Biol Chem 263, 10401–10407.

Gnansounou, E. (2010). Production and use of lignocellulosic bioethanol in Europe: Current situation and perspectives. Bioresour. Technol. 101, 4842–4850.

Goldman, A.D., Beatty, J.T., Landweber, L.F., David, A., Joshua, G., and Laura, T.B. (2016). The TIM Barrel Architecture Facilitated the Early Evolution of Protein-Mediated Metabolism. J. Mol. Evol. 82, 17–26.

Gong, L., Cao, W., Chi, H., Wang, J., Zhang, H., Liu, J., and Sun, B. (2018). Whole cereal grains and potential health effects: Involvement of the gut microbiota. Food Res. Int. 103, 84–102.

Grabber, J.H. (2005). How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. Crop Sci. 45, 820–831.

Grabber, J.H., Ralph, J., and Hatfield, R.D. (2000). Cross-linking of maize walls by ferulate dimerization and incorporation into lignin. J. Agric. Food Chem. 48, 6106–6113.

Grondin, J.M., Tamura, K., Déjean, G., Abbott, D.W., and Brumer, H. (2017). Polysaccharide Utilization Loci: Fuelling microbial communities. J. Bacteriol. 199, e00860-16.

Guillén, D., Sánchez, S., Rodríguez-Sanoja, R., Guillen, D., Sanchez, S., and Rodriguez-Sanoja, R. (2010). Carbohydrate-binding domains: multiplicity of biological roles. Appl. Microbiol. Biotechnol. 85, 1241–1249.

Guo, Z., Duquesne, S., Bozonnet, S., Cioci, G., Nicaud, J.-M., Marty, A., and O'Donohue, M.J. (2017). Conferring cellulose-degrading ability to Yarrowia lipolytica to facilitate a consolidated bioprocessing approach. Biotechnol. Biofuels 10, 132.

Hachem, M.A., Karlsson, E.N., Bartonek-Roxâ, E., Raghothama, S., Simpson, P.J., Gilbert, H.J., Williamson, M.P., and Holst, O. (2000). Carbohydrate-binding modules from a thermostable Rhodothermus marinus xylanase: cloning, expression and binding studies. Biochem. J. 345, 53–60.

Harris, G.W., Jenkins, J. a, Connerton, I., Cummings, N., Lo Leggio, L., Scott, M., Hazlewood, G.P., Laurie, J.I., Gilbert, H.J., and Pickersgill, R.W. (1994). Structure of the catalytic core of the family F xylanase from Pseudomonas fluorescens and identification of the xylopentaose-binding sites. Structure 2, 1107–1116.

Hashimoto, H. (2006). Recent structural studies of carbohydrate-binding modules. Cell. Mol. Life Sci. 63, 2954–2967.

He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R.H., Kyrpides, N.C., Warnecke, F., Tringe, S.G., and Hugenholtz, P. (2013). Comparative Metagenomic and Metatranscriptomic Analysis of Hindgut Paunch Microbiota in Wood- and Dung-Feeding Higher Termites. PLoS One 8.

Henrissat, B. (1991). A classification of glycosyl hydrolases based sequence similarities. Biochem. J. 280, 309–316.

Henrissat, B., and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. 316 (Pt 2, 695–696.

Henshaw, J., Horne-Bitschy, A., Van Bueren, A.L., Money, V.A., Bolam, D.N., Czjzek, M., Ekborg, N.A., Weiner, R.M., Hutcheson, S.W., Davies, G.J., et al. (2006). Family 6 carbohydrate binding modules in β-agarases display exquisite selectivity for the non-reducing termini of agarose chains. J. Biol. Chem. 281, 17099–17107.

Hervé, C., Rogowski, A., Blake, A.W., Marcus, S.E., Gilbert, H.J., Knox, J.P., Herve, C., Rogowski, A., Blake, A.W., Marcus, S.E., et al. (2010). Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. Proc. Natl. Acad. Sci. U. S. A. 107, 15293–15298.

Hinze, B., Crailsheim, K., and Leuthold, R.H. (2002). Polyethism in food processing and social organisation in the nest of Macrotermes bellicosus (Isoptera, Termitidae). Insectes Soc. 49, 31–37.

Hogan, M., Veivers, P.C., Slaytor, M., and Czolij, R.T. (1988). The site of cellulose breakdown in higher termites (Nasutitermes walkeri and Nasutitermes exitiosus). J. Insect Physiol. 34, 891–899.

Hong, P.Y., Iakiviak, M., Dodd, D., Zhang, M., Mackie, R.I., and Cann, I. (2014). Two new xylanases with different substrate specificities from the human gut bacterium Bacteroides intestinalis DSM 17393. Appl. Environ. Microbiol. 80, 2084–2093.

Hongoh, Y. (2011). Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. Cell. Mol. Life Sci. 68, 1311–1325.

Hongoh, Y., Ohkuma, M., and Kudo, T. (2003). Molecular analysis of bacterial microbiota in the gut of the termite Reticulitermes speratus (Isoptera; Rhinotermitidae). FEMS Microbiol. Ecol. 44, 231–242.

Hongoh, Y., Deevong, P., Inoue, T., Moriya, S., Trakulnaleamsai, S., Ohkuma, M., Noparatnaraporn, N., and Kudo, T. (2005). Intra- and Interspecific Comparisons of Bacterial Diversity and Community Structure Support Coevolution of Gut Microbiota and Termite Host Intra- and Interspecific Comparisons of Bacterial Diversity and Community Structure Support Coevolution of Gut Micro. Appl. Environ. Microbiol. 71, 6590–6599.

Hongoh, Y., Vineet K., S., Tulika, P., Satoko, N., Todd D., T., Toshiaki, K., Yoshiyuki, S., Atsushi, T., Hattori, M., and Moriya, O. (2008a). Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. Proc. Natl. Acad. Sci. U. S. A. 105, 5555–5560.

Hongoh, Y., Sharma, V.K., Prakash, T., Noda, S., Toh, H., Taylor, T.D., Kudo, T., Sakaki, Y., Toyoda, A., Hattori, M., et al. (2008b). Genome of an endosymbiont coupling N2 fixation to cellulolysis within protist cells in termite gut. Science (80-.). 322, 1108–1109.

Hövel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y., and Schomburg, D. (2003). Crystal structure and snapshots along the reaction pathway of a family 51 α-L-arabinofuranosidase. EMBO J. 22, 4922–4932.

Howard, R.L., Abotsi, E., Jansen van Rensburg, E.L., and Howard, S. (2003). Lignocellulose biotechnology: issues of bioconversion and enzyme production. African J. Biotechnol. 2, 602–619.

Huang, P.-S., Feldmeier, K., Parmeggiani, F., Fernandez Velasco, D.A., Höcker, B., and Baker, D. (2016). De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. Nat. Chem. Biol. 12, 29–34.

Human Microbiome Project Consortium, T., Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., et al. (2012). Structure, function and diversity of the healthy human microbiome. Nature 486, 207–214.

Husseneder, C., McGregor, C., Lang, R.P., Collier, R., and Delatte, J. (2012). Transcriptome profiling of female alates and egg-laying queens of the Formosan subterranean termite. Comp. Biochem. Physiol. - Part D Genomics Proteomics 7, 14–27.

Hyodo, F., Inoue, T., Azuma, J.I., Tayasu, I., and Abe, T. (2000). Role of the mutualistic fungus in lignin degradation in the fungus-growing termite Macrotermes gilvus (Isoptera; Macrotermitinae). Soil Biol. Biochem. 32, 653–658.

Itoh, Y., Watanabe, J., Fukada, H., Mizuno, R., Kezuka, Y., Nonaka, T., and Watanabe, T. (2006). Importance of Trp59 and Trp60 in chitin-binding, hydrolytic, and antifungal activities of Streptomyces griseus chitinase C. Appl. Microbiol. Biotechnol. 72, 1176–1184.

Jamal-Talabani, S., Boraston, A.B., Turkenburg, J.P., Tarbouriech, N., Ducros, V.M.A., and Davies, G.J. (2004). Ab initio structure determination and functional characterization of CBM36: A new family of calcium-dependent carbohydrate binding modules. Structure 12, 1177–1187.

Janecek, S. (1997). Alpha-Amylase Family: Molecular Biology and Evolution. J. Prog. Biophys. Mol. Biol. 67, 67–97.

Jenkins, J., Lo Leggio, L., Harris, G., and Pickersgill, R. (1995). Beta-glucosidase, beta-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold beta/alpha architecture and with two conserved glutamates near the carboxy-terminal ends of beta-strands four. FEBS Lett. 362, 281–285.

Jespersen, H.M., Ann MacGregor, E., Henrissat, B., Sierks, M.R., and Svensson, B. (1993). Starch- and glycogen-debranching and branching enzymes: Prediction of structural features of the catalytic (β/α)8-barrel domain and evolutionary relationship to other amylolytic enzymes. J. Protein Chem. 12, 791–805.

Jonathan Knowles (1987). Cellulase families and their genes. Elsevier Publ. 5, 255–261.

Joshi, M.D., Sidhu, G., Pot, I., Brayer, G.D., Withers, S.G., and McIntosh, L.P. (2000). Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. J. Mol. Biol. 299, 255–279.

Joslyn, M.A. (1963). Advances in Food Research Volume 11 (Academic Press).

van der Kamp, J.W., Jones, J., MacCleary, B., and Topping, B. (2010). Dietary Fibre: New Frontiers for Food and Health (Wageningen Academic Publishers).

Katapodis, P., Vršanská, M., Kekos, D., Nerinckx, W., Biely, P., Claeyssens, M., Macris, B.J., and Christakopoulos, P. (2003). Biochemical and catalytic properties of an endoxylanase purified from the culture filtrate of Sporotrichum thermophile. Carbohydr. Res. 338, 1881–1890.

Katō, K. (1981). Ultrastructure of the Plant Cell Wall: Biochemical Viewpoint. In Plant Carbohydrates II, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 29–46.

Ke, J., Laskar, D.D., Gao, D., and Chen, S. (2012). Advanced biorefinery in lower termite-effect of combined pretreatment during the chewing process. Biotechnol. Biofuels 5, 11.

Khosla, C., and Harbury, P.B. (2001). Modular enzymes. Nature 409, 247–252.

Kishimoto, T., Chiba, W., Saito, K., Fukushima, K., Uraki, Y., and Ubukata, M. (2010). Influence of syringyl to guaiacyl ratio on the structure of natural and synthetic lignins. J. Agric. Food Chem. 58, 895–901.

Kittur, F.S., Mangala, S.L., Rus'd, A.A., Kitaoka, M., Tsujibo, H., and Hayashi, K. (2003). Fusion of family 2b carbohydrate-binding module increases the catalytic activity of a xylanase from Thermotoga maritima to soluble xylan. FEBS Lett. 549, 147–151.

Klein, C., and Schulz, G.E. (1991). Structure of cyclodextrin glycosyltransferase refined at 2.0 Å resolution. J. Mol. Biol. 217, 737–750.

Knauf, B.M., and Moniruzzaman, M. (2004). Lignocellulosic biomass processing : A perspective. Int. Sugar J. 106, 147–150.

Köhler, T., Dietrich, C., Scheffrahn, R.H., and Brune, A. (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (Nasutitermes spp.). Appl. Environ. Microbiol. 78, 4691–4701.

Kolenová, K., Vršanská, M., and Biely, P. (2006). Mode of action of endo-β-1,4-xylanases of families 10 and 11 on acidic xylooligosaccharides. J. Biotechnol. 121, 338–345.

Kormelink, F.J.M., and Voragen, A.G.J. (1993). Degradation of different [(glucurono)arabino]xylans by a combination of purified xylan-degrading enzymes. Appl. Microbiol. Biotechnol. 38, 688–695.

Koropatkin, N.M., and Smith, T.J. (2010). SusG: A Unique Cell-Membrane-Associated alpha-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. Structure 18, 200–215.

Koropatkin, N.M., Cameron, E. a., and Martens, E.C. (2012). How glycan metabolism shapes the human gut microbiota. Nat. Rev. Microbiol. 10, 323–335.

Koshland, D. (1953). Stereochemistry and the mechanism of enzymatic reactions. Biol. Rev. 28, 416–436.

Kristensen, J.B. (2009). Enzymatic hydrolysis of lignocellulosese.

Lam, K.N., Cheng, J., Engel, K., Neufeld, J.D., and Charles, T.C. (2015). Current and future resources for functional metagenomics. Front. Microbiol. 6, 1–8.

Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. (2000). Structural Evidence for Evolution of the beta /alpha Barrel Scaffold by Gene Duplication and Fusion. Science (80-.). 289, 1546–1550.

Larsen, T.M., Laughlin, L.T., Holden, H.M., Rayment, I., and Reed, G.H. (1994). Structure of rabbit muscle pyruvate kinase complexed with Mn2+, K+, and pyruvate. Biochemistry 33, 6301–6309.

Lee, S.F., Forsberg, C.W., and Rattray, J.B. (1987). Purification and characterization of two endoxylanases from Clostridium acetobutylicum ATCC 824. Appl. Environ. Microbiol. 53, 644–650.

Lehtiö, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., and Teeri, T.T. (2003). The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. Proc. Natl. Acad. Sci. U. S. A. 100, 484–489.

Lei, Q., Wen-Chao, L., Jia-Qing, Z., Bing-Zhi, L., and Ying-Jin, Y. (2017). Hydrolysis of lignocellulosic biomass to sugars. In Production of Platform Chemicals from Sustainable Resources, pp. 3–41.

Liu, N., Yan, X., Zhang, M., Xie, L., Wang, Q., Huang, Y., Zhou, X., Wang, S., and Zhou, Z. (2011). Microbiome of fungus-growing termites: a new reservoir for lignocellulase genes. Appl. Environ. Microbiol. 77, 48–56.

Liu, N., Zhang, L., Zhou, H., Zhang, M., Yan, X., Wang, Q., Long, Y., Xie, L., Wang, S., Huang, Y., et al. (2013). Metagenomic Insights into Metabolic Capacities of the Gut Microbiota in a Fungus-Cultivating Termite (Odontotermes yunnanensis). PLoS One 8, e69184.

Liu, Y., Yu, J., Li, F., Peng, H., Zhang, X., Xiao, Y., and He, C. (2017). Crystal structure of a raw-starch- degrading bacterial α -amylase belonging to subfamily 37 of the glycoside hydrolase family GH13. Sci. Rep. 7.

Lo, N., and Eggleton, P. (2011). Termite phylogenetics and co-cladogenesis with symbionts. In Biology of Termites: A Modern Synthesis, (Dordrecht: Springer Netherlands), pp. 27–50.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000). The Dynamic Plant Cell Wall.

Lopes, M.L., Paulillo, S.C. de L., Godoy, A., Cherubin, R.A., Lorenzi, M.S., Giometti, F.H.C., Bernardino, C.D., de Amorim Neto, H.B., and de Amorim, H.V. (2016). Ethanol

production in Brazil: a bridge between science and industry. Brazilian J. Microbiol. 47, 64–76.

Lynd, L.R., Weimer, P.J., Zyl, W.H. Van, and Isak, S. (2002). Microbial Cellulose Utilization : Fundamentals and Biotechnology. Microbiol. Mol. Biol. Rev. 66, 506–577.

Lynd, L.R., Van Zyl, W.H., McBride, J.E., and Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: An update. Curr. Opin. Biotechnol. 16, 577–583.

Lynd, L.R., Liang, X., Biddy, M.J., Allee, A., Cai, H., Foust, T., Himmel, M.E., Laser, M.S., Wang, M., and Wyman, C.E. (2017). Cellulosic ethanol: status and innovation. Curr. Opin. Biotechnol. 45, 202–211.

MacGregor, E.A., Jespersen, H.M., and Svensson, B. (1996). A circularly permuted α-amylase-type α/β-barrel structure in glucan-synthesizing glucosyltransferases. FEBS Lett. 378, 263–266.

Mackenzie, C.R., and Bilous, D. (1988). Ferulic Acid Esterase Activity from Schizophyllum commune. Appl. Environ. Microbiol. 54, 1170–1173.

MacLeod, A.M., Tull, D., Rupitz, K., Warren, R.A.J., and Withers, S.G. (1996). Mechanistic Consequences of Mutation of Active Site Carboxylates in a Retaining β-1,4-Glycanase from Cellulomonas fimi. Biochemistry 35, 13165–13172.

Madhavan, A., Sindhu, R., Parameswaran, B., Sukumaran, R.K., and Pandey, A. (2017). Metagenome Analysis: a Powerful Tool for Enzyme Bioprospecting. Appl. Biochem. Biotechnol. 183, 636–651.

Mahowald, M. a, Rey, F.E., Seedorf, H., Turnbaugh, P.J., Fulton, R.S., Wollam, A., Shah, N., Wang, C., Magrini, V., Wilson, R.K., et al. (2009). Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proc. Natl. Acad. Sci. U. S. A. 106, 5859–5864.

Martens, E.C., Chiang, H.C., and Gordon, J.I. (2008). Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. Cell Host Microbe 4, 447–457.

Martens, E.C., Koropatkin, N.M., Smith, T.J., and Gordon, J.I. (2009). Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. J. Biol. Chem. 284, 24673–24677.

Mathews, S.L., Pawlak, J., and Grunden, A.M. (2015). Bacterial biodegradation and bioconversion of industrial lignocellulosic streams. Appl. Microbiol. Biotechnol. 99, 2939–2954.

Mattevi, A., Valentini, G., Rizzi, M., Speranza, M.L., Bolognesi, M., and Coda, A. (1995). Crystal structure of Escherichia coli pyruvate kinase type I: molecular basis of the allosteric transition. Structure 3, 729–741.

Maureen, C.M., Brian, W., and Keith, R. (1992). Complexity in the spatial localization and length distribution of plant cell-wall matrix polysaccharides. J. Microsc. 166, 123–136.

Mayorga, O.L., Kingston-Smith, A.H., Kim, E.J., Allison, G.G., Wilkinson, T.J., Hegarty, M.J., Theodorou, M.K., Newbold, C.J., and Huws, S.A. (2016). Temporal metagenomic and metabolomic characterization of fresh perennial ryegrass degradation by rumen bacteria. Front. Microbiol. 7, 1–23.

Mccann, M.C., Wells, B., and Roberts, K. (1990). Direct visualization of cross-links in the primary plant cell wall. J. Cell Sci. 96, 323–334.

McCann, M.C., and Carpita, N.C. (2015). Biomass recalcitrance: A multi-scale, multi-factor, and conversion-specific property. J. Exp. Bot. 66, 4109–4118.

Mccarter, J.D., and Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. Curr. Opin. Struct. Biol. 4, 885–892.

McIntosh, L.P., Hand, G., Johnson, P.E., Joshi, M.D., Korner, M., Plesniak, L. a., Ziser, L., Wakarchuk, W.W., and Withers, S.G. (1996). The pKa of the General Acid / Base Carboxyl Group of a Glycosidase Cycles during Catalysis : A 13 C-NMR Study of Bacillus circulans Xylanase. Biochemistry 35, 9958–9966.

McKendry, P. (2002). Energy production from biomass (part 1): Overview of biomass. Bioresour. Technol. 83, 37–46.

Menetrez, M.Y. (2012). An overview of algae biofuel production and potential environmental impact. Env. Sci Technol 46, 7073–7085.

Merino, S.T., and Cherry, J. (2007). Progress and challenges in enzyme development for biomass utilization. Adv. Biochem. Eng. Biotechnol. 108, 95–120.

Mewis, K., Lenfant, N., Lombard, V., and Henrissat, B. (2016). Dividing the Large Glycoside Hydrolase Family 43 into Subfamilies: a Motivation for Detailed Enzyme Characterization. Appl. Environ. Microbiol. 82, 1686–1692.

Millward-Sadler, S.J., Davidson, K., Hazlewood, G.P., Black, G.W., Gilbert, H.J., and Clarke, J.H. (1995). Novel cellulose-binding domains, NodB homologues and conserved modular architecture in xylanases from the aerobic soil bacteria Pseudomonas fluorescens subsp. cellulosa and Cellvibrio mixtus. Biochem. J. 312 (Pt 1, 39–48.

Miyanaga, A., Koseki, T., Miwa, Y., Mese, Y., Nakamura, S., Kuno, A., Hirabayashi, J., Matsuzawa, H., Wakagi, T., Shoun, H., et al. (2006). The family 42 carbohydrate-binding

module of family 54 α-L-arabinofuranosidase specifically binds the arabinofuranose side chain of hemicellulose. Biochem. J. 399, 503–511.

Miyazaki, K., Martin, J.., Marinsek-Logar, R., and Flint, H.. (1997). Degradation and Utilization of Xylans by the Rumen AnaerobePrevotella bryantii(formerlyP. ruminicolasubsp.brevis) B14. Anaerobe 3, 373–381.

Miyazaki, K., Miyamoto, H., Mercer, D.K., Hirase, T., Martin, J.C., Kojima, Y., and Flint, H.J. (2003). Involvement of the Multidomain Regulatory Protein XynR in Positive Control of Xylanase Gene Expression in the Ruminal Anaerobe Prevotella bryantii B14. J. Bacteriol. 185, 2219–2226.

Mizutani, K., Sakka, M., Kimura, T., and Sakka, K. (2014). Essential role of a family-32 carbohydrate-binding module in substrate recognition by Clostridium thermocellum mannanase CtMan5A. FEBS Lett. 588, 1726–1730.

Mohnen, D. (2008). Pectin structure and biosynthesis. Curr Opin Plant Biol 11, 266–277.

Montanier, C., van Bueren, A.L., Dumon, C., Flint, J.E., Correia, M. a, Prates, J. a, Firbank, S.J., Lewis, R.J., Grondin, G.G., Ghinet, M.G., et al. (2009). Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. Proc. Natl. Acad. Sci. U. S. A. 106, 3065–3070.

Nagano, N., Orengo, C.A., and Thornton, J.M. (2002). One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J. Mol. Biol. 321, 741–765.

Najmudin, S., Guerreiro, C.I.P.D., Carvalho, A.L., Prates, J.A.M., Correia, M.A.S., Alves, V.D., Ferreira, L.M.A., Romão, M.J., Gilbert, H.J., Bolam, D.N., et al. (2006). Xyloglucan is recognized by carbohydrate-binding modules that interact with β-glucan chains. J. Biol. Chem. 281, 8815–8828.

Nakashima, K., Watanabe, H., Saitoh, H., Tokuda, G., and Azuma, J.I. (2002). Dual cellulose-digesting system of the wood-feeding termite, Coptotermes formosanus Shiraki. Insect Biochem. Mol. Biol. 32, 777–784.

Newstead, S.L., Watson, J.N., Bennet, A.J., and Taylor, G. (2005). Galactose recognition by the carbohydrate-binding module of a bacterial sialidase. Acta Crystallogr. Sect. D Biol. Crystallogr. 61, 1483–1491.

Ni, J., and Tokuda, G. (2013). Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. Biotechnol. Adv. 31, 838–850.

Ni, J., Takehara, M., and Watanabe, H. (2010). Identification of activity related amino acid mutations of a GH9 termite cellulase. Bioresour. Technol. 101, 6438–6443.

Noda, S., Hongoh, Y., Sato, T., and Ohkuma, M. (2009). Complex coevolutionary history of symbiotic Bacteroidales bacteria of various protists in the gut of termites. BMC Evol. Biol. 9, 158.

Notenboom, V., Birsan, C., Nitz, M., Rose, D.R., Warren, R. a, and Withers, S.G. (1998). Insights into transition state stabilization of the beta-1,4-glycosidase Cex by covalent intermediate accumulation in active site mutants. Nat. Struct. Biol. 5, 812–818.

NR.Gilkes, B.Henrissat, DG.kilburn, R.C.Miller, JR., and R.AJ.Warren (1991). Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. 55, 303–315.

Nurizzo, D., Nagy, T., Gilbert, H.J., and Davies, G.J. (2002). The structural basis for catalysis and specificity of the Pseudomonas cellulosa α-glucuronidase, GlcA67A. Structure 10, 547–556.

O'sullivan, A.C. (1997). Cellulose: the structure slowly unravels. Cellulose 4, 173–207.

Öhgren, K., Bura, R., Saddler, J., and Zacchi, G. (2007). Effect of hemicellulose and lignin removal on enzymatic hydrolysis of steam pretreated corn stover. Bioresour. Technol. 98, 2503–2510.

Ohkuma, M. (2003). Termite symbiotic systems: efficient bio-recycling of lignocellulose. Appl. Microbiol. Biotechnol. 61, 1–9.

Okkerse, C., and van Bekkum, H. (1999). From fossil to green. Green Chem. 1, 107–114.

Oppert, C., Klingeman, W.E., Willis, J.D., Oppert, B., and Jurat-Fuentes, J.L. (2010). Prospecting for cellulolytic activity in insect digestive fluids. Comp. Biochem. Physiol. - B Biochem. Mol. Biol. 155, 145–154.

Paës, G., Berrin, J.-G., and Beaugrand, J. (2012). GH11 xylanases: Structure/function/properties relationships and applications. Biotechnol. Adv. 30, 564–592.

Pantoom, S., Songsiriritthigul, C., and Suginta, W. (2008). The effects of the surface-exposed residues on the binding and hydrolytic activities of Vibrio carchariae chitinase A. BMC Biochem. 9, 2.

Pell, G., Taylor, E.J., M.Gloster, T., Turkenburg, J.P., M.G.A.Fontes, C., M.A.Ferreira, L., Nagy, T., J.Clark, S., J.Davies, G., J.Gilbert, H., et al. (2004a). The Mechanisms by Which Family 10 Glycoside Hydrolases Bind Decorated Substrates. J. Biol. Chem. 279, 9597–9605.

Pell, G., Szabo, L., Charnock, S.J., Xie, H., Gloster, T.M., Davies, G.J., and Gilbert, H.J. (2004b). Structural and biochemical analysis of Cellvibrio japonicus Xylanase 10C: How variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases. J. Biol. Chem. 279, 11777–11788.

Pereira, C.S., Silveira, R.L., Dupree, P., and Skaf, M.S. (2017). Effects of Xylan Side-Chain Substitutions on Xylan-Cellulose Interactions and Implications for Thermal Pretreatment of Cellulosic Biomass. Biomacromolecules 18, 1311–1321.

Phillips, D.C. (1967). The Hen Egg-White Lysozyme Molecule. Proc. Natl. Acad. Sci. U. S. A. 57, 483–495.

Polizeli, M.L.T.M., Rizzatti, a. C.S., Monti, R., Terenzi, H.F., Jorge, J. a., and Amorim, D.S. (2005). Xylanases from fungi: properties and industrial applications. Appl. Microbiol. Biotechnol. 67, 577–591.

Pollet, A., Delcour, J. a, and Courtin, C.M. (2010). Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. Crit. Rev. Biotechnol. 30, 176–191.

Prade, R.A. (1996). Xylanases : from Biology to BioTechnology. Biotechnol. Genet. Eng. Rev. 13, 101–131.

Preece, I.A. (1931). Studies on hemicelluloses. J. Inst. Brew. 37, 409–413.

Puranen, T., Alapuranen, M., and Vehmaanperä, J. (2014). Trichoderma Enzymes for Textile Industries. In Biotechnology and Biology of Trichoderma, pp. 351–362.

Qian, M., Haser, R., and Payan, F. (1993). Structure and Molecular Model Refinement of Pig Pancreatic α-Amylase at 2•1 Å Resolution. J Mol Biol 231, 785–799.

Rakshit, S.K. (2014). Second generation bio-ethanol and renewable chemicals from lignocellulosics. In Biofuel Technologies: Recent Developments, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 259–269.

Reale, S., Di Tullio, A., Spreti, N., and De Angelis, F. (2004). Mass spectrometry in the biosynthetic and structural investigation of lignins. Mass Spectrom. Rev. 23, 87–126.

Reardon, D., and Farber, G.K. (1995). The structure of alpha / beta barrel proteins. FASEB J. 9, 497–503.

Reeves, A.R., Wang, G., and Salyers, A.A. (1997). Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. J. Bacteriol. 179, 643–649.

Reilly, P.J. (1981). Xylanase: structure and function. In Trends in the Biology of Fermentations for Fuels and Chemicals, pp. 111–129.

Rey, F.E., Faith, J.J., Bain, J., Muehlbauer, M.J., Stevens, R.D., Newgard, C.B., and Gordon, J.I. (2010). Dissecting the in vivo metabolic potential of two human gut acetogens. J. Biol. Chem. 285, 22082–22090.

Ricard, G., McEwan, N.R., Dutilh, B.E., Jouany, J.P., Macheboeuf, D., Mitsumori, M., McIntosh, F.M., Michalowski, T., Nagamine, T., Nelson, N., et al. (2006). Horizontal gene transfer from bacteria to rumen ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. BMC Genomics 7, 1–13.

Ridley, B.L., O'Neill, M.A., and Mohnen, D. (2001). Pectins: Structure, biosynthesis, and oligogalacturonide-related signaling.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42, 320–324.

Rodríguez-Sanoja, R., Oviedo, N., Escalante, L., Ruiz, B., and Sánchez, S. (2009). A single residue mutation abolishes attachment of the CBM26 starch-binding domain from Lactobacillus amylovorus α-amylase. J. Ind. Microbiol. Biotechnol. 36, 341–346.

Rogowski, A., Baslé, A., Farinas, C.S., Solovyova, A., Mortimer, J.C., Dupree, P., Gilbert, H.J., and Bolam, D.N. (2014). Evidence that GH115 alpha-glucuronidase activity, which is required to degrade plant biomass, is dependent on conformational flexibility. J. Biol. Chem. 289, 53–64.

Rossmassler, K., Dietrich, C., Thompson, C., Mikaelyan, A., Nonoh, J.O., Scheffrahn, R.H., Sillam-Dussès, D., and Brune, A. (2015). Metagenomic analysis of the microbiota in the highly compartmented hindguts of six wood- or soil-feeding higher termites. Microbiome 3, 56.

Ryabova, O., Vršanská, M., Kaneko, S., van Zyl, W.H., and Biely, P. (2009). A novel family of hemicellulolytic alpha-glucuronidase. FEBS Lett. 583, 1457–1462.

Saab-Rincón, G., Olvera, L., Olvera, M., Rudiño-Piñera, E., Benites, E., Soberón, X., and Morett, E. (2012). Evolutionary walk between (β/α) 8 barrels: Catalytic migration from triosephosphate isomerase to thiamin phosphate synthase. J. Mol. Biol. 416, 255 ‑ 270.

Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jørgensen, P.L., Schülein, M., Dupont, C., Davies, G.J., and Wilson, K.S. (1999). Catalysis and specificity in enzymatic glycoside hydrolysis: a 2,5 B conformation for the glycosyl-enzyme intermediate revealed by the structure of the Bacillus agaradhaerens family 11 xylanase. Chem. Biol. 6, 483–492.

Saha, B.C. (2000). Alpha-L-arabinofuranosidases: biochemistry, molecular biology and application in biotechnology. Biotechnol. Adv. 18, 403–423.

Saha, B.C. (2001). Purification and characterization of an extracellular beta-xylosidase from a newly isolated Fusarium verticillioides. J. Ind. Microbiol. Biotechnol. 27, 241–245.

Saha, B.C. (2003). Hemicellulose bioconversion. J. Ind. Microbiol. Biotechnol. 30, 279–291.

Saini, J.K., Saini, R., and Tewari, L. (2015). Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. 3 Biotech 5, 337–353.

Sainz-Polo, M.A., González, B., Menéndez, M., Pastor, F.I.J., and Sanz-Aparicio, J. (2015). Exploring multimodularity in plant cell wall deconstruction: Structural and functional analysis of Xyn10C containing the CBM22-1-CBM22-2 tandem. J. Biol. Chem. 290, 17116–17130.

Sakka, M., Higashi, Y., Kimura, T., Ratanakhanokchai, K., and Sakka, K. (2011). Characterization of Paenibacillus curdlanolyticus B-6 Xyn10D, a xylanase that contains a family 3 carbohydrate-binding module. Appl. Environ. Microbiol. 77, 4260–4263.

Salis, A., Monduzzi, M., and Solinas, V. (2007). Industrial Enzymes.

Sanyika, T.W., Rashamuse, K.J., Hennessy, F., and Brady, D. (2012). Luminal hindgut bacterial diversities of the grass and sugarcane feeding termite Trinervitermes trinervoides. African J. Microbiol. Res. 6, 2639–2648.

Sapag, A., Wouters, J., Lambert, C., De Ioannes, P., Eyzaguirre, J., and Depiereux, E. (2002). The endoxylanases from family 11: Computer analysis of protein sequences reveals important structural and phylogenetic relationships. J. Biotechnol. 95, 109–131.

Sarkar, N., Ghosh, S.K., Bannerjee, S., and Aikat, K. (2012). Bioethanol production from agricultural wastes: An overview. Renew. Energy 37, 19–27.

Sarkar, P., Bosneaga, E., and Auer, M. (2009). Plant cell walls throughout evolution: Towards a molecular understanding of their design principles. J. Exp. Bot. 60, 3615–3635.

Schädel, C., Richter, A., Blöchl, A., and Hoch, G. (2010). Hemicellulose concentration and composition in plant cell walls under extreme carbon source-sink imbalances. Physiol. Plant. 139, 241–255.

Scharf, M.E. (2015a). Omic research in termites: an overview and a roadmap. Front. Genet. 6, 1–16.

Scharf, M.E. (2015b). Termites as targets and models for biotechnology. Annu. Rev. Entomol. 60, 77–102.

Scharf, M.E., Karl, Z.J., Sethi, A., and Boucias, D.G. (2011). Multiple levels of synergistic collaboration in termite lignocellulose digestion. PLoS One 6, 1–7.

Schauer, C., Thompson, C.L., and Brune, A. (2012). The bacterial community in the gut of the cockroach Shelfordella lateralis reflects the close evolutionary relatedness of cockroaches and termites. Appl. Environ. Microbiol. 78, 2758–2767.

Scheller, H.V., and Ulvskov, P. (2010). Hemicelluloses. Annu. Rev. Plant Biol. 61, 263–289.

Schmidt, A., GÜbitz, G.M., and Kratky, C. (1999). Xylan binding subsite mapping in the xylanase from Penicillium simplicissimum using xylooligosaccharides as cryo-protectant. Biochemistry 38, 2403–2412.

Selvendran, R.R. (1984). The plant cell wall as a source of dietary fiber: chemistry and structure. Am. J. Clin. Nutr. 39, 320–337.

Sermsathanaswadi, J., Baramee, S., Tachaapaikoon, C., Pason, P., Ratanakhanokchai, K., and Kosugi, A. (2017). The family 22 carbohydrate-binding module of bifunctional xylanase/β-glucanase Xyn10E from Paenibacillus curdlanolyticus B-6 has an important role in lignocellulose degradation. Enzyme Microb. Technol. 96, 75–84.

Shallom, D., and Shoham, Y. (2003). Microbial hemicellulases. Curr Opin Microbiol 6, 219–228.

Shao, W., Obi, S., Puls, J., Wiegel, J., and Puls, R. (1995). Purification and Characterization of the alpha -Glucuronidase from Thermoanaerobacterium sp . Strain JW / SL-YS485 , an Important Enzyme for the Utilization of Substituted Xylans. Appl. Environ. Microbiol. 61, 1077–1081.

Shipman, J.A., Berleman, J.E., and Salyers, A.A. (2000). Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. J. Bacteriol. 182, 5365–5372.

Simpson, P.J., Xie, H., Bolam, D.N., Gilbert, H.J., and Williamson, M.P. (2000). The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. J. Biol. Chem. 275, 41137–41142.

Slavin, J.L., Brauer, P.M., and Marlett, J. a (1981). Neutral detergent fiber, hemicellulose and cellulose digestibility in human subjects. J. Nutr. 111, 287–297.

Smith, D.C., and Forsberg, C.W. (1991). Alpha-Glucuronidase and other hemicellulase activities of Fibrobacter succinogenes S85 grown on crystalline cellulose or ball-milled barley straw. Appl. Environ. Microbiol. 57, 3552–3557.

Smith, M.M., and Hartley, R.D. (1983). Occurrence and nature of ferulic acid substitution of cell-wall polysaccharides in graminaceous plants. Carbohydr. Res. 118, 65–80.

Solomon, V., Teplitsky, A., Shulami, S., Zolotnitsky, G., Shoham, Y., and Shoham, G. (2007). Structure-specificity relationships of an intracellular xylanase from Geobacillus stearothermophilus. Acta Crystallogr. Sect. D Biol. Crystallogr. 63, 845–859.

Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredez, A., Persson, S., Raab, T., et al. (2004). Toward a systems approach to understanding plant cell walls. Science (80-.). 306, 2206–2211.

Sorimachi, K., Gal-Coëffet, M.-F. Le, Williamson, G., Archer, D.B., and Williamson, M.P. (1997). Solution structure of the granular starch binding domain of Aspergillus niger glucoamylase bound to β-cyclodextrin. Structure 5, 647–661.

Southall, S.M., Simpson, P.J., Gilbert, H.J., Williamson, G., and Williamson, M.P. (1999). The starch-binding domain from glucoamylase disrupts the structure of starch. FEBS Lett. 447, 58–60.

Stam, M.R., Danchin, E.G.J., Rancurel, C., Coutinho, P.M., and Henrissat, B. (2006). Dividing the large glycoside hydrolase family 13 into subfamilies: Towards improved functional annotations of α-amylase-related proteins. Protein Eng. Des. Sel. 19, 555–562.

Stephen, A.M. (1983). Other Plant Polysaccharides. In The Polysaccharides, VOL.2, pp. 97–193.

Sticklen, M.B. (2008). Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. Nat. Rev. Genet. 9, 433–443.

Straus, D., and Gilbert, W. (1985). Genetic engineering in the Precambrian: structure of the chicken triosephosphate isomerase gene. Mol. Cell. Biol. 5, 3497–3506.

Suh, J.-H., and Choi, Y.-J. (1996a). Synergism among endo-xylanase, beta-xylosidase, and acetyl xylan esterase from Bacillus stearothermophilus. J. Microbiol. Biotechnol. 6, 173–178.

Suh, J.-H., Cho, S.-G., and Choi, Y.-J. (1996b). Synergic effects among endo-xylanase, β-xylosidase, and α-L-arabinofuranosidase from Bacillus stearothermophilus. J. Microbiol. Biotechnol. 6, 179–183.

Sun, J., Ding, S.-Y., and Doran-peterson, J.O.Y. (2014). Biomass and its Biorefinery: Novel Approaches from Nature-Inspired Strategies and Technology. In Biological Conversion of Biomass for Fuels and Chemicals: Explorations from Natural Utilization Systems, pp. 1–13.

Sun, J.Y., Liu, M.Q., Weng, X.Y., Qian, L.C., and Gu, S.H. (2007). Expression of recombinant Thermomonospora fusca xylanase A in Pichia pastoris and xylooligosaccharides released from xylans by it. Food Chem. 104, 1055–1064.

Sunna, A., and Antranikian, G. (1997). Xylanolytic enzymes from fungi and bacteria. Crit. Rev. Biotechnol. 17, 39–67.

Suresh, C., KITAOKA, M., and HAYASHI, K. (2003). A Thermostable Non-xylanolytic α-Glucuronidase of Thermotoga maritima MSB8. Biosci. Biotechnol. Biochem. 67, 2359–2364.

Suzuki, R., Fujimoto, Z., Ito, S., Kawahara, S.I., Kaneko, S., Taira, K., Hasegawa, T., and Kuno, A. (2009). Crystallographic snapshots of an entire reaction cycle for a retaining xylanase from Streptomyces olivaceoviridis E-86. J. Biochem. 146, 61–70.

Sweeney, M.D., and Xu, F. (2012). Biomass Converting Enzymes as Industrial Biocatalysts for Fuels and Chemicals: Recent Developments. Catalysts 2, 244–263.

Takenishi, S., Tsujisaka, Y., and Fukumoto, J. (1973). Studies on hemicellulases. IV. Purification and properties of the Beta-xylosidase produced by Aspergillus niger van Tieghem. J. Biochem. 73, 335–343.

Tanakas, Y., Taoo, W., Blanchards, J.S., and Hehren, E.J. (1994). Transition State Structures for the Hydrolysis of a-D-Glucopyranosyl Fluoride by Retaining and Inverting Reactions of Glycosylases. J. Biol. Chem. 269, 32306–32312.

Tanner, W., and Loewus, F.A. (1981). Plant Carbohydrates II : Extracellular Carbohydrates (Springer Berlin Heidelberg).

Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Leclerc, M., et al. (2010). Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome Res. 20, 1605–1612.

Teeri, T.T. (1997). Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. Trends Biotechnol. 15, 160–167.

Tenkanen, M., and Siika-aho, M. (2000). An alpha-glucuronidase of Schizophyllum commune acting on polymeric xylan. J. Biotechnol. 78, 149–161.

Terrapon, N., Lombard, V., Gilbert, H.J., and Henrissat, B. (2015). Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. Bioinformatics 31, 647–655.

Todaka, N., Inoue, T., Saita, K., Ohkuma, M., Nalepa, C.A., Lenz, M., Kudo, T., and Moriya, S. (2010). Phylogenetic analysis of cellulolytic enzyme genes from representative lineages of termites and a related cockroach. PLoS One 5, 1–10.

Tomme, P., Warren, R. a, and Gilkes, N.R. (1995). Cellulose hydrolysis by bacteria and fungi.

Tomme, P., Kwan, E., Gilkes, N.R., Kilburn, D.G., Antony, R., and Warren, J. (1996). Characterization of CenC, an enzyme from Cellulomonas fimi with both endo- and exoglucanase activities. J. Bacteriol. 178, 4216–4223.

Topping, D. (2007). Cereal complex carbohydrates and their contribution to human health. J. Cereal Sci. 46, 220–229.

Tormo, J., Lamed, R., Chirino, A.J., Morag, E., Bayer, E.A., Shoham, Y., and Steitz, T.A. (1996). Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. EMBO J. 15, 5739–5751.

Tull, D., and Withers, S.G. (1994). Mechanisms of cellulases and xylanases: a detailed kinetic study of the exo-beta-1,4-glycanase from Cellulomonas fimi. Biochemistry 33, 6363–6370.

Tuohy, M.G., Puls, J., Claeyssens, M., Vrsanská, M., and Coughlan, M.P. (1993). The xylan-degrading enzyme system of Talaromyces emersonii: novel enzymes with activity against aryl beta-D-xylosides and unsubstituted xylans. Biochem. J. 290, 515–523.

Uchiyama, T., and Miyazaki, K. (2009). Functional metagenomics for enzyme discovery: challenges to efficient screening. Curr. Opin. Biotechnol. 20, 616–622.

Urfer, R., and Kirschner, K. (1992). The importance of surface loops for stabilizing an eightfold beta alpha barrel protein. Protein Sci. 1, 31–45.

Valentin Dibangou (2012). Spatial distribution and density of fungus-growing termite Pseudacanthotermes militaris (Isoptera: Macrotermitinae) in the congo republic. Asian J. Biol. Sci. 5, 406–416.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010). Lignin biosynthesis and structure. Plant Physiol. 153, 895–905.

Várnai, A., Siika-Aho, M., and Viikari, L. (2013). Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs. Biotechnol. Biofuels 6, 30.

Varner, J.E., Louis, S., Lin, L.-S., Louis, S., and Lin, L.-S. (1989). Plant cell wall architecture. Cell 56, 231–239.

Veivers, P.C., Musca, A.M., Brien, R.W.O., and Slaytor, M. (1982). Digestive Enzymes of the Salivary glands and gut of Mastotermes darwiniensis. Insect Biochem. 12.

Viegas, A., Brás, N.F., Cerqueira, N.M.F.S.A., Fernandes, P.A., Prates, J.A.M., Fontes, C.M.G.A., Bruix, M., Romão, M.J., Carvalho, A.L., Ramos, M.J., et al. (2008). Molecular determinants of ligand specificity in family 11 carbohydrate binding modules - An NMR, X-ray crystallography and computational chemistry approach. FEBS J. 275, 2524–2535.

de Vries, R.P., Kester, H.C.M., Poulsen, C.H., Benen, J. a E., and Visser, J. (2000). Synergy between accessory enzymes from Aspergillus in the degradation of plant cell wall polysaccharides. Carbohydr. Res. 327, 401–410.

Vuong, T. V., and Wilson, D.B. (2010). Glycoside hydrolases: Catalytic base/nucleophile diversity. Biotechnol. Bioeng. 107, 195–205.

Wan, Q., Parks, J.M., Hanson, B.L., Fisher, S.Z., Ostermann, A., Schrader, T.E., Graham, D.E., Coates, L., Langan, P., and Kovalevsky, A. (2015). Direct determination of protonation

states and visualization of hydrogen bonding in a glycoside hydrolase with neutron crystallography. Proc. Natl. Acad. Sci. 112, 12384–12389.

Wang, J., Zhang, T., Liu, R., Song, M., Wang, J., Hong, J., Chen, Q., and Liu, H. (2017). Recurring sequence-structure motifs in (βα)8-barrel proteins and experimental optimization of a chimeric protein designed based on such motifs. Biochim. Biophys. Acta - Proteins Proteomics 1865, 165–175.

Wang, K., Pereira, G. V., Cavalcante, J.J. V., Zhang, M., Mackie, R., and Cann, I. (2016). Bacteroides intestinalis DSM 17393, a member of the human colonic microbiome, upregulates multiple endoxylanases during growth on xylan. Sci. Rep. 6, 1–11.

Wang, L., Zhang, Y., and Gao, P. (2008). A novel function for the cellulose binding module of cellobiohydrolase I. Sci. China, Ser. C Life Sci. 51, 620–629.

Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N., et al. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450, 560–565.

Warren, R.A.J. (1996). Microbial Hydrolysis of polysaccharides. Annu. Rev. Microbiol 50, 183–212.

Watanabe, H., and Tokuda, G. (2001). Animal cellulases. Cell. Mol. Life Sci. 58, 1167–1178.

Watanabe, H., and Tokuda, G. (2010). Cellulolytic systems in insects. Annu. Rev. Entomol. 55, 609–632.

Watanabe, H., Noda, H., Tokuda, G., and Lo, N. (1998). A cellulase gene of termite origin. Nature 394, 330–331.

Werpy, T.A., Holladay, J.E., and White, J.F. (2004). Top Value Added Chemicals from Biomass.

White, A., Tull, D., Johns, K., Withers, S.G., and Rose, D.R. (1996). Crystallographic observation of a covalent catalytic intermediate in a beta-glycosidase. Nat. Struct. Biol. 3, 149–154.

White, B. a, Lamed, R., Bayer, E. a, and Flint, H.J. (2014). Biomass Utilization by Gut Microbiomes. Annu. Rev. Microbiol. 68, 279–296.

Wierenga, R.K. (2001). The TIM-barrel fold: A versatile framework for efficient enzymes. FEBS Lett. 492, 193–198.

Wilkens, C., Andersen, S., Dumon, C., Berrin, J.-G., and Svensson, B. (2017). GH62 arabinofuranosidases: Structure, function and applications. Biotechnol. Adv. 35, 792–804.

Williams, B.A., Grant, L.J., Gidley, M.J., and Mikkelsen, D. (2017). Gut Fermentation of Dietary Fibres : Physico-Chemistry of Plant Cell Walls and Implications for Health. Int. J. Mol. Sci. 18.

Withers, S.G., Dombroski, D., Berven, L.A., Kilburn, D.G., Miller, R.C., J. Warren, R.A., and Gilkes, N.R. (1986). Direct 1H N.M.R. determination of the stereochemical course of hydrolyses catalysed by glucanase components of the cellulase complex. Biochem. Biophys. Res. Commun. 139, 487–494.

Wong, K.K., Tan, L.U., and Saddler, J.N. (1988). Multiplicity of beta-1,4-xylanase in microorganisms: functions and applications. Microbiol. Rev. 52, 305–317.

Wyman, C.E. (1999). BIOMASS ETHANOL: Technical Progress, Opportunities, and Commercial Challenges. Annu. Rev. Energy Environ. 24, 189–226.

Wyman, C.E., Cai, C.M., and Kumar, R. (2017). Bioethanol from Lignocellulosic Biomass. In Encyclopedia of Sustainability Science and Technology, pp. 1–27.

Xu, J., Bjursell, M.K., Himrod, J., Dend, S., Carmichael, L.K., Chiang, H.C., Hooper, L. V., and Gordon, J.I. (2003). A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis. Science (80-.). 299, 2074–2076.

Xu, J., Mahowald, M.A., Ley, R.E., Lozupone, C.A., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., et al. (2007). Evolution of Symbiotic Bacteria in the Distal Human Intestine. PLoS Biol. 5, 1574–1586.

Yan, Q.J., Wang, L., Jiang, Z.Q., Yang, S.Q., Zhu, H.F., and Li, L.T. (2008). A xylose-tolerant β-xylosidase from Paecilomyces thermophila: Characterization and its co-action with the endogenous xylanase. Bioresour. Technol. 99, 5402–5410.

Yoon, K.A., Kim, J.H., Hwang, C.E., Kim, Y.H., Lee, W.H., and Lee, S.H. (2015). Cellulase gene expression profiles in termites according to habitat and diet. J. Asia. Pac. Entomol. 18, 369–375.

You, T., and Feng, X. (2016). Applications of of Molecular Methods to Applications to the Elucidation of Lignin Structure the Elucidation of Lignin Structure. In Applications of Molecular Spectroscopy to Current Research in the Chemical and Biological Sciences, pp. 235–260.

Yu, P., Maenz, D.D., McKinnon, J.J., Racz, V.J., and Christensen, D. a. (2002). Release of ferulic acid from oat hulls by Aspergillus ferulic acid esterase and Trichoderma xylanase. J. Agric. Food Chem. 50, 1625–1630.

Zanoelo, F.F., Polizeli, M. de L.T. de M., Terenzi, H.F., and Jorge, J.A. (2004). Purification and biochemical properties of a thermostable xylose-tolerant β-D-xylosidase from Scytalidium thermophilum. J. Ind. Microbiol. Biotechnol. 31, 170–176.

Zeikus, J.G., Jain, M.K., and Elankovan, P. (1999). Biotechnology of succinic acid production and markets for derived industrial products. Appl. Microbiol. Biotechnol. 51, 545–552.

Zhang, J., and Viikari, L. (2012). Xylo-oligosaccharides are competitive inhibitors of cellobiohydrolase I from Thermoascus aurantiacus. Bioresour. Technol. 117, 286–291.

Zhang, C., Zhuang, X., Wang, Z.J., Matt, F., John, F.S., and Zhu, J.Y. (2013). Xylanase supplementation on enzymatic saccharification of dilute acid pretreated poplars at different severities. Cellulose 20, 1937–1946.

Zhang, M., Chekan, J.R., Dodd, D., Hong, P.-Y.P.-Y.P.-Y.P.-Y., Radlinski, L., Revindran, V., Nair, S.K., Mackie, R.I., and Cann, I. (2014). Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. Proc. Natl. Acad. Sci. U. S. A. 111, E3708-17.

Zhang, Z., Song, J., and Han, B. (2017). Catalytic Transformation of Lignocellulose into Chemicals and Fuel Products in Ionic Liquids. Chem. Rev. 117, 6834–6880.

Zhao, G., Ali, E., Araki, R., Sakka, M., Kimura, T., and Sakka, K. (2005). Function of the family-9 and family-22 carbohydrate-binding modules in a modular β-1,3-1,4-glucanase/xylanase derived from Clostridium stercorarium Xyn10B. Biosci. Biotechnol. Biochem. 69, 1562–1567.

Zhou, X., Smith, J. a. J., Oi, F.F.M., Koehler, P.P.G., Bennett, G.G.W., and Scharf, M.E. (2007). Correlation of cellulase gene expression and cellulolytic activity throughout the gut of the termite Reticulitermes flavipes. Gene 395, 29–39.

Zolotnitsky, G., Cogan, U., Adir, N., Solomon, V., Shoham, G., and Shoham, Y. (2004). Mapping glycoside hydrolase substrate subsites by isothermal titration calorimetry. Proc. Natl. Acad. Sci. U. S. A. 101, 11275–11280.

Zverlov, V. V., Volkov, I.Y., Velikodvorskaya, G.A., and Schwarz, W.H. (2001). The binding pattern of two carbohydrate-binding modules of laminarinase Lam16A from Thermotoga neapolitana: Differences in β-glucan binding within family CBM4. Microbiology 147, 621–629.

Van Zyl, W.H., Lynd, L.R., Den Haan, R., and McBride, J.E. (2007). Consolidated bioprocessing for bioethanol production using saccharomyces cerevisiae. Adv. Biochem. Eng. Biotechnol. 108, 205–235.

ARTICLE I

Investigating the multimodularity of a Xyn10C-like protein found in termite gut

ABSTRACT:

The functional screening of a the *Pseudacantotermes militaris* termite gut metagenomic library revealed an array of multi-modular enzymes, including a novel GH10 encoding sequence, *Pm25*. Sequence analysis showed details of the unusual domain organization of this enzyme. It consists of one catalytic domain, which is intercalated by two Carbohydrate Binding Modules (CBMs). Homologue of this enzyme sharing the same architecture, Xyn10C, was vastly distributed in different xylan utilization loci found in gut Bacteroidetes, which indicated its importance in the glycan foraging for the gut microbiota. In an effort to understand its unusual multi-modularity, detailed characterization of *Pm*25, including the catalytic domain GH10 separated from the CBM4, and CBM domain was performed. Here we show that the *Pm*25 is dedicated to the degradation of xylan. The binding results showed that the catalytic domain and CBMs act independently, and the tandem CBM4s act synergistically with each other and assist the release of D-xylose towards complex polysaccharide but not soluble polysaccharide. These findings contribute to a better understanding of the potential role of Xyn10C-like proteins in xylan utilization system of gut bacteria.

INTRODUCTION:

Xylan is the most abundant hemicellulose present in cell walls of higher plants, especially cereal grains and hardwoods (Stephen, 1983). Usually, xylan main chains are composed of β-1,4-linked D- xylopyranosyl residues that often bear substitutions at O-2 and/or O-3 positions. Moieties such as L-arabinofuranosyl (L-Ara*f*), 4-O-methyl glucuronyl (D-MeGlcA*p*) and acetyl residues are frequent main chain substituents, and substituting L-Ara*f* moieties can be esterified by ferulate at their O-5 position. The exact nature and pattern of xylan main chain decorations depends on the botanical source. Moreover, when xylan is extracted from a single plant species, the tissue source and the developmental status will determine the type of xylan that is observed (Scheller et al., 2010). Graminaceous plants are generally rich in glucuronoarabinoxylan (GAX), whereas the dominant xylan in typical dicotyledons is glucuronoxylan (GX), the difference between these two categories being the relative amounts of L-Ara*f* and D-MeGlcA*p* present. Complete xylan degradation requires an extensive arsenal of enzymes that often work in synergy. The main chain is depolymerized by β-D-xylanases that hydrolyze internal β-1, 4 bonds, while decorations are removed by a wide variety of accessory enzymes, including α-L-arabinofuranosidases, α-D-glucuronidases, feruloyl and acetyl esterases. Finally, β-D-xylosidases breakdown xylo-oligosaccharides, removing D-Xyl from the reducing extremity (Merino et al., 2007).

According to the Pfam database (http://pfam.xfam.org/), 20-30% of β-D-xylanases are multi-domain proteins, comprising catalytic domains associated with accessory, or helper domains such as carbohydrate binding modules (CBMs). These latter have been attributed various roles, including the ability to target specific regions in substrates (Bolam et al., 1998), disrupt polysaccharide structure (Din et al., 1994) or anchor enzymes to bacterial surfaces (Montanier et al., 2009). In the CAZy database (www.cazy.org, Lombard et al. 2014), carbohydrate-degrading enzymes and CBMs are grouped into families that are defined on the basis of primary amino acid similarities and, by inference, structural relatedness (Cantarel et al., 2009). β-D-xylanases are mainly found in glycoside hydrolase (GH) families 5, 8, 10 and 11, with family 5, 10 and 11 members performing catalysis via a retaining mechanism (Collins et al., 2005). The canonical 3D structure of GH5 and 10 families is a TIM barrel - $(\beta/\alpha)_8$, which is the most commonly known (2077 occurrences) protein fold in the Protein Data Bank (PDB). Other xylan-degrading enzymes are found in GH families, such as GH43, 51, 62 (α-L-arabinofuranosidase activities) and GH115 (α-glucuronidase activity). Regarding CBMs, these

are grouped into 81 CAZy families, 15 of which contain members that bind to xylan and display β-sheet architectures.

Bioinformatics reveal that 60 and 80 % of prokaryotic and eukaryotic proteins respectively are formed from multiple domains, which are defined as the structural, functional or evolutionary units of proteins (Aroul-Selvam et al., 2004). In multi-domain proteins, individual domains can be regarded as biological equivalents of components in complex devices, whose parts can be interchanged. For protein engineers, multi-domain architectures are particularly interesting, because they provide clues about how new biocatalysts can be made. Mostly, domains in proteins are sequentially organized, with one domain following another one. However, 9% of domain combinations are discontinuous, with one domain being inserted into another one.

Termites are wood-feeding animals that are considered to be an abundant source of biomass-degrading enzymes (Ohkuma, 2003). Although termites do not actually produce many endogenous biomass-degrading enzymes, their guts contain microorganisms that are responsible for their ability to capture nutrients and energy from biomass (Brune, 2007; Scharf and Tartar, 2012). Over the last decade, this fact has underpinned numerous metagenomics studies that have aimed to characterize the enzyme arsenals of termite gut microbiomes and detect promising enzymes for industrial use (Bastien et al., 2013; Otani et al., 2014; Su et al., 2016; Warnecke et al., 2007).

Detailed studies of gut ecosystems have revealed that gut symbionts deploy different strategies to capture nutrients. While members of the Firmicutes (Gram-positive) phylum deploy vast arsenals of extracellular enzymes (Gilbert, 2007), those of the Bacteroidetes (Gram-negative), the dominant phylum in many animal digestive systems (Bignell, 2010; Bryant et al., 1957; Chassard et al., 2007; Dehority, 1966; Kudo, 2009), possess glycan utilization systems. The paradigm for this type of system is provided by the well-studied Starch Utilization System (Sus). In Sus-like systems, several proteins encoded by a gene cluster (known as a polysaccharide utilization loci or PUL) act in a coordinated manner, binding and hydrolyzing complex sugars (Martens et al., 2009). Many PULs have been described, including a xylan utilization system (Xus) that is composed of two outer-membrane polysaccharide-binding proteins (XusB and XusD), two transporter proteins (XusA and XusC) and two outer-membrane proteins (XusE and Xyn10C) (Dodd et al., 2011). Each of these proteins is expressed from a cluster of tandem genes that are organized as follows: *xus A*/*B*/*C*/*D* (or sometimes only *xus C*/*D*), followed by *xus E* and *xyn10C*, the latter encoding a CBM-containing GH10 β-D-xylanase (Dodd, Mackie, and Cann 2011). According to

previous data, the CBMs in Xyn10C are inserted into the polypeptide sequence of the GH10 catalytic domain, between structural elements β3 and α3 (Flint et al., 1997). The expression of Xyn10C was shown to be induced by xylan (Despres et al., 2016) and so were the other xylanases XynA and XynB, the most effective inducer of which is demonstrated to be with a degree of polymerization (DP) similar to the hydrolysates of Xyn10C (around 35)(Miyazaki et al., 2005), consistent with the hypothesis that Xyn10C might serve as a functional homologue of the *Bacteroides thetaiotaomicron* VPI-5482 SusG protein, initiating xylan metabolism through extracellular hydrolysis of polymeric substrates (Dodd et al., 2011). In this regard, it has been proposed that Xyn10C could be used as a functional marker of xylan degradation in the human gut (Despres et al., 2016; Dodd et al., 2011; Rogowski et al., 2015). The potential roles and distributions of the Xyn10C have recently attracted considerable attention, but have not yet been fully described (Despres et al., 2016; Flint et al., 1997; Rogowski et al., 2015; Zhang et al., 2014).

Previously, a putative *xus* locus assigned to the genus *Bacteroides* was identified in a metagenomic library from the microbiome of a fungus-growing termite, *Pseudacanthotermes militaris* (Bastien et al., 2013). This *xus* is composed of eight different ORFs encoding putative XusC/D-like proteins, GH10|CBM4, GH115, GH11, a putative transporter protein, GH10 and GH43 (Fig.1A). The GH10|CBM4, designated hereafter as *Pm*25, presents an insertional modular structure, homologous to Xyn10C protein (Flint et al., 1997; McNulty et al., 2013; Zhang et al., 2014; Rogowski et al., 2011; Despres et al., 2016). Here we describe the characterization of *Pm*25 and discuss its activity with respect to its unusual multi-domain organization.

MATERIAL AND METHODS:

Materials

Beechwood GX, D-Xyl and most other reagents were purchased from Sigma-Aldrich (Saint-Quentin Fallavier, France). Low viscosity wheat arabinoxylan (Ara:Xyl=38:62, LVWAX), acid-debranched wheat arabinoxylan (Ara:Xyl=22:78, ADWAX), enzyme-debranched wheat arabinoxylan (Ara:Xyl=30:70, EDWAX), rye arabinoxylan (Ara:Xyl=38:62, RAX), galactomannan (carob; low viscosity), xyloglucan (tamarind), arabinan (sugar beet), β –glucan (barley; medium viscocity), xylobiose (X2), xylotriose (X3), xylotetraose (X4), xylopentaose (X5), xylohexaose (X6), p-nitrophenyl-β-D-xylopyranoside (pNPX), p-

nitrophenyl-β-D-xylobiose ($pNPX_2$), p-nitrophenyl-β-D-xylotriose ($pNPX_3$), p-nitrophenyl-β-D-xylotetraose ($pNPX_4$) were all purchased from Megazyme (Bray, Ireland). Cellulose nanocrystals from cotton linters were prepared as previously described (Martinez et al., 2015). Oligonucleotide primers were purchased from Eurogentec (Liège, Belgium) (Table S1).

Sequence analysis

The Genbank accession number for the clone containing Pm25 is HF548280.1, and the protein ID for Pm25 is CCO21036.1. Putative signal peptide sequence analysis was performed using SignalP 4.1 server (Petersen et al., 2011). The domain annotation of Pm25 was done using InterPro protein sequence analysis (https://www.ebi.ac.uk/interpro/) and the accession S0DFK9. Multiple protein sequence alignment of CBM4s was done using Clustal Omega at EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/) and the alignment of the secondary structure elements of Pm25 with other structurally-characterized GH10 family members was achieved using both Clustal Omega and ESPript 3 at http://espript.ibcp.fr/ESPript/ESPript/ (Robert et al., 2014).

Construction of sequence similarity networks (SSNs)

Amino acid sequences of GH10 family members were extracted using InterPro (https://www.ebi.ac.uk/interpro/) on May 17, 2017, and the entry number is IPR001000. To remove redundant sequences, the 8339 sequences were winnowed to 5068 by a sequence identity cut-off of 0.95 (Li and Godzik, 2006) and length cut-off of 250 (Gerlt et al., 2015). SSNs were constructed using the Enzyme Function Initiative-Enzyme similarity Tool (EFI-EST) (Gerlt et al., 2015) and visualized using Cytoscape 3.4 (Shannon et al., 2003). The proteins in one node share over 60% identity and the edge between two nodes indicates that they share over 35% identity (the e-value threshold is10^{-47}).

Expression, purification and site-directed mutagenesis of wild type Pm25 and its truncations

Cloning of the plasmid pDEST17 containing the Pm25-encoding sequence was described in detail elsewhere (G. Arnal et al. manuscript in preparation, (Vincentelli and Romier, 2013). All mutants were constructed using the QuikChange Site-Directed Mutagenesis Kit (Strategene, La Jolla, CA, USA) and pairs of complementary oligonucleotide primers (Table S1). The M6 construct, which corresponds to Pm25 deprived of its CBMs, was obtained by gene synthesis. The mutation of E546 to A in M6 yielded M7, while M8 and M9 were

constructed by ligating the sequences encoding CBM4-1 and CBM4-2 respectively to pET28a (+) expression vector. Likewise, the construct M10 is the pET32a (+) expression vector containing the sequence encoding both CBM4-1 and CBM4-2 cloned in frame with the thioredoxin tag using the In-Fusion cloning kit (Clontech, Takara, Shiga, Japan). Wild type Pm25 and the mutants M1, 2, 3, 4 and 5 (Fig. 4A) were expressed in Escherichia coli Rosetta^{TM} (DE3) pLysS grown in ZYP medium at 25°C overnight. Constructs M6 to M10 were transformed into E. coli Tuner^{TM} (DE3) and cultured 2 hours at 37°C until the OD_{600} reached 0.6. At this point, IPTG (200 µM final concentration) was added and growth was pursued at 16°C overnight. Cell pellets were collected by centrifugation, washed and lyzed using sonication (Fisherbrand^{TM} Q700; tip diameter, 13mm; output, 40W) and the clarified cell lysate was applied to TALON® metal affinity resin. After elution, protein purity was estimated by SDS/PAGE to be 95%. Protein concentrations were determined by measuring absorbance at 280 nm and applying the Lambert-Beer equation. Theoretical molar extinction coefficients were calculated using ProtParam online software (Gasteiger et al., 2005).

Determination of pH and temperature optima

The apparent optimal pH of Pm25 was determined in the pH range 3.0 to 11.0, measuring the enzyme activity (35 mg/L final enzyme concentration) on 1% (w/v) beechwood GX at 37°C. The buffers used were 50 mM citrate buffer for pH 3.0 to 6.0, 50 mM phosphate buffer for pH 6.0 to 8.0, 20 mM bicine buffer for pH 8.0 to 9.0 and 20 mM glycine-NaOH for pH 9.0 to 11.0. Xylanase activity was determined by measuring the release of reducing sugars using the 3,5-dinitrosalicylic acid (DNS) assay (Arnal et al., 2014; Miller, 1959). Reactions were performed in triplicate at 37°C in the different buffers from pH 3.0 to 11.0, containing bovine serum albumin (BSA, 1mg/ml). At regular intervals (0, 3, 6, 9, 12, 15, 18, 21, and 24 min), 100 µl of the reaction were removed and added to 100 µl of DNS and kept on ice until all samples were ready. After, all samples were heated at 95°C for 10 min and cooled on ice, before adding 1 ml of deionized water and recording the absorbance at 540 nm using a spectrophotometer. A D-xylose series (0 to 1 mg/ml) was used to prepare a standard curve. The apparent optimal temperature was determined over the range 21 to 90°C in 50 mM phosphate buffer (pH 7.5). Thermostability was monitored by pre-incubating the enzyme in the absence of substrate in 50 mM phosphate buffer (pH 7.5) at 45, 50, 55 and 60°C from 0 to 24 h. Residual enzyme activity in each case was then assayed as described above.

Enzyme specificity and kinetics

The final concentration of Pm25 used to degrade beechwood GX and other soluble substrates is 35 and 7 mg/L, respectively. Regarding the measurement of reaction kinetics on polysaccharides, initial rates (the concentration of D-Xyl equivalent released in mg/ml/min) were determined using a range of substrate concentrations (0.5 to 40 g/L beechwood GX and 0.25 to 10 g/L for other soluble substrates) under optimal conditions. The DNS assay was used to monitor reducing sugar release as described earlier. The kinetic parameters (k_{cat} and apparent K_M) were calculated using non-linear regression in SigmaPlot (Systat Software, San Jose, CA). One unit of xylanase activity was defined as the amount of enzyme that catalyses the release of 1 µmol of D-Xyl equivalents per min.

To study the hydrolysis of XOS (X_4-X_6), reactions were performed using various XOS concentrations (0.05 to 0.8 mM) and the optimal reactions conditions. Assays began upon the addition of enzyme, its final concentration being fixed to account for the nature of the substrate. Accordingly, 2.60, 0.26 and 2.60 µM enzyme was used for X_4, X_5, and X_6 respectively. At regular intervals (0, 5, 10, 15, 20, 30, 40, 50, and 60 min) aliquots were removed and immediately heated at 95°C for 10 min to stop the reaction. The hydrolyzed products were then analyzed by high-performance anionic-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) using an ICS 3000 dual (Dionex, France) equipped with Carbo-Pac PA-100 guard and analytical column (2*50mm and 2*250mm, respectively) as described before (Arnal et al., 2014). 10 µl of sample were injected, and separation was achieved by applying a gradient of 0 to 85 mM sodium acetate, 150 mM NaOH from 0 to 30 min, isocratic elution with 500 mM sodium acetate 150 mM NaOH from 30 to 33 min, then re-equilibrate the column with 50 mM sodium acetate, 150mM NaOH for another 10 min at a flow rate of 0.25 ml/min. Calibration was achieved using D-Xyl and XOS (X_2, X_3, X_4, X_5 and X_6) at concentrations from 5 to 50 µM. Plotting hydrolysis rate (µM/min) versus oligosaccharide substrate concentration (µM) yielded a linear relationship meaning that the catalytic constant k_{cat}/K_M could be calculated from the slope k using Eq. (1) where [E] is the final concentration of enzyme.

$$k = k_{cat}/K_{M*}[E] \quad (1)$$

All experiments were performed in triplicate, and reported values are the means of three experiments.

Determination of subsite affinities

The binding affinities of glycon subsites was calculated using Eq. (2) (Matsui et al., 1991; Song et al., 2014).

$$A_{-i} = \frac{RT}{4183} \ln \frac{(\frac{k_{cat}}{K_M})_{(pNPX_i)}}{(\frac{k_{cat}}{K_M})_{(pNPX_{i-1})}} (kcal \cdot mol^{-1}) \quad (2)$$

where A_{-i} is the subsite affinity at -*i* subsite, k_{cat}/K_M of *p*NPX$_i$ is the performance constant for *p*NP-labeled XOS with a DP of *i* (where *i* is a whole number) and *R* is the universal gas constant (8.314 J mol^{-1}K^{-1}).

To determine the catalytic parameters of reactions containing pNP-XOS, the final concentration of *Pm*25 used was 6, 27 and 136 nM for *p*NPX$_4$, *p*NPX$_3$, and *p*NPX$_2$ respectively. Similarly, the final concentration of M6 was 10 nM for *p*NPX$_4$ and *p*NPX$_3$, and 54 nM for *p*NPX$_2$. The concentration range of substrate was 0.025, 0.05 and 0.1 mM for *p*NPX$_3$ and *p*NPX$_4$, and 0.5, 2, and 5 mM for *p*NPX$_2$. All experiments were performed in duplicate. The plot of hydrolysis rate against *p*NP substrate concentration was linear, which indicated that substrate concentration was far below K_M. Therefore, the k_{cat}/K_M of reactions containing aryl β-xylosides was determined under optimum condition using equation (3) (Charnock et al., 1999; Matsui et al., 1991). Briefly, the substrate concentrations at the beginning of the reaction ([S$_0$]) and at specific times ([S$_t$]) were fitted to equation (3) where $k = (k_{cat}/K_M)$ [Enzyme]. And [Enzyme] is the final concentration of enzyme.

$$k = \ln[S_0]/[S_t] \quad (3)$$

The molar extinction coefficient of *p*NP (15 570 M^{-1}cm^{-1}) was determined experimentally by measuring the absorbance at 404 nm for a standard curve ranging from 0 to 0.12 mM of *p*NP at pH 7.5 and 50°C.

Affinity gel electrophoresis (AGE)

The binding of CBM4-1 and CBM4-2 to soluble polysaccharides was evaluated by affinity gel electrophoresis, using 7.5% (w/v) acrylamide gels containing various amounts of polysaccharide (for the concentration range of RAX refer to Table S2). AD-, ED- and LVWAX samples, and beechwood GX were used at 0.006 to 0.06% (w/v), while other polysaccharides were used at 0.5% (w/v). Pure protein (6 µg) was migrated (10 mA/gel for about 1 h at room temperature) on gels in 25 mM Tris, 250 mM glycine buffer, pH 8.3. BSA (15 µg) was also included in the experiment as a negative, non-interacting control. Proteins

were visualized by Coomassie Blue staining. The dissociation constant K_d was calculated as previously described (Takeo, 1984). In Eq (4),

$$\frac{1}{R_0 - r} = \frac{1}{R_0 - R_c}(1 + \frac{K}{c}) \quad (4)$$

R_0 is the relative protein migration distance compared to that of BSA in the control gel (without ligand). The variable r is the relative protein migration distance compared to BSA in ligand-containing gels. R_c is the relative protein migration distance of complex between protein and ligand. c is the concentration of ligand. When Eq (4) was plotted taking 1/ (R_0 - r) as the ordinate and 1/c as the abscissa, a straight line was obtained. The intercept of the line on the abscissa provided a negative reciprocal value of the dissociation constant (-1/K). All experiments were performed in triplicate, and reported values are the means of three experiments.

MicroScale electrophoresis (MST)

MST is a novel biophysical technology that is employed to study interactions between biomolecules (Wu et al., 2017). To fluorescently label the amine groups of exposed lysines (Fig. 1D) lying in the vicinity of the ligand binding clefts in M7, 8 and 9, 100 µl of pure proteins (20 µM) were treated with the reagents in the protein labeling kit (RED-NHS), following the manufacturer's instructions. The labeled proteins were recovered and purified using TALON® metal affinity resin and the concentration of the labelled proteins was estimated using SDS/PAGE and serial dilutions of a protein solution of known concentration. A total of 16 dilutions (350 to 11 mM) of a solution of X_6 containing either 0.075 µM of M7, 0.13 µM of M8 or 0.07 µM M9, in 50 mM phosphate buffer pH 7, 0.05% Pluronic® F-127 were loaded on to 16 standard capillaries. The initial fluorescence of all 16 samples was obtained by performing a capillary scan with LED power of 25% for M8, and 20% for M7 and M9. The dissociation constant K_d was calculated by selecting the tab "Initial Fluorescence Analysis Set " in the Affinity Analysis software (Jerabek-Willemsen et al., 2014). To perform the SDS-denaturation (SD)-test, 10 µl of samples 1-3 and 14-16 were mixed with 10 µl of 4% SDS, 40 mM DTT after 10 min centrifugation at 15,000 g, followed by 5 min incubation of the mixture at 95°C to denature the protein. Then the samples were loaded into the capillaries to measure their fluorescence intensity.

Solid depletion assay

The ability of inactivated M1, 7, 8 and 9 to bind wheat bran was investigated by incubating 100 µg of protein with 4 mg of wheat bran in 200 µl of reaction buffer (50mM sodium phosphate, pH 7). Reactions were performed in 0.2 ml PCR tubes, and incubated at 10°C for 2h with agitation in an Eppendorf Thermomixer™ R at 1400 rpm. For each reaction, the supernatant containing the unbound enzyme fraction was recovered after centrifugation using a bench top microcentrifuge. The pelleted substrate was washed 3 times with reaction buffer. Finally, 20 µl of Laemmli sample buffer were added to the pellet and heated at 95°C for 10 min to denature the protein (bound fraction). All the fractions were verified by SDS-PAGE. BSA was used as a negative control.

Hydrolysis of wheat arabinoxylan and wheat bran

The product profile were generated by *Pm*25, M3, 4, 5 and 6 on either LVWAX (0.5% (w/v)) or wheat bran (20 mg/ml of wheat bran pre-hydrated for 12 h at 37°C, 1400 rpm using the Eppendorf Thermomixer™ R). The enzymes (final concentration 0.5 µM) were incubated with the respective substrate in 50 mM phosphate buffer (pH 7.5) and 1 mg/ml of BSA. Enzymatic reactions were incubated at 37°C for either 24 h for LVWAX or 14 h for wheat bran and aliquots were removed at regular time intervals and heated at 95°C for 10 min to terminate the reaction. Each sample was centrifuged at 20,000 x g for 5 min and quantified by HPAEC-PAD on a Dionex PA1 column equipped with Carbo-Pac PA-1 guard and analytical columns (4*50mm and 4*250mm, respectively). Separation of oligosaccharides was achieved by isocratic elution with 100 mM NaOH at a flow rate of 1 ml/min from 0 to 5 min, a gradient of 0 to 120 mM sodium acetate in 100 mM NaOH from 5 min to 25 min, and isocratic elution with 500 mM sodium acetate in 100 mM NaOH from 25 min to 35 min. Then the column was re-equilibrated with 100 mM NaOH for another 10 min. Calibration was achieved using D-Xyl and XOS (X_2, X_3, X_4, X_5 and X_6) at concentrations from 5 to 100 µM.

RESULTS:

Bioinformatic analysis of the Pm25-encoding gene sequence

Analysis using SignalP 4.1 revealed that the first 32 amino acids of *Pm*25 probably constitute a signal peptide. Moreover, further analysis of the sequence revealed that the

catalytic domain of *Pm*25 (amino acid residues 66 to 151, and 514 to 753) contains two putative tandem CBM4s (CBM4-1, residues 161 to 321 and CBM4-2, residues 324 to 486) that constitute insertional domains . Alignment of the amino acid sequence of *Pm*25 with those of other GH10 family members for which structural data is available revealed that CBM4-1 and CBM4-2 are inserted between β3 and α3 (Fig.1B). Moreover, this alignment allowed the identification of E546 and E663 to be putative catalytic acid/base and nucleophile, respectively. Alignment of the amino acid sequences of CBM4-1/CBM4-2 with the well-characterized tandem CBMs, 2Y6G from Rhodothermus marinus Xyn10A, 1GU3 from cellulomonas fimi Cel9B, 1GUI Thermotoga maritima Lam16, 3K4Z from Clostridium thermocellum Cbh9A, and 3P6B from Clostridium thermocellum Cel9K, revealed modest similarity (16/19,15/15, 16/19, 20/21, and 18/21 % similarity respectively). Moreover, using this similarity and previous reports, it was possible to assign possible ligand binding role to residues Y213, Q216, and Y257 in CBM4-1, and Y378, Q381, and Y422 in CBM4-2 (Alahuhta et al., 2010, 2011; Boraston et al., 2002b; Fisher et al., 2015; Kormos et al., 2000; Von Schantz et al., 2012) (Fig.1C).

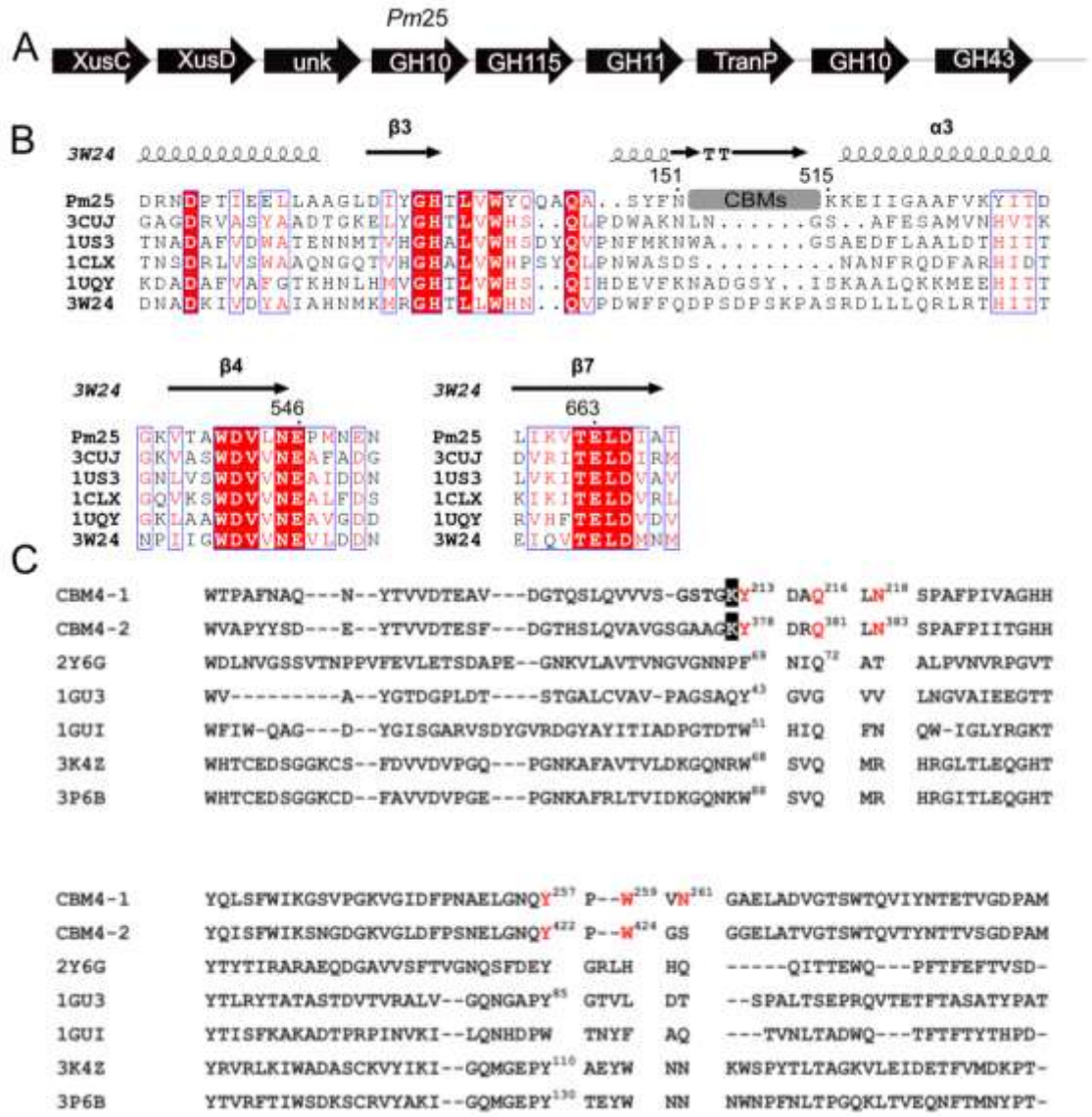


Figure 1. Modular architecture of Xyn10C-like protein. (A) Schematic representation of the putative xylan-active gene cluster. (B) Amino acid alignment of *Pm*25 with other native GH10 xylanases designated by their PDB codes. PDB: 3CUJ, *Cellulomonas fimi* Xylanase/Cellulase Cex (*Cf*Xyn10A); 1US3, *Cellvibrio japonicus* xylanase 10C; 1CLX, *Pseudomonas fluorescens* xylanase A; 1UQY, *Cellvibrio mixtus* Xylanase B; 3W24, *Thermoanaerobacterium saccharolyticum* xylanase A; The β strands and α helices are shown above the aligned sequences. Numbering corresponds to the sequences of *Pm*25. Residues in red background are conserved and within blue frames are conservatively substituted. (C) Amino acid alignment of CBM4s in *Pm*25 with well characterized CBM4s in literature. CBM4-1 is the first CBM4 in *Pm*25; CBM4-2 is the second CBM4 in Pm25; PDB: 2Y6G, CBM4 from Rhodothermus marinus xylanase; 1GU3, CBM4 from Cellulomonoas fimi endoglucanase C; 1GUI, CBM4 from Thermotoga maritima laminarinase 16A; 3K4Z, CBM4

from Clostridium thermocellum cellulase CbhA; 3P6B, CBM4 from Clostridium thermocellum CelK. Residues in red were used for mutagenesis. Numbered residues in PDB sequences indicated that they are responsible for ligand binding according to literatures. The K in white is potentially labelled with NHS-dye which was described in the methods for MST experiment.

SSN analysis

To identify *Pm*25 homologs and study their distribution, a SSN was created for all GH10 sequences (5068 non-redundant sequences) present in the InterPro database (Fig.2). Archaea, Eukaryota and Bacteria contribute 1, 33, and 65 % respectively of GH10s, the remaining 1% being unassigned sequences. Bacterial GH10s belong mainly to four bacterial phyla, Actinobacteria (20%), Firmicutes (13%), Bacteroidetes (11%), and Proteobacteria (11%).

The SSN showed five main clusters, numbered following each cluster size (Fig.2). 61% and 35% of Eukaryotic GH10s are grouped in cluster 1 and 2, respectively. Archaeal GH10s are mostly gathered together in cluster 2 closer to Eukaryotic ones. Although bacterial sequences are present in all the 5 clusters, most of GH10s from Actinobacteria (92%), Firmicutes (83%), and Proteobacteria (75%) are grouped in cluster 1 while 45% of GH10s from Bacteroidetes are grouped in cluster 1 and 30% are clustered in cluster 4. Interestingly, cluster 4 apart from one unclassified sequences (UniprotKB: A0A0H5Q2Q6), comprises only Bacteroidetes sequences including *Pm*25. In this cluster that is composed of 79 nodes, 48 of them display two insertional CBMs, rather like *Pm*25, and 29 contain just one insertional CBM. Regarding the two remaining nodes in the cluster, one lacks GH10a (i.e. the N terminal part of GH10, UniprotKB: W7XZD0) and the other corresponds to a single GH10 catalytic domain (e.g. UniprotKB: A0A1H3ZQ12). Furthermore, all the GH10s with insertional architecture based on InterPro domain classification were found in the cluster 4. In complement, a separate SSN analysis of the 589 bacterial non-redondant CBM4s revealed a high level of scattering, no group formed containing CBM4s in *Pm*25 probably due to poor sequence similarity (data not-shown), reinforcing that the cluster 4 (Fig. 2) is grouped based on the GH10 catalytic domain.

Regarding the genetic context of the Bacteroidetes GH10-encoding sequences (in cluster 4), 73 out of the 79 nodes are consecutive to hypothetical *susC-susD*-(*unk*) gene clusters, meaning that these are likely to be PUL components. The search for CBM4-containing GH10

enzymes in the extended PULDB (http://www.cazy.org/PULDB/) further confirmed that *Pm*25 homologs are always located consecutive to *sus*C-*sus*D(-unk) clusters.

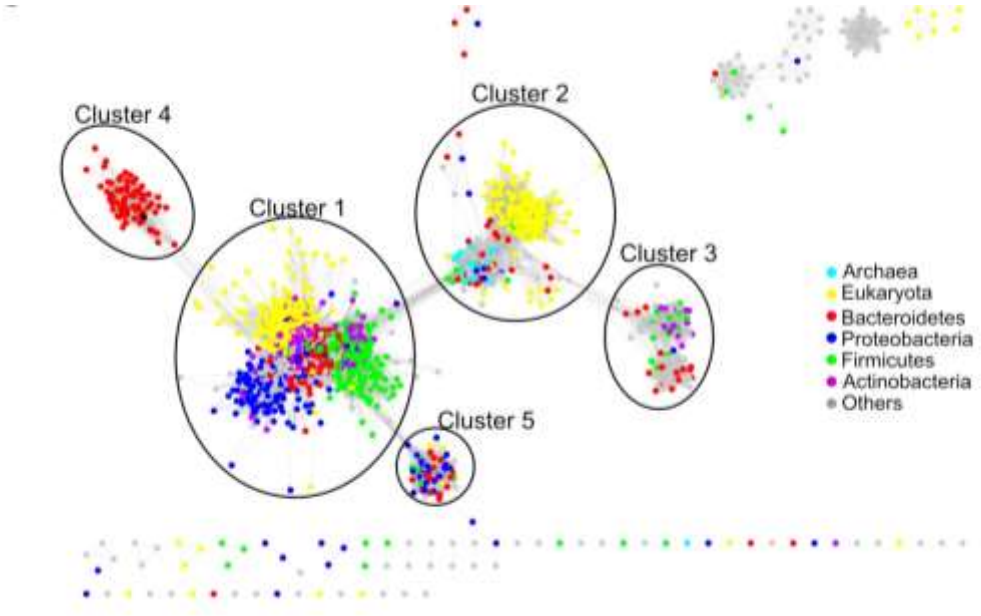


Figure 2. Taxonomic distribution of GH10 xylanase. The nodes are colored by the taxonomic annotation listed on the right.

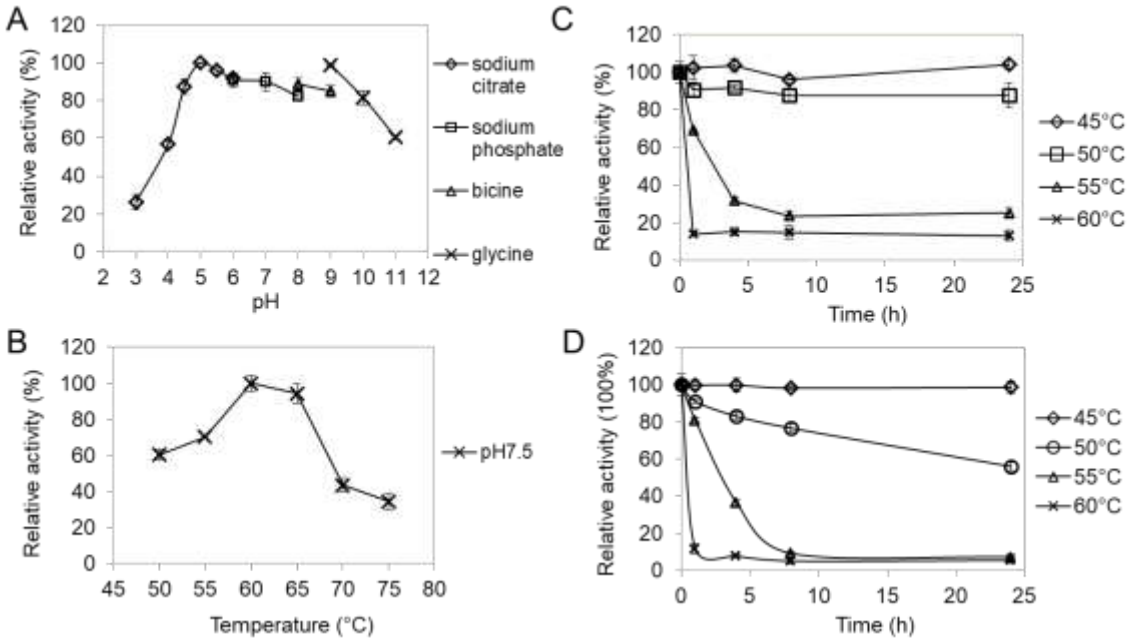


Figure 3. Optimal physiochemical parameters. (A). Optimum pH. (B) Thermoactivity at pH 7.5. (C) Thermostability at pH 7.5. (D) Thermostablity at pH 9.0.

Optimal pH and temperature

The optimal pH of *Pm*25 was determined using beechwood GX as the substrate. The purified enzyme retained greater than 80% of its activity in the pH range from 4.5 to 9.0 (Fig.3A). The activity was also measured at temperatures from 50 to 75°C, with maximum activity being observed at 60°C (Fig.3B). However, since *Pm*25 was rather unstable at 60°C, thermostability was measured at both pH 7.5 and pH 9 (Fig.3C and 3D respectively). *Pm25* remained fully active for over 24 hours at either 50 °C pH 7.5 or 45 °C pH 9. Overall, optimal conditions for routine *Pm*25 assays were defined as pH 7.5 and 50°C.

Enzyme assays and Kinetic Analysis

Determining the hydrolytic activity of *Pm*25 on various substrates revealed that it is 3 to 4-fold more active on arabinoxylans than beechwood GX, indicating that the latter is the least suitable substrate among those tested (Table 1). When comparing the relative activities of *Pm*25 on different arabinoxylans, no significant differences were detected, although RAX, which possess more O-3 substitutions (Comino et al., 2014), qualifies as the best substrate tested. Testing the activity of the inactive mutants (M1 and 2, Fig. 4) on different substrates revealed that they both displayed much lower (2 to 3 orders of magnitude) activity when compared to *Pm*25 (Table S3).

Table 1. Kinetic parameters of *Pm*25.

Substrate	Activity (U/mg)	K_{Mapp} (mg/ml)	k_{cat} (/s)	k_{cat}/K_M (/min/mM)
Beechwood GX	7.4 ± 0.1	3.1 ± 0.2	10.3 ± 0.2	
LVWAX	25.8 ± 0.8	2.0 ± 0.2	36.2 ± 1.1	
ADWAX	21.7 ± 0.5	1.0 ± 0.1	30.4 ± 0.7	
EDWAX	21.4 ± 0.4	1.7 ± 0.1	29.9 ± 0.5	
RAX	30.3 ± 0.6	1.4 ± 0.1	42.4 ± 0.8	
X_6				292.8 ± 36.8
X_5				35.5 ± 2.4
X_4				3.75 ± 0.1
*p*NPX_4				2381.7 ± 381.7

*p*NPX$_3$	1041.4 ± 138.2
*p*NPX$_2$	14.09 ± 0.66

K_{Mapp}, apparent K_M.

Hydrolysis product analysis with XOSs

HPAEC-PAD analysis of reactions containing different XOS and *Pm*25 revealed that the hydrolysis of X_6 (after 60 min) produced a mixture of X_2, X_3 and X_4 in a ratio of 1:2:1 (Fig.5A). Likewise, the hydrolysis of X_5 led to the production of X_2 and X_3 as major products (Fig.5B), while the hydrolysis of X_4 yielded X_1 and X_3 (Fig.5C). Moreover the activity of *Pm*25 was directly correlated with the DP of the XOS used, with higher DP XOS leading to higher activity (Table 1). This suggests that the *Pm*25 substrate binding cleft is quite large and can accommodate at least six xylosyl residues. To further study binding cleft subsite interactions, reactions were performed using *Pm*25 and different aryl-β-xylosides.

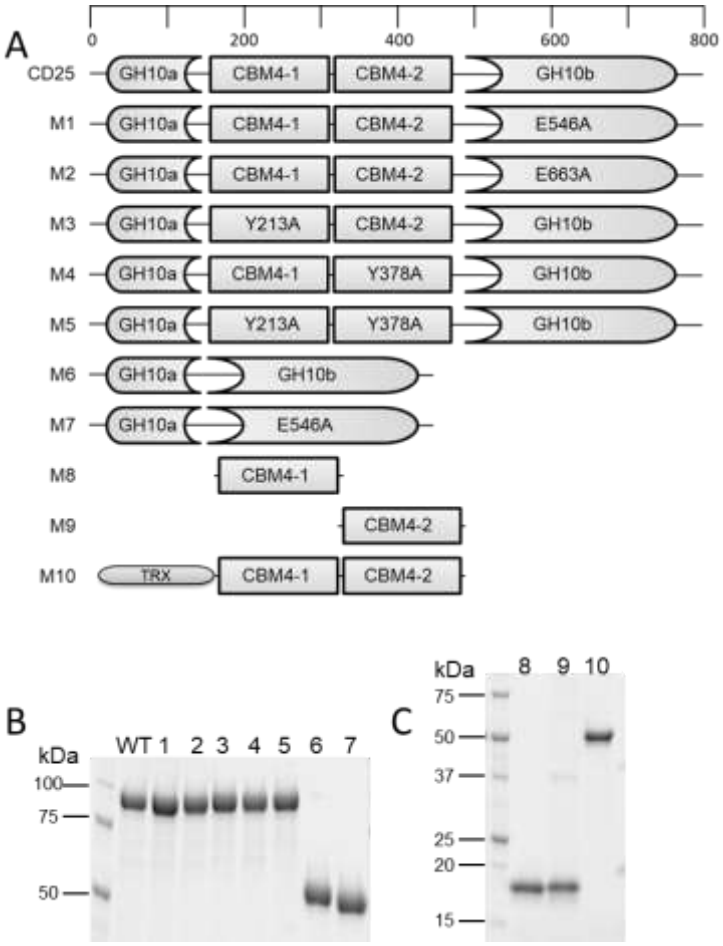


Figure 4. Mutational and truncation constructs. (A) Domain organization of *Pm*25 (WT) and modified constructs. (B) SDS-PAGE of the purified WT and M1-M7. (C) SDS-PAGE of the purified CBM constructs M8-M10.

Accordingly, the efficiency constant (k_{cat}/K_M) obtained when using *p*NPX$_4$ as substrate was approximately 2.3-fold greater than that measured when using *p*NPX$_3$ (Table 1). These data provided the basis to calculate the binding affinity at -4 subsite, revealing a value of 0.53 kcal/mol. Likewise, the -3 subsite binding affinity is 2.76 kcal/mol, which correlates with the fact that the k_{cat}/K_M value for the hydrolysis of *p*NPX$_3$ is over 70-fold higher than that measured for the hydrolysis of *p*NPX$_2$ (Table 1). Furthermore, the use of construct M6 (i.e. *Pm*25 devoid of CBMs) (Fig.4) to hydrolyze aryl-β-xylosides revealed similar affinities at subsites -4 and -3 (0.75 and 2.69 kcal/mol, respectively), inferring that the two CBM4 domains do not influence catalysis when using these substrates.

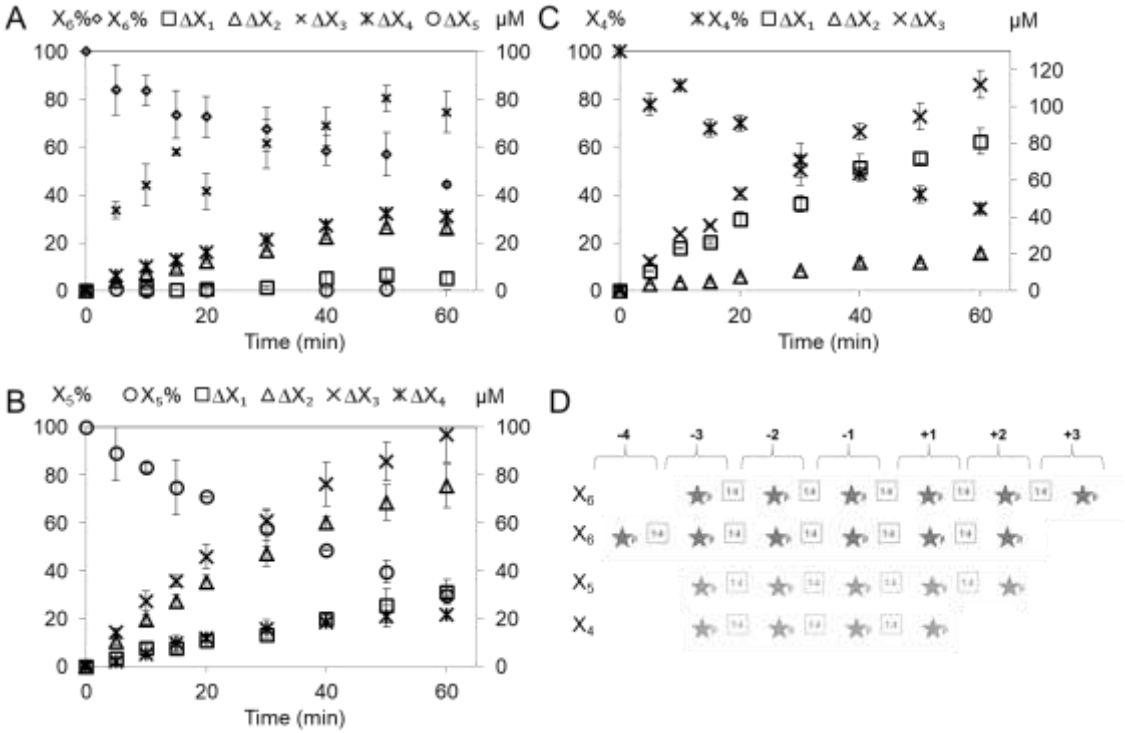


Figure 5. Progressive degradation of 0.2mM of XOS, (A) X6, (B) X5, (C) X4 by Pm25. (D) The potential subsite mapping of *Pm*25 with X6, X5 and X4.

Measuring CBM ligand specificity

To probe the function of CBM4-1 and -2, various constructions M1, M7, M8, M9 and M10 were designed and expressed (Fig.4) and affinity gel electrophoresis experiments were

performed. Since the results of bioinformatics analysis putatively assign CBM4-1 and -2 to CBM family 4, initial tests were performed on xylan and glucan, before performing complementary tests with other polysaccharides (Table 2). All of the proteins displayed strong affinity with the different xylans tested. However, very low or no affinity was detected for β-glucan and xyloglucan and none of the proteins bound to arabinan, nanocellulose and galactomannan (Table 2). Importantly, in the presence of different xylans the K_d of M10 (CBM4-1 and -2 in tandem) was much lower than that of M8 (CBM4-1 alone) and M9 (CBM4-2 alone), suggesting that there might be cooperativity between the two CBMs. Comparing the behavior of the inactive mutants M1 and M7 with that of M10 revealed that the affinity of the inactive *Pm*25 M1 for xylans was similar to that exhibited by the M10, CBM4-1/-2 tandem, whereas the K_d values obtained for the inactive enzyme devoid of CBMs (i.e. M7) were significantly higher. This suggests that xylan binding is largely driven by the two CBM4 domains. Comparing M8 and M9 revealed that M9 systematically exhibited lower (1.4 to 1.8 times) K_d values for tests involving xylans, inferring that its binding ability is stronger (Table 2). However, the results obtained with other polysaccharides indicate that CBM4-1 weakly binds to xyloglucan, while M9 (CBM4-2 alone) does not (Fig.S1). Overall, the results suggest that binding specificities of CBM4-1 and -2 are not identical.

To further probe the ligand binding ability of the two *Pm*25-associated CBM4 domains, different residues putatively involved in ligand binding were mutated. Accordingly, the binding abilities of both CBM4-1|Y213A and CBM4-2|Y378A were lost, confirming that these residues play essential roles in the CBM-ligand interaction (Fig.6). The binding abilities of CBM4-1|Y257A and CBM4-2|Y422A were also diminished, but ligand interactions were still observable, indicating that these tyrosines play less critical roles compared to Y213 and Y378 respectively (Fig.6A and B). Likewise, other mutants such as CBM4-1|Q216A and CBM4-1|W259A, and CBM4-2|Q381A and CBM4-2|W424A also diminished to some extent ligand binding, but the mutation of asparagines (N218, N261 in CBM4-1 and N383 in CBM4-2) had no apparent effect on binding althought these residues are close to the essential ones

The determination of K_d values for the interaction of M7, M8 and M9 with X_6 was achieved using MST. The sigmoidal titration curves were used to calculate K_d values (Fig.S2). The K_d values obtained when using M7, 8 and 9 were 1.5 ±0.2 mM, 4.7 ± 0.4 mM, and 1.8 ± 0.1 mM, respectively. Recalling that M7 is devoid of CBM domains, it is noteworthy that this construction displayed the highest binding ability, with M9 (i.e. CBM4-2) exhibiting a similar ligand binding ability.

Finally, substrate depletion experiments performed using M1, 7, 8 and 9 on wheat bran revealed that M8 and M9 (CBM4-1 and CBM4-2 respectively) exhibited binding ability. Similarly, M1 (i.e. *Pm*25|E546A) was also able to bind to wheat bran despite its catalytic impotency (Fig.S3). However, this was not the case for M7 (i.e. the inactivated catalytic domain alone) clearly demonstrating the role of the CBM in the binding.

Table 2. Binding of the recombinant proteins to soluble substrates as determined in affinity gel electrophoresis.

The unit for K_d is in mg/ml; + , retardation observable; -, no retardation.

Ligand	M1	M7	M8	M9	M10
			K_d (10^{-2})		
RAX	2.8 ± 0.1	18.8 ± 3.5	31.9 ± 0.8	18.1 ± 2.2	5.1 ± 0.4
LVWAX	3.7 ± 0.3	27.1 ± 1.9	22.7 ± 1.1	15.9 ± 1.9	3.7 ± 0.2
ADWAX	2.3 ± 0.2	15.6 ± 0.2	10.9 ± 0.7	8.0 ± 0.1	6.4 ± 0
EDWAX	2.0 ± 0.1	16.7 ± 0.4	12.6 ± 0.8	9.2 ± 0.4	3.2 ± 0.0
Beechwood GX	4.9 ± 0.3	62.0 ± 1.9	30.3 ± 1.2	17.5 ± 0.1	9.0 ± 0.5
Nanocellulose	-	-	-	-	-
Galactomannan	-	-	-	-	-
Xyloglucan	+	+	+	-	+
Arabinan	-	-	-	-	-
β-glucan	+	-	+	+	+

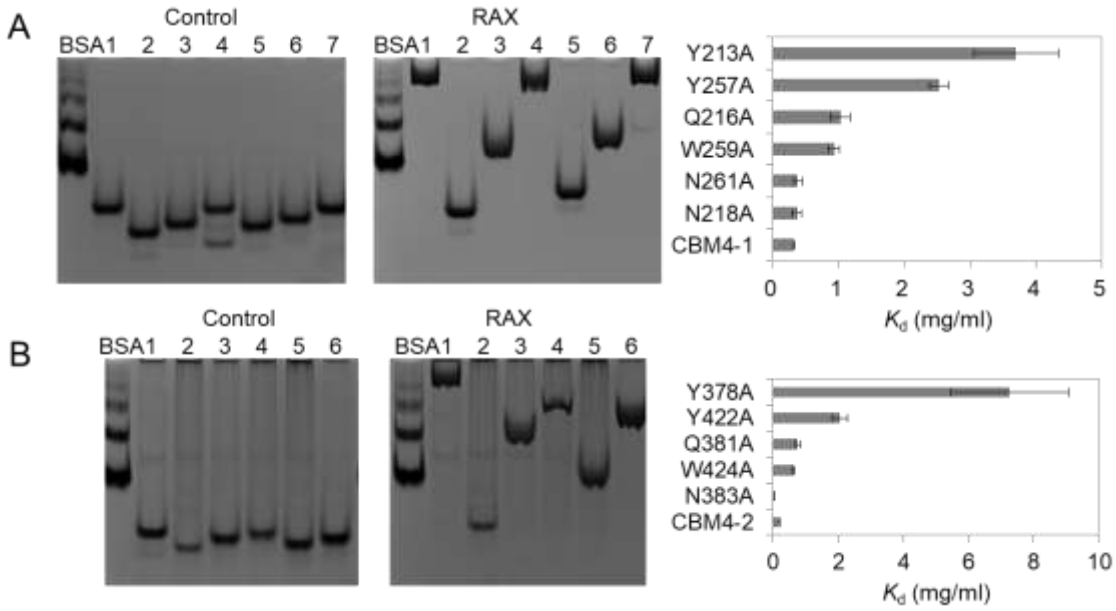


Figure 6. Affinity gel electrophoresis of different CBM mutations in Pm25. The control gel is lack of ligand, while the other one is with 0.06% (w/v) of RAX. (A) Different retardation of CBM4-1(M8) and its mutants. 1, M8; 2, Y213A; 3, Q216A; 4,N218A; 5,Y257A; 6,W259A; 7,N261A. The K_d values for each constructs are on the right. (B) Different retardation of CBM4-2 (M9) and its mutants. 1, M9; 2, Y378A; 3, Q381A; 4, N383A; 5, Y422A; 6, W424A. The K_d values for each constructs are on the right.

Analysis of polysaccharide and wheat bran hydrolysis

HPAEC-PAD of soluble polysaccharide hydrolysis mediated by Pm25 and its variants M5 and M6 failed to reveal significant differences in D-Xyl and XOS release (Fig.7 and S4). This infers that the CBM4 domains do not enhance degradation of soluble polysaccharides, since M6 is devoid of CBM4 domains. Performing a similar analysis using wheat bran as the substrate revealed that (after 14h incubation) D-Xyl and XOS release by Pm25 was approximately twice that of the variants M3, 4, 5 and 6 (Fig.8). Significantly, in this experiment the consequences of the point mutations CBM4-1|Y213A and CBM4-1|Y378A were approximately equivalent to those produced by the ablation of the two CBM4 domains.

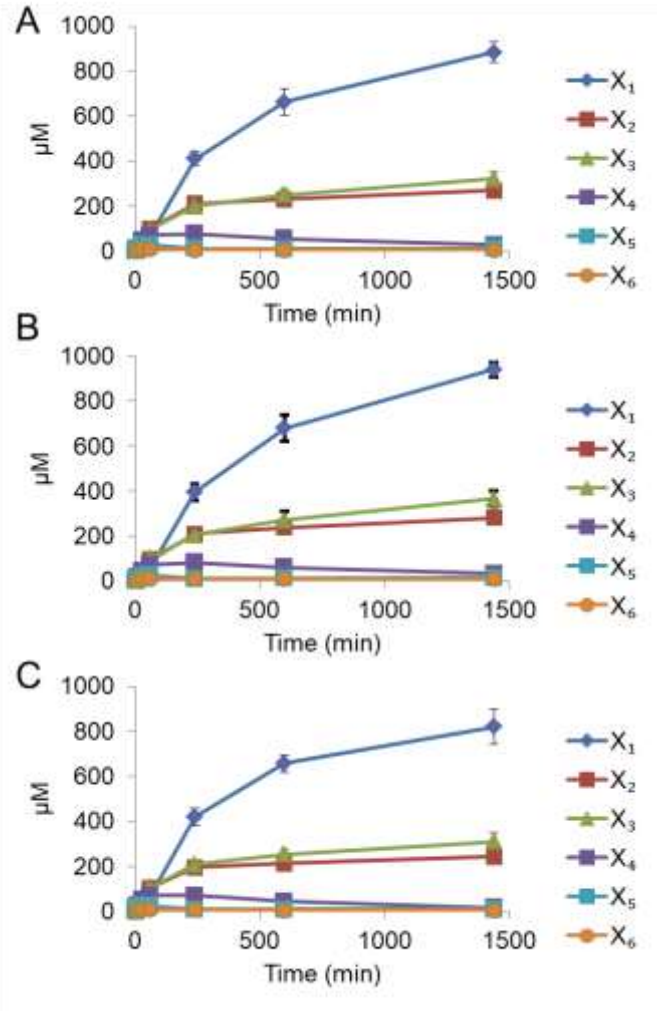


Figure 7. Hydrolysis of LVWAX by Pm25 and its mutants. The time-dependent degradation of LVWAX by (A) *Pm*25 (WT), (B) M5 and (C) M6.

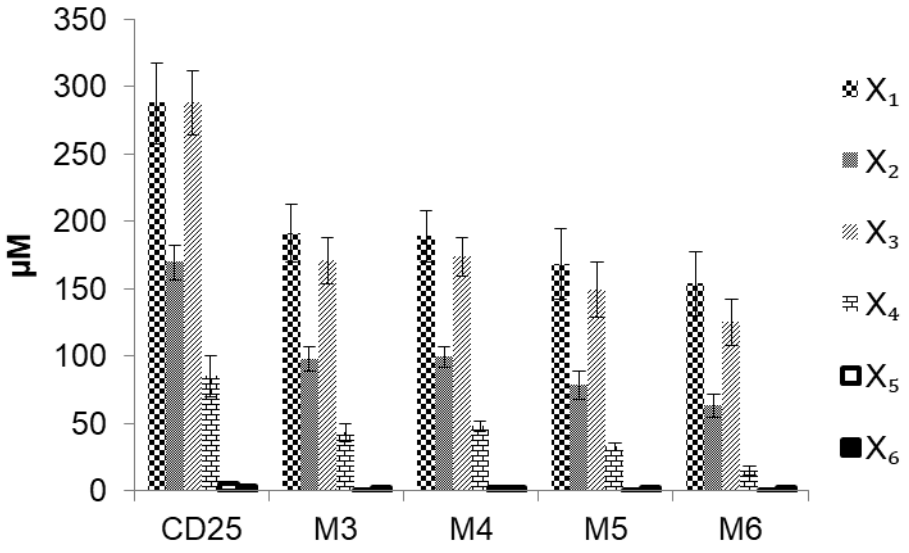


Figure 8. Hydrolytic patterns of *Pm*25 (WT) and mutants towards wheat bran.

DISCUSSION:

Unlike the vast majority of multi-modular enzymes that display a sequential arrangement of their modules, the enzyme described herein is characterized by a discontinuous organization that involves the insertion of two CBM domains into one GH10 xylanase domain. In this regard, it is significant that the SSN analysis performed using the amino acid sequence of M6 replacing *Pm*25 located the sequence within the same cluster, even though the CBMs were omitted. This infers that the *Pm*25 GH10 domain forms part of a distinct group and implies that the intercalated GH10 arrangement is robust from an evolutionary standpoint. Moreover, the biochemical data described herein demonstrate that, despite its discontinuous organization, *Pm*25 is a fully functional xylanase.

The first *Pm*25 analog was identified in a rumen-based member of the Bacteroidetes phylum (Flint et al., 1997). More have since been found in human gut bacteria (Despres et al., 2016; Dodd et al., 2010; Rogowski et al., 2015), with *Pm*25 being the first described in termite gut. Several studies have revealed the importance of *Pm*25-like GH10 in xylan utilization systems (Despres et al., 2016; Flint et al., 1997; Rogowski et al., 2015; Zhang et al., 2014). Using SSN analysis, we have shown that *Pm*25-like xylanases are exclusively linked to Bacteroidetes and are almost always (73 out 79 cases studied) adjacent to a *susC-susD*-(*unk*) cluster. This evidence of strong conservation is consistent with the fact that in their native host the genes encoding *Pm*25 homologs are highly induced/expressed during growth on xylan (Despres et al., 2016; Dodd et al., 2010). Taken together, one can conclude that Xyn10C-like proteins are essential for xylan utilization by members of the Bacteroidetes phylum in the gut ecosystem.

The *in vivo* function of *Pm*25 homologs in gut Bacteroidetes has not yet been fully established, although it has been suggested that it may be a functional homolog of SusG (Dodd et al., 2011). SusG is a cell surface-bound GH13 α-amylase that catalyzes the initial cleavage of polysaccharides (Shipman et al., 1999). In our study, we also predicted that *Pm*25 bears a N-terminal signal peptide that might direct it to the cell surface, consistent with a proposal that was previously made for a *Pm*25 homolog (Rogowski et al., 2015). Moreover, SusG displays negligible activity compared to periplasmic α-amylases (Shipman et al., 1999), an observation that is consistent with our findings (unpublished). Indeed, when compared to other xylanases (Chakdar et al., 2016; Pell et al., 2004a), both *Pm*25 and its alike display quite poor catalytic efficiency towards polysaccharides (Wang et al., 2016a; Zhang et al., 2014) and oligosaccharides (Rogowski et al., 2015). This trend is also observed in other polysaccharide

degrading systems, such as mannan utilization loci from members of the Bacteroidetes phylum (Bågenholm et al., 2016) and the xylan degrading system in Proteobacteria (Pell et al., 2004b) phylum. The underlying reason for such low activity no doubt reflects its function. SusG-like proteins probably play a carbohydrate surveillance function, while highly active intracellular enzymes are charged with complete oligosaccharide breakdown prior to sugar metabolism. This clever and "selfish" strategy ensures that readily metabolizable sugars are not released into the environment where they can be used by other bacteria that lack specific glycan utilization machinery (Martens et al., 2014).

Remarkably, we found that *Pm*25 remains active over a broad pH range, maintaining more than 80% of its maximum activity at pH 9.0. This observation correlates well with results obtained for the *Pm*25 homologs *Bi*Xyn10C and *Bi*Xyn10A, which were identified in the human gut microbiome (Wang et al., 2016a). Accounting for the fact that alkaline-stable xylanases are sought after for use in applications such as paper pulp biobleaching, *Pm*25 might constitute a useful starting point for enzyme engineering aimed at improving its hydrolytic properties.

So far, we have been unable to obtain structural data pertaining to *Pm*25 and none is available for its closest homologs. Therefore, at this stage it is tricky to speculate on the exact topology and molecular determinants of its active site. Nevertheless to gain some understanding we have examined similarities with the family GH10 xylanase *Cj*Xyn10C, which displays approximately 30% identity with *Pm*25 and whose structure is known (PDB: 1US3). Like *Pm*25, *Cj*Xyn10C exhibits rather poor activity on XOS ascribed to weak substrate binding in subsite -2 (Pell et al., 2004b). Unlike most other GH10 enzymes, in *Cj*Xyn10C subsite -2 contains G295 in the place of E, whose sidechain can hydrogen bond to the substrate According to sequence alignment, *Pm*25 also lacked the vital E residue in subsite -2, an observation that might explain its poor ability to hydrolyze X_4 (Charnock et al., 1999; Pell et al., 2004b). Therefore, the -3 subsite with rather strong affinity value (2.76 kcal/mol) compared to others (Charnock et al., 1999) could probably be involved in the glycone subsite in the degradation of X_4 in order to in complement with the poor -2 subsite. Taken together, a hypothetical subsite mapping of the active site of *Pm*25 with XOS is proposed for *Pm*25 (Fig.5D).

The two CBM4s that are inserted into *Pm*25 clearly contribute to the binding and degradation of complex biomass. Our results reveal that this is especially true when both CBMs are functional and suggest that binding of large ligands involves a cooperativity phenomenon (Fig 8). However, some *Pm*25 homologs only contain one CBM domain. Similarly, database searching reveals that other *Pm*25 homologs contain two members of CBM family 22 in place of the CBM4 domains. Recently a multi-domain GH10 discovered from *Bacteroides clarus* in

chicken caecum was integrated into UniProtKB (accession number A0A1Y3YZ70) and shown to incorporate three CBM4s inserted into GH10 amino acid sequences. These exceptional enzymes are also found in putative XULs from gut Bacteroidetes isolate (http://www.cazy.org/PULDB/, McNulty et al. 2013). This infers that the SusG-assimilated functions can be fulfilled by enzymes that are not configured in an identical way. Moreover, it also confirms that the TIM-barrel fold in the GH10 family is quite accommodating in terms of insertions at the β3/α3 loop.

Apparently, unlike many highly active periplasmic endoglucanases, such as SusA (Shipman et al., 1999) and *Cj*Xyn10D (Pell et al., 2004b), extracellular enzymes such as SusG, *Cj*Xyn10A and *Cj*Xyn10C are generally appended to CBMs (Pell et al., 2004b). Therefore, it is of interest to discuss the reason for this. The CBM58 in SusG (Koropatkin et al., 2010) and the CBM4s in *Pm*25 appear to improve the ability of the enzymes to hydrolyze insoluble substrates (Fig 8), while CBM15 in *Cj*Xyn10C does not play an important role in catalysis, irrespective of whether the substrate is soluble or not (Pell et al., 2004b). However, our data suggest that the affinity of *Pm*25 for soluble substrates was mostly derived from the binding ability of the CBMs (Table 2). In light of this observation, we propose that CBMs in membrane-associated enzymes might temporarily withhold soluble oligosaccharides before their importation into the cell. The implication of this is that the function of the CBM4 domains would be relatively independent of that of the GH10 domain. In this regard, it is noteworthy that the first structure of a SusG protein has been reported (Koropatkin et al., 2010). This structure reveals that the CBM58 domain is inserted into the B domain of the GH13 α-amylase domain. However, from a spatial point of view it is remarkable that the CBM58 does not form hydrogen-bonds with the catalytic domain, an observation that argues in favor of an independent function. Regarding *Pm*25, evidence for an independent function of the CBM4 domains is provided by the fact that the xylan-degrading profile of *Pm*25 wild type was almost identical to that of the CBM deleted version, M6 (Fig 7) and the fact that the xylan binding affinity of CBMs was relatively unaltered when the CBM domains were separated from the GH10 domain (Table 2). Finally, it is also useful to recall that the affinity values determined for subsites -4 and -3 of *Pm*25 and M6 were also nearly identical. Therefore, we believe that the catalytic center of *Pm*25 and the binding surfaces of the CBM4 domains are disconnected, an organization that corresponds to independent functions and contributes to low enzyme reaction rates (Khosla et al., 2001).

In conclusion, focusing on a termite gut-derived enzyme, we have provided further insight into the properties and function of Xyn10C-like enzymes that form part of core xylan utilization systems. This system seems to be rather efficient in term of evolution since it is

conserved in termite gut, rumen, and human gut. Therefore the role of the CBM insertion is an interesting question. In this respect, we have thoroughly succeeded in characterizing the enzyme and shown that the CBM4 domains can be successfully excised without loss of catalytic function. Regarding the enzyme's substrate specificity, although it is difficult to speculate on the group of polysaccharides that might be preferential substrates in the termite gut environment, we have shown that it is better adapted for the hydrolysis of arabinoxylans than glucuronoxylans, which is consistent with the fact that the host termite feeds on crops such as sugarcane, rather than wood.

Reference

Alahuhta, M., Xu, Q., Bomble, Y.J., Brunecky, R., Adney, W.S., Ding, S.Y., Himmel, M.E., and Lunin, V. V. (2010). The unique binding mode of cellulosomal CBM4 from clostridium thermocellum cellobiohydrolase A. J. Mol. Biol. 402, 374–387.

Alahuhta, M., Luo, Y., Ding, S.Y., Himmel, M.E., and Lunin, V. V. (2011). Structure of CBM4 from Clostridium thermocellum cellulase K. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 67, 527–530.

Arnal, G., Bastien, G., Monties, N., Abot, A., Leberre, V.A., Bozonnet, S., O'Donohue, M., and Dumon, C. (2014). Investigating the function of an Arabinan utilization locus isolated from a termite gut community. Appl. Environ. Microbiol. 81, 31–39.

Aroul-Selvam, R., Hubbard, T., Sasidharan, R., Aroul-Selvam, R., Hubbard, T., and Sasidharan, R. (2004). Domain insertions in protein structures. J. Mol. Biol. 338, 633–641.

Bågenholm, V., Reddy, S.K., Bouraoui, H., Morrill, J., Kulcinskaja, E., Bahr, C.M., Aurelius, O., Rogers, T., Xiao, Y., Logan, D.T., et al. (2016). Galactomannan catabolism conferred by a polysaccharide utilisation locus of Bacteroides ovatus : enzyme synergy and crystal structure of a β-mannanase. J. Biol. Chem. 292, 229–243.

Bastien, G.G., Arnal, G.G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., Henrissat, B., Lefèvre, F., Robe, P., Bouchez, O., et al. (2013). Mining for hemicellulases in the fungus-growing termite Pseudacanthotermes militaris using functional metagenomics. Biotechnol. Biofuels 6, 78.

Bignell, D.E. (2010). Morphology, Physiology, Biochemistry and Functional Design of the Termite Gut: An Evolutionary Wonderland. In Biology of Termites: A Modern Synthesis, (Dordrecht: Springer Netherlands), pp. 375–412.

Bolam, D.N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M.P., Rixon, J.E., Boraston, A., Hazlewood, G.P., and Gilbert, H.J. (1998). Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem. J. 331, 775–781.

Boraston, A.B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D.R., Kilburn, D.G., and Davies, G.J. (2002). Differential oligosaccharide Recognition by Evolutionarily-related β-1,4 and β-1,3 Glucan-binding Modules. J. Mol. Biol. 319, 1143–1156.

Brune, A. (2007). Woodworker's digest. Nature 450, 487–488.

Bryant, M.P., Small, N., Bouma, C., and Chu, H. (1957). Bacteroides ruminicola n. sp. and Succinimonas amylolytica; the new genus and species; species of succinic acid-producing anaerobic bacteria of the bovine rumen. J. Bacteriol. 15–23.

Cantarel, B.I., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. 37, 233–238.

Chakdar, H., Kumar, M., Pandiyan, K., Singh, A., Nanjappan, K., Kashyap, P.L., and Srivastava, A.K. (2016). Bacterial xylanases: biology to biotechnology. 3 Biotech 6, 1–15.

Charnock, S.J., Spurway, T.D., Xie, H., Beylott, M.H., Virden, R., Warren, R. a J., Hazlewood, G.P., and Gilbert, H.J. (1999). The topology of the substrate binding clefts of glycosyl hydrolase family 10 xylanases are not conserved. J. Biol. Chem. 273, 32187–32199.

Chassard, C., Goumy, V., Leclerc, M., Del'homme, C., and Bernalier-Donadille, A. (2007). Characterization of the xylan-degrading microbial community from human faeces. FEMS Microbiol. Ecol. 61, 121–131.

Collins, T., Gerday, C., and Feller, G. (2005). Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiol. Rev. 29, 3–23.

Comino, P., Collins, H., Lahnstein, J., Beahan, C., and Gidley, M.J. (2014). Characterisation of soluble and insoluble cell wall fractions from rye, wheat and hull-less barley endosperm flours. Food Hydrocoll. 41, 219–226.

Dehority, B.A. (1966). Characterization of Several Bovine Rumen Bacteria Isolated with a Xylan Medium. J. Bacteriol. 91, 1724–1729.

Despres, J., Forano, E., Lepercq, P., Comtet-Marre, S., Jubelin, G., Chambon, C., Yeoman, C.J., Berg Miller, M.E., Fields, C.J., Martens, E., et al. (2016). Xylan degradation by the human gut Bacteroides xylanisolvens XB1AT involves two distinct gene clusters that are linked at the transcriptional level. BMC Genomics 17, 326.

Din, N., Damude, H.G., Gilkes, N.R., Miller, R.C., Warren, R. a, and Kilburn, D.G. (1994). C1-Cx revisited: intramolecular synergism in a cellulase. Proc. Natl. Acad. Sci. U. S. A. 91, 11383–11387.

Dodd, D., Moon, Y.-H.H., Swaminathan, K., Mackie, R.I., and Cann, I.K.O. (2010). Transcriptomic analyses of xylan degradation by Prevotella bryantii and insights into energy acquisition by xylanolytic bacteroidetes. J. Biol. Chem. 285, 30261–30273.

Dodd, D., Mackie, R.I., and Cann, I.K.O. (2011). Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. Mol. Microbiol. 79, 292–304.

Fisher, S.Z., Von Schantz, L., Hakansson, M., Logan, D.T., and Ohlin, M. (2015). Neutron Crystallographic Studies Reveal Hydrogen Bond and Water-Mediated Interactions between a Carbohydrate-Binding Module and Its Bound Carbohydrate Ligand. Biochemistry 54, 6435–6438.

Flint, H.J., Whitehead, T.R., Martin, J.C., and Gasparic, A. (1997). Interrupted catalytic domain structures in xylanases from two distantly elated strains of Prevotella ruminicola. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1337, 161–165.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, (Totowa, NJ: Humana Press), pp. 571–607.

Gerlt, J.A., Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkhin, B., Slater, D.R., and Whalen, K.L. (2015). Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. Biochim. Biophys. Acta - Proteins Proteomics 1854, 1019–1037.

Gilbert, H.J. (2007). Cellulosomes: Microbial nanomachines that display plasticity in quaternary structure. Mol. Microbiol. 63, 1568–1576.

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P., and Breitsprecher, D. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. J. Mol. Struct. 1077, 101–113.

Khosla, C., and Harbury, P.B. (2001). Modular enzymes. Nature 409, 247–252.

Kormos, J., Johnson, P.E., Brun, E., Tomme, P., McIntosh, L.P., Haynes, C.A., and Kilburn, D.G. (2000). Binding Site Analysis of Cellulose Binding Domain CBD N1 from Endoglucanse C of Cellulomonas fimi by Site-Directed Mutagenesis. Biochemistry 39, 8844–8852.

Koropatkin, N.M., and Smith, T.J. (2010). SusG: A Unique Cell-Membrane-Associated alpha-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. Structure 18, 200–215.

Kudo, T. (2009). Termite-microbe symbiotic system and its efficient degradation of lignocellulose. Biosci. Biotechnol. Biochem. 73, 2561–2567.

Li, W., and Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, 490–495.

Martens, E.C., Koropatkin, N.M., Smith, T.J., and Gordon, J.I. (2009). Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. J. Biol. Chem. 284, 24673–24677.

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N., et al. (2011). Recognition and degradation of plant

cell wall polysaccharides by two human gut symbionts. PLoS Biol. 9.

Martens, E.C., Kelly, A.G., Tauzin, A.S., and Brumer, H. (2014). The devil lies in the details: How variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. J. Mol. Biol. 426, 3851–3865.

Martinez, T., Texier, H., Nahoum, V., Lafitte, C., Cioci, G., Heux, L., Dumas, B., O'Donohue, M., Gaulin, E., and Dumon, C. (2015). Probing the functions of carbohydrate binding modules in the cbel protein from the oomycete phytophthora parasitica. PLoS One 10, 1–14.

Matsui, I., Ishikawa, K., Matsui, E., Miyairi, S., Fukui, S., and Honda, K. (1991). Subsite structure of Saccharomycopsis alpha-amylase secreted from Saccharomyces cerevisiae. J Biochem 109, 566–569.

McNulty, N.P., Wu, M., Erickson, A.R., Pan, C., Erickson, B.K., Martens, E.C., Pudlo, N.A., Muegge, B.D., Henrissat, B., Hettich, R.L., et al. (2013). Effects of Diet on Resource Utilization by a Model Human Gut Microbiota Containing Bacteroides cellulosilyticus WH2, a Symbiont with an Extensive Glycobiome. PLoS Biol. 11, e1001637.

Merino, S.T., and Cherry, J. (2007). Progress and challenges in enzyme development for biomass utilization. Adv. Biochem. Eng. Biotechnol. 108, 95–120.

Miller, G.L. (1959). Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal. Chem. 31, 426–428.

Miyazaki, K., Hirase, T., Kojima, Y., and Flint, H.J. (2005). Medium- to large-sized xylo-oligosaccharides are responsible for xylanase induction in Prevotella bryantii B14. Microbiology 151, 4121–4125.

Montanier, C., van Bueren, A.L., Dumon, C., Flint, J.E., Correia, M. a, Prates, J. a, Firbank, S.J., Lewis, R.J., Grondin, G.G., Ghinet, M.G., et al. (2009). Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. Proc. Natl. Acad. Sci. U. S. A. 106, 3065–3070.

Ohkuma, M. (2003). Termite symbiotic systems: efficient bio-recycling of lignocellulose. Appl. Microbiol. Biotechnol. 61, 1–9.

Otani, S., Mikaelyan, A., Nobre, T., Hansen, L.H., Koné, N.A., Sørensen, S.J., Aanen, D.K., Boomsma, J.J., Brune, A., and Poulsen, M. (2014). Identifying the core microbial community in the gut of fungus-growing termites. Mol. Ecol. 23, 4631–4644.

Pell, G., Taylor, E.J., M.Gloster, T., Turkenburg, J.P., M.G.A.Fontes, C., M.A.Ferreira, L., Nagy, T., J.Clark, S., J.Davies, G., J.Gilbert, H., et al. (2004a). The Mechanisms by Which Family 10 Glycoside Hydrolases Bind Decorated Substrates. J. Biol. Chem. 279, 9597–9605.

Pell, G., Szabo, L., Charnock, S.J., Xie, H., Gloster, T.M., Davies, G.J., and Gilbert, H.J. (2004b). Structural and biochemical analysis of Cellvibrio japonicus Xylanase 10C: How variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases. J. Biol. Chem. 279, 11777–11788.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42, 320–324.

Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Baslé, A., Morland, C., Day, A.M., Zheng, H., et al. (2015). Glycan complexity dictates microbial resource allocation in the large intestine. Nat. Commun. 6, 1–15.

Von Schantz, L., Hkansson, M., Logan, D.T., Walse, B., Österlin, J., Nordberg-Karlsson, E., and Ohlin, M. (2012). Structural basis for carbohydrate-binding specificity-A comparative assessment of two engineered carbohydrate-binding modules. Glycobiology 22, 948–961.

Scharf, M.E., and Tartar, A. (2012). Termite digestomes as sources for novel lignocellulases. Biofuels, Bioprod. Biorefining 6, 246–256.

Scheller, H.V., and Ulvskov, P. (2010). Hemicelluloses. Annu. Rev. Plant Biol. 61, 263–289.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A software Environment for integrated models of biomolecular interaction networks. Am. Assoc. Cancer Res. Educ. B. 13, 2498–2504.

Shipman, J. a, Cho, K.H., Siegel, H. a, and Salyers, A. a (1999). Physiological Characterization of SusG , an Outer Membrane Protein Essential for Starch Utilization by Bacteroides thetaiotaomicron. J. Bacteriol. 181, 7206–7211.

Song, L., Dumon, C., Siguier, B.B., André, I., Eneyskaya, E., Kulminskaya, A., Bozonnet, S., and O'Donohue, M.J. (2014). Impact of an N-terminal extension on the stability and activity of the GH11 xylanase from Thermobacillus xylanilyticus. J. Biotechnol. 174, 64–72.

Stephen, A.M. (1983). Other Plant Polysaccharides. In The Polysaccharides, VOL.2, pp. 97–193.

Su, L., Yang, L., Huang, S., Su, X., Li, Y., Wang, F., Wang, E., Kang, N., Xu, J., and Song, A. (2016). Comparative Gut Microbiomes of Four Species Representing the Higher and the Lower Termites. J. Insect Sci. 16, 97.

Takeo, K. (1984). Affinity electrophoresis: Principles and applications. Electrophoresis 5,

187–195.

Vincentelli, R., and Romier, C. (2013). Expression in Escherichia coli: Becoming faster and more complex. Curr. Opin. Struct. Biol. 23, 326–334.

Wang, K., Pereira, G. V., Cavalcante, J.J. V., Zhang, M., Mackie, R., and Cann, I. (2016). Bacteroides intestinalis DSM 17393, a member of the human colonic microbiome, upregulates multiple endoxylanases during growth on xylan. Sci. Rep. 6, 1–11.

Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.C., Djordjevic, G., Aboushadi, N., et al. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450, 560–565.

Wu, H., Montanier, C.Y., and Dumon, C. (2017). Quantifying CBM Carbohydrate Interactions Using Microscale Thermophoresis. In Methods in Molecular Biology, (Humana Press, New York, NY), pp. 129–141.

Zhang, M., Chekan, J.R., Dodd, D., Hong, P.-Y.P.-Y.P.-Y.P.-Y., Radlinski, L., Revindran, V., Nair, S.K., Mackie, R.I., and Cann, I. (2014). Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. Proc. Natl. Acad. Sci. U. S. A. 111, E3708-17.

FOOTNOTES:

The abbreviations used are : AGE, affinity gel electrophoresis; BSA, bovine serum albumin; CBM, carbohydrate binding module; D-MeGlcAp, D-4-O-methyl glucuronyl; D.P, degree of polymerization; D-Xyl, D- xylopyranosyl; GH, glycoside hydrolyase; GAX, glucuronoarabinoxylan; GX, glucuronoxylan; L-Araf, L -arabinofuranosyl; PUL, polysaccharide utilization loci; MST, microscale thermophoresis; SD test, SDS-denaturation test; SSN, sequence similarity network; XOS, xylooligosaccharide.

SUPPLEMENTARY MATERIAL:

Materials and Methods:

Enzyme specificity of Pm25 and mutants thereof. To compare the specific activity of Pm25 and mutants thereof, RAX (0.75%, w/v), LVWAX (0.4%, w/v) and beechwood GX (1%, w/v) were employed. The final concentration of Pm25 used to degrade beechwood GX and other soluble substrates is 35 and 7 mg/L, respectively. Owing to the weaker activity of the mutants (M1 and 2), for activity assays the final enzyme concentration was increased 100-fold compared to assays using the wild type Pm25.

Table S1. Primes used in this study.

	Target	Orientation	Sequence (5' → 3')
primers for site-directed mutagenesis	*Pm25*|E546A	F	GTGGGATGTGCTGAACGCGCCCATGAACGAGAAC
		R	GTTCTCGTTCATGGGCGCGTTCAGCACATCCCAC
	Pm25|E663tga	F	GGCAAACTCATCAAGGTGACGTGACTCGATATTGCCATATCCAC
		R	GTGGATATGGCAATATCGAGTCACGTCACCTTGATGAGTTTGCC
	Pm25|tga663A	F	GCAAACTCATCAAGGTGACGGCGCTCGATATTGCCATATCC
		R	GGATATGGCAATATCGAGCGCCGTCACCTTGATGAGTTTGC
	Pm25|H634A	F	GCATCGGCACGCAAATGGCGCTCAACCTCAACTGG
		R	CCAGTTGAGGTTGAGCGCCATTTGCGTGCCGATGC
	CBM4-1|Y213A	F	GGCTCGACCGGAAAAGCGGACGCACAACTCAACTCTCC
		R	GGAGAGTTGAGTTGTGCGTCCGCTTTTCCGGTCGAGCC
	CBM4-1|Q216A	F	CGGAAAATACGACGCAGCGCTCAACTCTCCGGCG
		R	CGCCGGAGAGTTGAGCGCTGCGTCGTATTTTCCG
	CBM4-1|N218A	F	TACGACGCACAACTCGCGTCTCCGGCGTTCCCG
		R	CGGGAACGCCGGAGACGCGAGTTGTGCGTCGTA
	CBM4-1|Y257A	F	CGGAACTGGGTAATCAAGCGCCATGGGTGAACGGC
		R	GCCGTTCACCCATGGCGCTTGATTACCCAGTTCCG
	CBM4-1|W259A	F	GGTAATCAATATCCAGCCGTGAACGGCGCAGAACTGGC
		R	GCCAGTTCTGCGCCGTTCACGGCTGGATATTGATTACC
	CBM4-1|N261A	F	CAATATCCATGGGTGGCGGGCGCAGAACTGGCCG
		R	CGGCCAGTTCTGCGCCCGCCACCCATGGATATTG
	CBM4-2|Y378A	F	GGTGCGGCTGGCAAAGCGGACCGCCAACTCAAC
		R	GTTGAGTTGGCGGTCCGCTTTGCCAGCCGCACC
	CBM4-2|Q381A	F	GGCAAATACGACCGCGCGCTCAACAGCCCGGCATTTCC
		R	GGAAATGCCGGGCTGTTGAGCGCGCGGTCGTATTTGCC
	CBM4-2|N383A	F	GCAAATACGACCGCCAACTCGCGAGCCCGGCATTTCCG
		R	CGGAAATGCCGGGCTCGCGAGTTGGCGGTCGTATTTGC
	CBM4-2|Y422A	F	GAACTCGGCAACCAGGCGCCGTGGGGAAGCGGC
		R	GCCGCTTCCCCACGGCGCCTGGTTGCCGAGTTC
	CBM4-2|W424A	F	GGCAACCAGTATCCGGCCGGAAGCGGCGGCGAG
		R	CTCGCCGCCGCTTCCGGCCGGATACTGGTTGCC
primers for in-fusion clone	*Pm25*_CBM4-1_L167-L320	F	CAGCCATATGGCTAGCTTGTGGACGGAAGTTTCGAAG
		R	GGTGGTGGTGCTCGATTAGAGGTCTGTGATCTCGACG
	*Pm25*_CBM4-2_D332-D484	F	CAGCCATATGGCTAGCCTTGTGGACGGCAATTTCGAAGGCGGAG
		R	GGTGGTGGTGCTCGAGTTATTAGAGGTCGGTGACCTCAACGGCGTC
	*Pm25*_CBM4-1+2_L167-D484	F	GACGACGACGACAAGCTTGTGGACGGAAGTTTCGAAGAGGGGATG
		R	GTGGTGGTGCTCGAGGTCGGTGACCTCAACGGCGTCGATC
	PET32a_linearization	F	CTCGAGCACCACCACCACCACCAC
		R	CTTGTCGTCGTCGTCGCCAGAACCAGAACCGGCCAGGTTAG

Table S2. The concentration series of RAX used for different constructs.

RAX‰ (w/v)	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
M8	0.9	0.6	0.3	0.15	0.105	0.06	0.03					
M8_Y213A	2.1	1.5	1.2	0.9	0.6	0.3	0.15	0.105	0.06	0.03		
M8_Q216A	1.5	1.2	0.9	0.6	0.45	0.3	0.15	0.105	0.06	0.03		
M8_N218A	0.9	0.6	0.3	0.15	0.105	0.06	0.03					
M8_Y257A	2.1	1.5	1.2	0.9	0.6	0.3	0.15	0.105	0.06	0.03		
M8_W259A	1.5	1.2	0.9	0.6	0.3	0.15	0.105	0.06	0.03			
M8_N261A	0.9	0.6	0.3	0.15	0.105	0.06	0.03					
M9	1.5	1.2	0.9	0.6	0.3	0.15	0.126	0.105	0.06	0.03		
M9_Y378A	3	2.1	1.5	1.2	0.9	0.6	0.45	0.3	0.15	0.105	0.06	
M9_Q381A	1.5	1.2	0.9	0.6	0.3	0.15	0.105	0.06				
M9_N383A	1.5	1.2	0.9	0.6	0.3	0.15	0.126	0.105	0.09	0.06		
M9_Y422A	2.1	1.5	1.2	0.9	0.6	0.45	0.3	0.15	0.105	0.06		
M9_W424A	1.5	1.2	0.9	0.6	0.3	0.15	0.105	0.06				
M11	0.45	0.3	0.15	0.126	0.105	0.09	0.06	0.045	0.03	0.0225	0.015	0.01
M1	0.45	0.3	0.15	0.126	0.105	0.09	0.06	0.045	0.03	0.0225	0.015	0.01
M8	0.45	0.3	0.15	0.126	0.105	0.0225	0.015					

Table S3. Relative activity (%) of Pm25 and mutants thereof towards polysaccharides.

Substrate	*Pm*25	M 1	M 2
RAX	100 ± 0.89	0.19 ± 0	0.24 ± 0
LVWAX	100 ± 6.57	1.08 ± 0.03	0.76 ± 0.02
Beechwood GX	100 ± 4.67	4.50 ± 0.15	3.08 ± 0.13

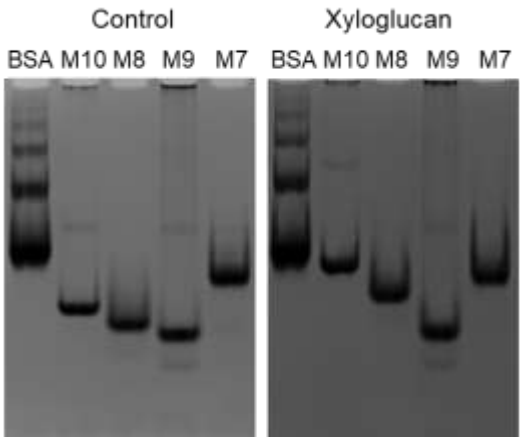


Figure S1. Affinity gel electrophoresis of CBMs in *Pm*25 towards 0.5% (w/v) xyloglucan.

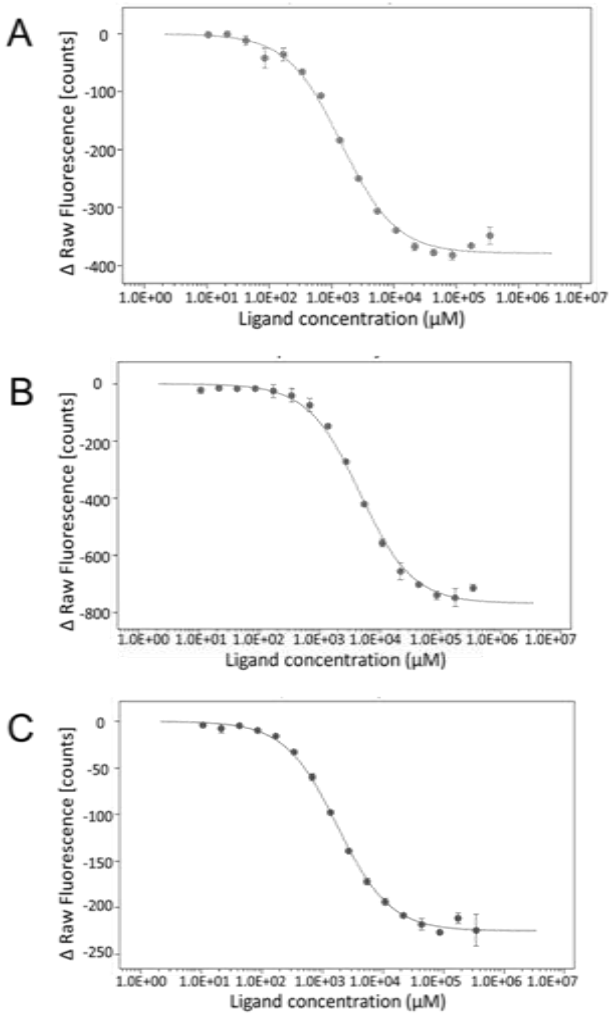


Figure S2. Measurement of Kd based on ligand induced effect. (A) The inactive catalytic domain, M7 with X_6. (B) CBM4-1, M8 with X_6. (C) CBM4-2, M9 with X_6.

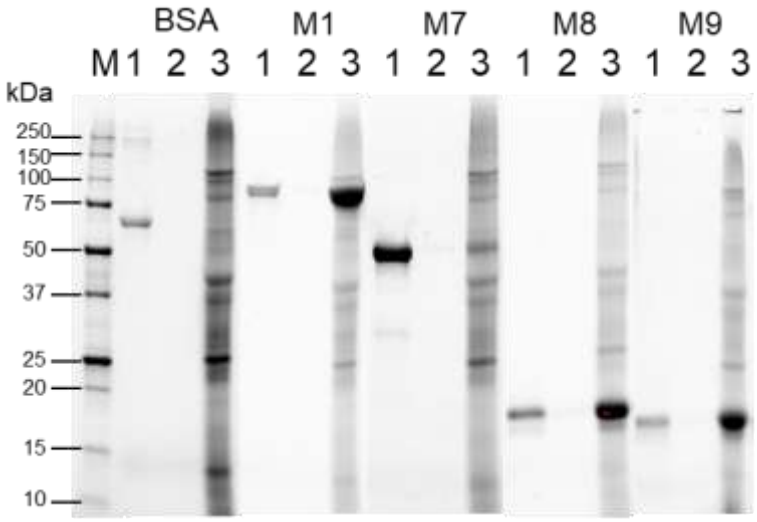


Figure S3. Binding of inactive Pm25 and its truncated derivatives to wheat bran. BSA was used as negative control. Lane M, molecular mass markers. Lane 1, unbound fraction; lane 2, the third wash of the pellet; lane 3, bound fraction.

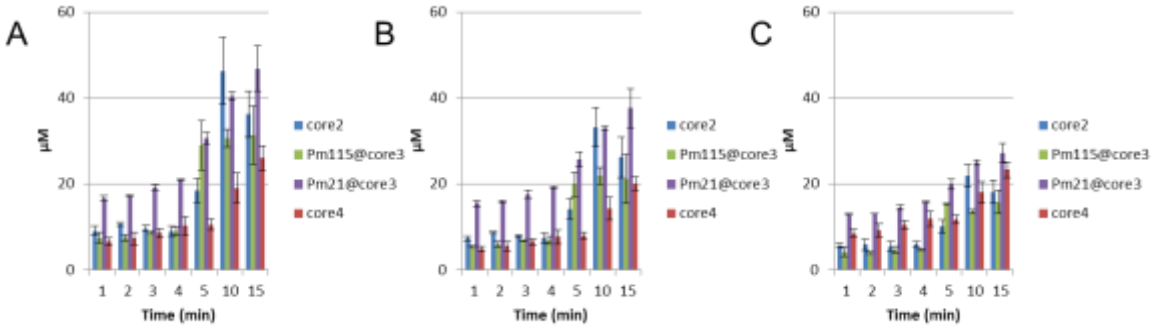


Figure S4. The time-dependent degradation of LVWAX by (A) M3 and (B) M4.

ARTICLE II

Study of the role of CBMs in the context of PUL enzymes

ABSTRACT:

CBMs can be appended to GHs, GTs, toxins and so forth and different synergistic effects were observed in different combinations and orientations of CBM and catalytic domains. In many case CBM didn't bring additive effect on the catalytic efficiency towards soluble substrates. However, these studies focused on either specific activity or xylo-oligosaccharide (XOS, DP<7), which is not sufficient to completely deny the contribution of CBMs. Based on an usual multi-domain xylanases, *Pm*25, with one catalytic domain (GH10) and two CBM4 intercalated, we compared the product profile from D-xylose to XOS (DP=28) of wild type *Pm*25 and the CBM4-abrogated mutants with/without other xylan-active enzymes from the same putative PUL. The results showed that GH10 and GH11 from the PUL are the main xylan active enzymes, small or negligible effects and support the negligible contribution of CBMs in terms of soluble substrates under the test conditions.

INTRODUCTION:

The biodegradation of plant cell walls is of considerable interest to scientists and technologists alike. This is because plant cell walls constitute a highly abundant reservoir of renewable carbon in the form of lignocellulose. This composite carbon-rich material is principally composed of cellulose, lignin and heteroxylans, these latter constituting the most widespread and abundant form of hemicellulose.

Xylanases are carbohydrate-active enzymes that hydrolyze heteroxylans, bringing about their depolymerization. According to the classification of carbohydrate-active enzymes (the CAZy database, Lombard et al., 2014), the two archetypal xylanase families are GH10 and GH11, families that are composed of both bacterial and fungal origins whereas a majority of GH10s are from bacteria. Typically, GH10 xylanases display high molecular mass, low pI, multiple functions and accessory domains (40%), whilst GH11 members generally display lower molecular mass, higher pI, a unique function and are less likely to be associated with other domains (20%) (NR.Gilkes et al., 1991; Sunna and Antranikian, 1997). Similar to enzymes within families GH1, 2, and 5, the catalytic domain of GH10 xylanases display a TIM-barrel fold. This is in contrast to GH11 xylanases, which are characterized by a β-jelly roll structure. Nevertheless, despite these differences, both GH10 and GH11 xylanases perform catalysis via a retaining mechanism that requires a catalytic dyad composed of a catalytic nucleophile and a proton donor (Gebler et al., 1992; Withers et al., 1986).

Regarding xylooligosaccharide (XOS), GH10 generally display good aptitude for the hydrolysis of xylotriose i.e. DP 3 (Pell et al., 2004). In contrast, members of family GH11 are generally much less active on such XOS, requiring a minimal size of DP4 (xylotetraose) (Pollet et al., 2010). When acting on branched heteroxylans, substrate recognition and binding by members of GH10 requires the presence of one consecutive unsubstituted xylopyranosyl residues at substite -1 and at substite -2, it can tolerate the O-3 linked L-arabinofuranoside (Biely et al., 1997; Kolenová et al., 2006; Pell et al., 2004). In contrast, GH11 family members require three consecutive unsubstituted xylopyranosyl residues, which are accommodated in subsites -2 to +1 (Biely et al., 1997; Katapodis et al., 2003). Consequently, GH10 xylanase can produce smaller hydrolysis products from glucuronoxylan and arabinoxylan than GH11 xylanases (Pollet et al., 2010), and generally performed better in biomass deconstruction than GH11 xylanase (Hu and Saddler, 2018). In addition, GH10 xylanase is more favorable to associate with carbohydrate binding module (CBM) constituting

multi-domain proteins compared to GH11 xylanase, with 18% and 3% respectively according to Pfam database (http://pfam.xfam.org/). CBM has been recognized as one of the helper agents enhancing performance of lignocellulases although it is devoid of hydrolytic properties (Kim et al., 2014).

In order to completely degrade xylan, xylanases operate in cooperation with other so-called accessory enzymes. Prominent among these are α-L-arabinofuranosidases (e.g. from families GH43 and GH51), α-D-glucuronidases (e.g. from families GH67 and GH115), acetylxylan esterases , feruloyl esterases (e.g. from family CE1). All of these enzymes remove main-chain decorations, while β-D-xylosidases act on XOS released by xylanases, converting these into xylose. The concerted action between xylanase and accessory enzymes is characteristic of xylanolytic systems and thus such enzymes are widespread in fungi and bacteria (Coughlan and Hazlewood, 1993; Filho et al., 1991; Sunna et al., 1997; Weaver et al., 1992).

Bioprospection within microbial xylanolytic systems can be conveniently performed using functional metagenomics, a technique that can be used to probe microbiologically-complex environmental samples without the need for prior cultivation of individual species (Bastien et al., 2013; Tasse et al., 2010). Using this type of approach it is possible to simultaneously pinpoint individual genes encoding xylanolytic enzyme components and probe their local genomic context. This provides a means to discover functional gene clusters, such as polysaccharide utilization loci which are characteristic of the bacterial phylum *Bacteroidetes* (Grondin et al., 2017).

The extraction of monomeric sugars (saccharification) from lignocellulosic biomass is a fundamental step in the production of 2nd generation biofuels. To achieve this step, a series of unit operations are deployed, which include the hydrolysis of extracted polysaccharides using highly complex enzyme cocktails. These contain a wide range of cellulose-active enzymes and many accessory enzymes, including xylanases. The latter are beneficial because it has been shown that the presence of xylanases and other xylanolytic enzymes not only leads to the release of xylose, but also enhances cellulase action and thus accentuates glucose production, while reducing enzyme loadings (Dondelinger et al., 2016; Gao et al., 2011; Tabka et al., 2006). For this reason, the search for efficient xylanolytic enzymes and cocktails containing both main-chain cleaving (MCEs) and side-chain cleaving enzymes (SCEs) has been intensive (Kormelink and Voragen, 1993; Sørensen et al., 2007; Suh et al., 1996; Vardakou et al., 2004; de Vries et al., 2000).

Within xylanolytic cocktails the different enzymes interact in a variety of ways. When MCEs or SCEs act in synergy between themselves, this can be described as homosynergy, whereas the concerted action of both MCEs and SCEs is referred to as heterosynergy. Sometimes, MCEs require the prior action of debranching enzymes (Khandke et al., 1989), in other cases MCE action is required as a first step to release branched oligosaccharides that constitute preferential targets for some SCEs (Mackenzie and Bilous, 1988; Poutanen and Sundberg, 1988; Puls et al., 1987). In this way, the sequential and simultaneous action of multiple enzymes can favorably modulate the catalytic efficiency of individual enzyme components (Poutanen et al., 1991). On the other hand, the concomitant action of cocktails components can also lead to negative effects, which are known as antisynergy. This last phenomenon can occur when the action of one enzyme prevents or hampers the action of another (Coughlan et al., 1993). Finally, considering single enzymes, intramolecular synergy is achieved through the functional interplay between catalytic domains and associated CBMs (Boraston et al., 2004; Din et al., 1994). Moreover, CBMs could also be involved in synergism between exo-lytic enzymes, so called exo-exo synergy in cellulolytic enzymes (Badino et al., 2017).

Enzymes encoded by PUL systems are interesting, because they form part of finely-tuned catabolic machinery in which strong enzyme synergy provides the basis for efficient polysaccharide catabolism (Arnal et al., 2014; Bågenholm et al., 2016; Wang et al., 2016a). Therefore, PULs might be considered as interesting paradigms for the creation of enzyme cocktails, especially since it has been suggested that the synergy between PUL enzyme components is greater than that between enzymes from different species (Gilbert et al., 1993; Tauzin et al., 2016).

This study focuses on a putative PUL that contains two GH10 family members (*Pm*12 and *Pm*25), one GH11 (*Pm*22), one GH115 (*Pm*115) and one GH43 (*Pm*21) (Bastien et al., 2013) (Fig1). The *Pm*115 is a putative α-L-glucuronidase that removes the methylated glucuronic acids in side chains. The *Pm*21 could have both β-xylosidase and α-L-arabinofuranosidase, as its amino acid sequence shares 62% identity with CoXyl43, a β-xylosidase/α-arabinofuranosidase isolated from a compost metagenome (Matsuzawa et al., 2017). In addition, Xus C and Xus D are located upstream in the cluster, and one hypothetical transporter protein is between *Pm*22 and *Pm*12. Among the three xylanases, *Pm*22 and *Pm*25 are multi-domain xylanases. The C-terminal sequence of *Pm*22 is annotated as GH11 (*Pm*24) while N-terminal sequence is poorly annotated and currently unknown. *Pm*25 is a Xyn10C-

like enzyme carrying two CBM4s, which corresponds to a ubiquitous PUL feature that has attracted a lot of attention recently and has been the subject of several studies (Despres et al., 2016; Dodd et al., 2011; Flint et al., 1997; Rogowski et al., 2015; Zhang et al., 2014; Wu, in preparation). It is believed that Xyn10C-like proteins, which contain CBMs, provide early stage hydrolysis of polysaccharides, producing oligosaccharides that are then withheld by the CBMs thus avoiding diffusion. Subsequently, the products of Xyn10C-like proteins are further processed and imported into the bacterial cell by other PUL components (White et al., 2014). In this study we have investigated the role of the *Pm*25 CBM, notably with the aim to understand how it modulates the release of short-to-medium sized oligosaccharides (DP 1 to 28) when *Pm*25 operates in synergy with the other PUL enzymes.

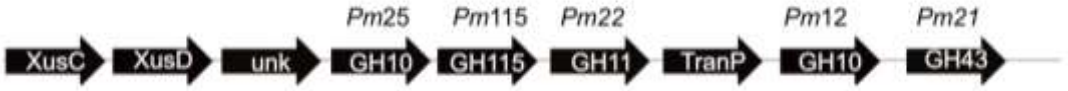


Fig 1. Presence of xylan active enzymes in the putative PUL. *Pm*25, *Pm*22 and *Pm*12 are endo-acting on the main chain of xylan; *Pm*115 can remove the glucuronic acid decorations, and *Pm*21 can liberate xylose from xylobiose as well as arabinofuranose from side chain of xylan.

MATERIAL AND METHODS:

Protein expression and purification.

The amino acid sequence of *Pm*12, *Pm* 22 and *Pm* 21 were rooted from its original sequences with the GenBank accession number CCO21040.1, CCO21038.1, and CCO21041.1, respectively. *Pm* 12 and *Pm*22 were cloned after removal of signal peptide predicted by SignalP4.1. Due to unsuccessful expression of *Pm*22, *Pm* 24, the truncation version of the GH11 module (residues 637-841) was also constructed. All three sequences were cloned into pDEST17 as described elsewhere (G. Arnal et al. manuscript in preparation, Vincentelli and Romier 2013). *Pm*115 (CCO21037.1) was cloned without the signal peptide into pET28 using the forward primers 5'-CAGCCATATGGCTAGCCAGTCCCGGGAGAGTAGA TTC-3', and reverse primer 5'-GGTGGTGGTGCTCGAGTCATTTGTCAATGGTTTTAC-3'. The recombinant proteins

harboring His_6 tags at their N-terminal extremities, were expressed in RosettaDE3 with auto-inducing medium, except for *Pm* 115 which was expressed within TunerDE3 in LB (Luria-Bertani) medium induced by IPTG (isopropyl-β-D-l-thiogalactopyranoside, 0.2 mM final). Cells were collected by centrifugation at 7,000 × *g* for 20 min and disrupted by sonication. The purification step was done by applying cell lysates to Talon metal affinity resin (Clontech, USA) following the manufacturer's instructions. All purified proteins were conserved at 4°C in 50 mM phosphate buffer (pH 7.0) after dialysis. The expression and purification of *Pm* 25, *Pm*25Y213AY378A and *Pm* 25ΔCBMs were as described before (H. Wu, to be published).

Biochemical characterization of PUL enzymes

Beechwood glucuronoxylan (BGX) and wheat arabinoxylan (WAX, low viscosity) were purchased from Megazyme International (Bray, Ireland), while and aldopentauronic acid (UXXR and XUXXR) were kindly provided by Magazyme. Xylanase activity was estimated by quantifying the reducing sugars released over time using a dinitrosalicylic acid (DNS)-based assay (Miller, 1959). Working in triplicate, reactions were initiated by the addition of 90 µl of enzyme (*Pm*12 and *Pm*24, 0.2 µM) into 810 µl of BGX (20 mg/ml) in appropriate buffer containing bovine serum albumin (BSA; 1 mg/ml) and were incubated for 18 min at 37 °C. 100 µl of reaction was removed every 3 min and boiled with 100 µl of DNS for 10 min at 95°C, and absorbance was measured at 540 nm in a plate reader after adding 1 ml of MilliQ-water (Arnal et al., 2014). The amount of reducing sugar liberated was determined via a standard curve using D-xylose (X1).

The apparent optimal pH for activity of the *Pm*12, *Pm*24 was determined by measuring the enzyme activity on BGX in the pH range from 4.0 to 10.0. Buffers used were, Phosphate-citrate buffer pH range 4.0 to 7.0, citric acid (50 mM)-Na_2HPO_4 (100 mM) (pH according to *buffer reference center*, Sigma-Aldrich protocol), trizma buffer (50 mM, pH 7.0 to 9.0), glycine-NaOH buffer (50 mM, pH 9.0 to 10.0). The optimum temperature was determined by incubating enzymes at optimum pH over a temperature range of 30-65 ºC. Thermal stability was evaluated by incubating the enzyme solution at a certain temperature for different durations (0, 0.5, 1, 4, 8, and 12 h) before the residual activity was assayed at the same temperature. One unit (IU) was defined as the quantity of enzyme in mg required to liberate 1 µmol of D-xylose equivalent per minute under optimum condition.

Enzyme kinetics parameters were calculated using BGX (0.4 to 20 mg/ml) and WAX (0.5 to 16 mg/ml) as substrates at the optimum pH and temperature. Apparent K_m and V_{max} values were determined by fitting Michaelis-Menten curves to one-site saturation under ligand-binding equation using Sigma Plot software version 11.0.

HPAEC-PAD and MALDI-TOF analysis of hydrolyzed products

BGX was chosen for this analysis as it provides the substrate for *Pm*115. Hydrolysis of BGX was performed by incubating 1350 µl of 2% (w/v) BGX (citric acid 50 mM, dibasic sodium phosphate 100 mM, 1 mg/ml BSA) with 150 µl of enzyme mix in an Eppendorf™ Thermomixer™ R at 1400rpm, 37 °C. The enzyme amounts were set to make sure only around 10% of substrates is consumed in the first 15 min. Enzyme doses in the reaction were as follows, xylanase (*Pm*12, *Pm*24, *Pm*25, *Pm*25Y213AY378A and *Pm*25ΔCBMs), 2 nM; xylosidase (Pm21), 200 nM; α-Glucuronidase (*Pm*115), 10 nM. The enzyme combinations were made with "core enzymes" mixing with *Pm*25 and mutants thereof. There are four different core enzyme combinations, *Pm*12 and *Pm*24 (**core2**); core2 + *Pm*115 (**Pm115@core3**), core2+ *Pm*21 (**Pm21@core3**) and the four enzymes mentioned above *Pm*12, *Pm*24, *Pm*21+ *Pm*115 (**core4**). Reactions were conducted for 15 min, and after each time point (1, 2, 3, 4, 5, 10, 15 min), 190 µl of reaction was taken out and heated at 95 °C for 10 min to stop the reaction. The samples were ready to be analyzed after centrifuge at maximum speed for 10 min.

The released X1 and XOS were quantified by HPAEC-PAD on a Dionex PA1 column equipped with Carbo-Pac PA-1 guard and analytical columns (4*50 mm and 4*250 mm, respectively). Separation of oligosaccharides was achieved by isocratic elution with 100 mM NaOH at a flow rate of 1 ml/min from 0 to 5 min, a gradient of 0 to 120 mM sodium acetate in 100 mM NaOH from 5 min to 25 min, and isocratic elution with 500 mM sodium acetate in 100 mM NaOH from 25 min to 35 min. Then the column is re-equilibrated with 100 mM NaOH for another 10 min. Calibration was achieved using X1 and XOS (X2, X3, X4, X5 and X6) at concentrations from 5 to 100 µM.

For the analysis of xylan hydrolysis products by MALDI-TOF MS, 1 µl of hydrolysates was mixed with 1µl of matrix solution (10 mg/ml 2, 5-dihydroxybenzoic acid dissolved in acetonitrile-water [1:1, v/v], and 0.1 % [w/v] trifluoroacetic acid). One microliter of the mixture was mixed with matrix (DMA/DHB) and spotted onto the MALDI-TOF MS

plate and allowed to dry before the analysis. Positive mass spectra were collected with a AUTOFLEX Speed instrument (linear mode, laser 1000Hz).

RESULTS AND DISCUSSION:

Biochemical characterization of PUL enzymes

Table 1. Biochemical properties and kinetic parameters of *Pm*12, *Pm*24 and *Pm*25.

Enzyme		*Pm*12	*Pm*24	*Pm*25[a]
Final yield of purified protein (mg/L)		126	12.8	34.9
Theoretical pI		5.16	5.43	4.34
Mol mass (kDa)		43.23	26.57	83.97
Optimal pH[b]		6.5-7.5 (6.5)	5.0-9.0 (6.0)	4.5-9.0 (7.5)
Optimal T°C[b]		45-60 (50)	60-65 (40)	60-65 (50)
K_m^{app}(g/L)	**BGX**	1.2 ± 0.2	1.5 ± 0.2	3.1 ± 0.2
	WAX	3.0 ± 0.6	3.2 ± 0.4	2.0 ± 0.2
K_{cat}(/s)	**BGX**	296.2 ± 11.8	179.7 ± 7.6	10.3 ± 0.2
	WAX	608.1 ± 43.4	799.0 ± 29.0	36.2 ± 1.1

[a] The biochemical properties of *Pm*25 have been reported elsewhere [article 1]

[b] Values indicate that the relative enzyme activity was > 80% of the activity observed at optimum conditions. Values in parentheses are the optimal conditions defined for biochemical characterization.

All the glycoside hydrolases from the PUL were produced and purified to homogeneity. The physiochemical parameters for *Pm*12 and *Pm*24 were determined (Table 1). The pH optimum was different for each enzyme, with *Pm*25 showing the widest pH optimum range and *Pm*12 the narrowest. It is noteworthy that the activity of *Pm*24 is also stable over a wide range, from pH 5.0 to 9.0 (Table 1). Regarding temperature, all three enzymes achived highest activities in the range 60-65°C. However, enzyme stabilities, measured over a 24h-period were lower, at

40°C in the case of *Pm*24 and 50°C for *Pm*12 and *Pm*25. Overall the study of individual physicochemical parameters provided the means to define optimal conditions (pH 6.5-7.5 and 40°C) in which the three xylanases could operate together, without unduly compromising their activity. The values $K_{m\ app}$ on BGX and WAX were similar for *Pm*12 and *Pm*24 respectively, although a small difference was observed in the case of *Pm*25. Regarding K_{cat}, *Pm*12 and *Pm*24 were much more active than *Pm*25. This is not surprising considering that *Pm*25, like other Xyn10C-type enzyme, is a rather low active xylanase (Rogowski et al., 2015; Wang et al., 2016a). The activities of purified *Pm*21 and *Pm*115 were assessed using different substrates, revealing that *Pm*21 is more active against *p*NP-xylopyranose than *p*NP-arabinofuranose, which is consistent with results obtained for a homolog, CoXyn43 (Matsuzawa et al., 2017). Moreover, it was shown that *Pm*21 displays a Tm of 45.6 °C and that its optimum activity on *p*NP-xylopyranose occurs between pH 6.5 and 7.0 (data not shown). *Pm*115 was partly characterized, revealing for example that it displays a Tm of 39.5 °C and it is active on both both U^{4m2}XXR and XU^{4m2}XXR (Fig S1), consistent with its putative description as a glucuronidase (Wang et al., 2016b).

Comparison of the product profile of Pm25 and its mutants

In order to study the intramolecular synergy between the CBM4s and GH10 in *Pm*25, the product profile of the wild type and mutant *Pm*25 against BGX were compared. After 30 min of reaction, wild type *Pm*25 released 1.5-fold less X3 compared to the CBM-deleted constructs (Fig 2A). Nevertheless, after 14h the total release of D-xylose and XOS was the same for both wild type and mutated *Pm*25 (Fig 2B), which agrees with the results in article I except using another soluble substrate, WAX.

In order to characterize the effects of the above enzymes on medium-sized XOS, mass spectroscopy analysis was performed. Oligosaccharides from DP 5 to 28 were monitored and grouped by size range (Fig 2C, D and E). The intact enzyme released 1.5-fold less DP 5-8 than the CBM-truncated construct, while the double mutant (lacking both CBMs) produced an intermediate result (Fig 2C, D and E).

Taking into account the similarity of these results, we suggest that interplay between the catalytic domain and the CBMs in *Pm*25 is not an advantage in xylan depolymerization, and may even somehow hinder the hydrolysis at least in the beginning of the reaction (i.e. antisynergy). Similar behavior were found for another CBM-insertional enzyme, SusG (Koropatkin and Smith, 2010).

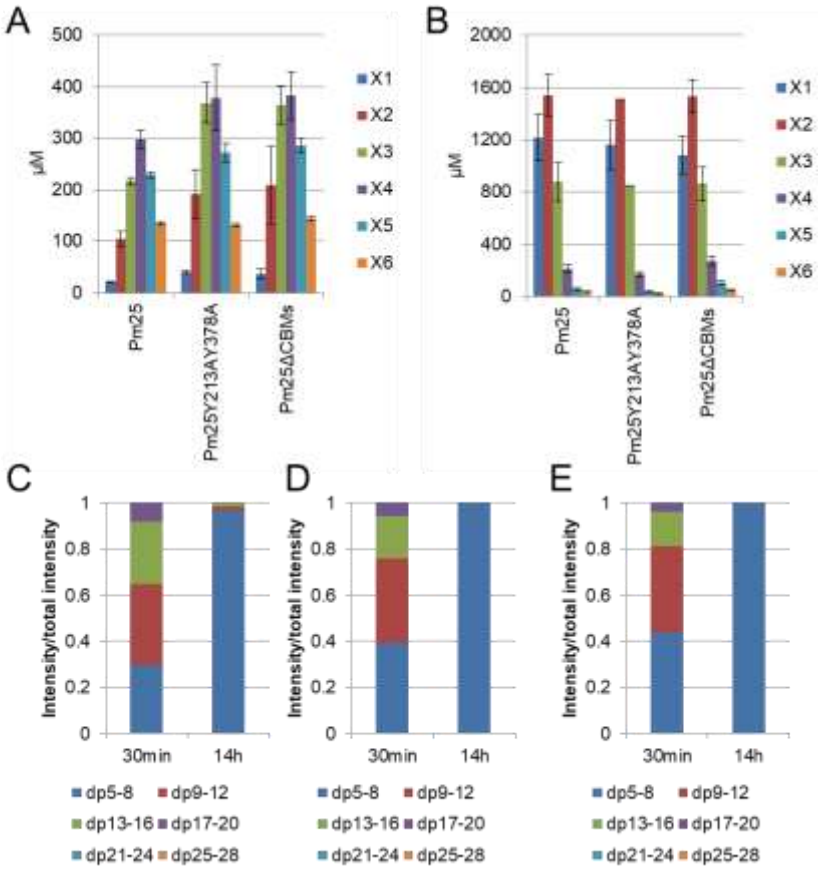


Fig 2. The product (DP from 1 to 28) profile of by *Pm*25 and its mutants. The amounts of xylose and XOS in products were quantified by HPAEC-PAD after reactions of 30 min (A) and 14h (B). The variety of medium-sized products after treatment of *Pm*25, *Pm*25Y213AY378A and *Pm*25ΔCBMs were shown in (C), (D) and (E) respectively.

Study the cooperation of MCE (Pm12 and Pm24) and SCE (Pm21 and Pm115) in PUL enzymes

In order to investigate synergy or complementarity between PUL components, four different core enzyme combinations, core2, *Pm*21@core3, *Pm*115@core3, and core4 respectively, were compared (Fig 3 and 4) using BGX as the substrate. Although *Pm*12, *Pm*24, *Pm*21 all display activity on WAX and BGX, BGX was chosen because it contains significant amounts of glucuronic acid substitutions (around 13% of glucuronic acid substitutions, Megazyme), necessary to investigate *Pm*115 activity. The inclusion of *Pm*21 (*Pm*21@core3 and core4) in the reaction mixture drastically increased in xylose release, consistent with the fact that we showed it is a xylosidase (Fig 3A). Core4 yielded to a higher amount of xylose compared to *Pm*21@core3 (Fig 3A), indicating that *Pm*115 may facilitate the action of *Pm*21 by removing substitutions (i.e. homosynergy between SCEs). Likewise, the amounts of oligosaccharides X2-6 released were lowest with core4, followed by *Pm*115@core3 (Fig 3B-F), suggesting that *Pm*115 doesn't complement *Pm*12 and *Pm*24 activities (i.e. no significant hetersynergy between SCE and MCE).

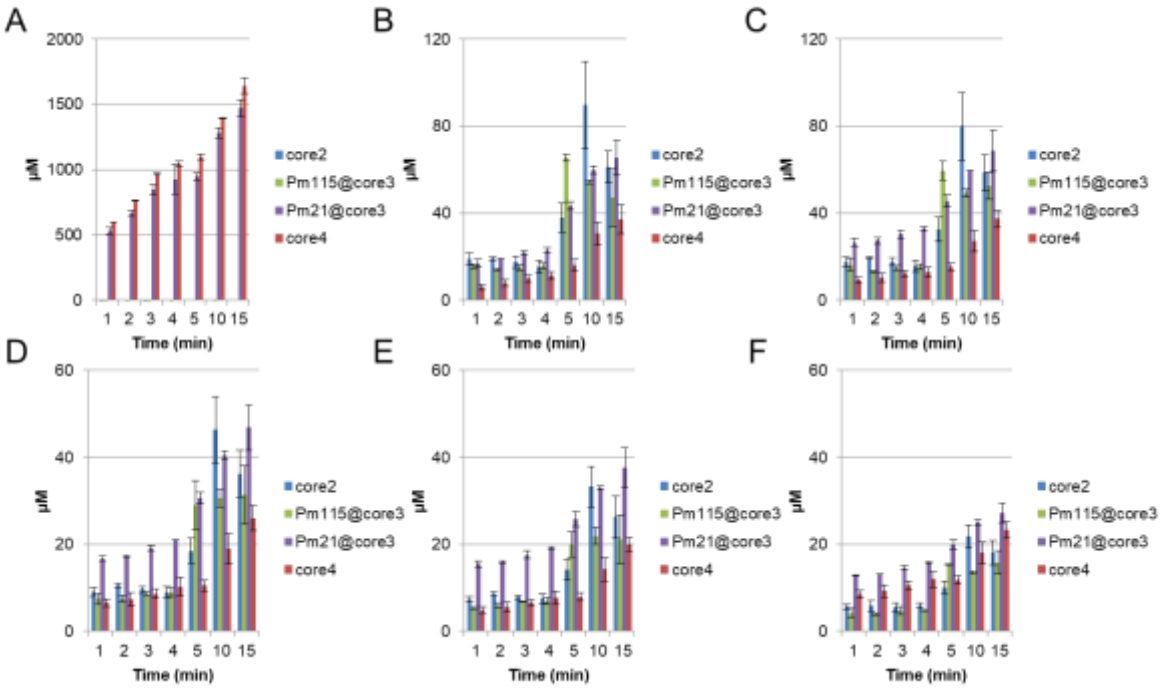


Fig 3. The product (DP from 1 to 6) profile of different combinations of core enzymes. The amount (in µM) of xylose (A), xylobiose (B), xylotriose (C), xylotetraose (D), xylopentaose (E), and xylohexaose (F) was quantified by HPAEC-PAD.

Regarding medium-sized XOS products, short time reactions were analysed (up to 15min) by mass spectrometry. The main products are from DP5 to DP20, only few conditions display DP up to 28. The presence of *Pm*21 (*i.e*, in *Pm*21@core3 and core4) led to a similar XOS size distribution, producing almost identical chain length abundance for all XOS groups from 1 to 15 min (Fig 4C and D). Moreover, the *Pm*21-containing mixtures yielded XOS of higher DP compared to core2 and *Pm*115@core3 (Fig 4A-D) which is rather inexpected. After 10 min, core 2 and *Pm*115@core3 also produced a similar profile, with XOS DP 6 to 12 and XOS DP 13 to 20 representing approximately 60% and 40% of the population respectively (Fig 4A and B). Calculating the amount of glucuronic acid substitutions per XOS group revealed a correlation with XOS length (Fig 4E-H). Also, comparing reactions catalyzed by core2 and *Pm*115@core3, it is noteworthy that the former produced more highly substituted XOS, consistent with the hypothesis that *Pm*115 is a glucuronidase (Fig 4E and F). In case of mixture with *Pm*21 it is possible that the increased length of XOS produced leads to increased number of substitutions (Fig 4C, D, G and H).

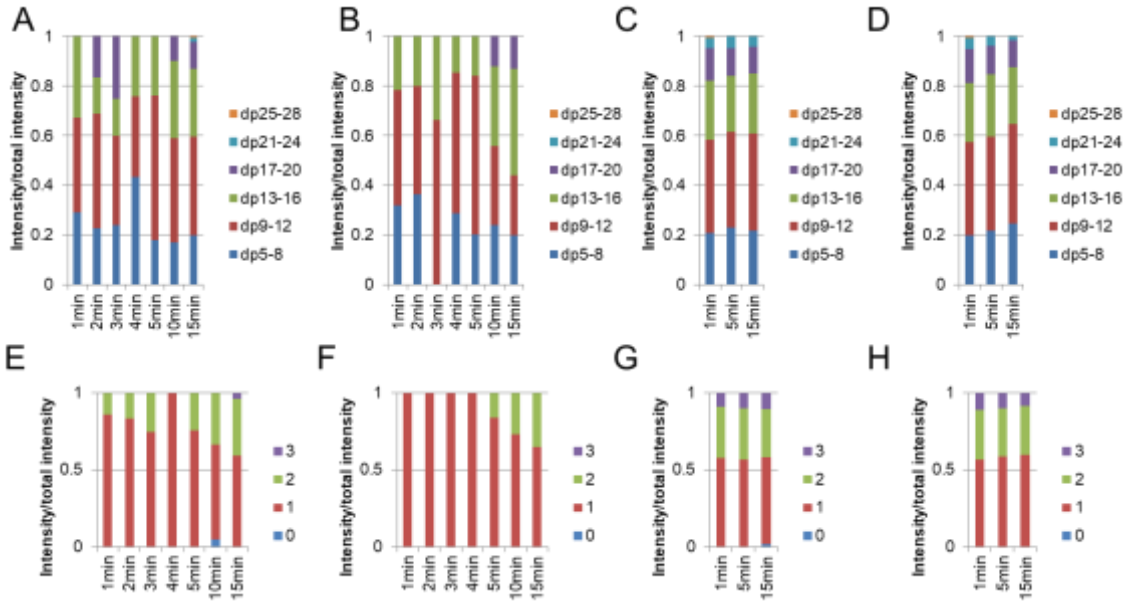


Fig 4. The product (DP from 5 to 28) profile of by different combinations of core enzymes. The populations of medium-sized products after treatment of core2 (A), *Pm*115@core3 (B), *Pm*21@core3 (C) and core4 (D) were accessed by MALDI-TOF. The number of methylated glucuronic acids in the products were counted after the treatment of core2(E) , *Pm*115@core3 (F), *Pm*21@core3 (G) and core4 (H).

Comparison of the product profile of Pm25 and Pm25ΔCBMs in the context of PUL enzymes

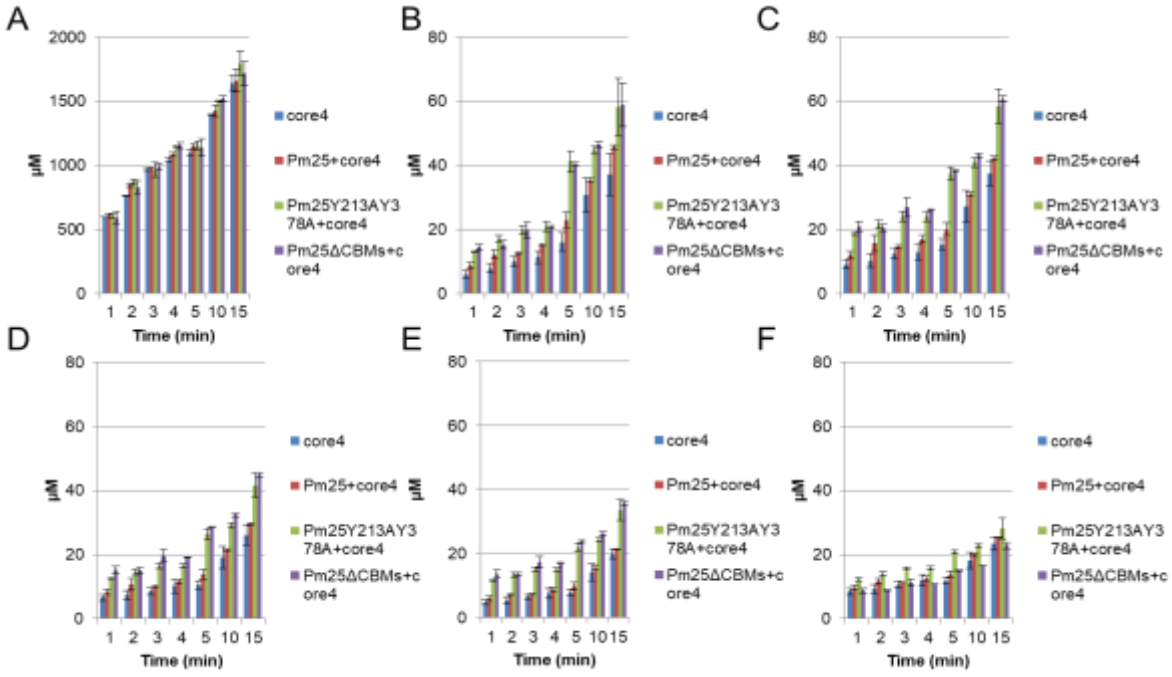


Fig 5. The product profile (DP from 1 to 6) of Pm25 and mutants in combination with core4. The amount of xylose (A), xylobiose (B) and xylotriose (C), xylotetraose (D), xylopentaose (E) and xylohexaose (F) were quantified by HPAEC-PAD.

In order to study whether the *Pm*25-borne CBM4s positively or negatively contribute to synergy between PUL enzymes, the product profile of the wild type and mutant *Pm*25 with core4 (Fig 5) and core2 (Fig 6) were compared. As mentioned before, *Pm*21 plays a role in shaping the product profile (Fig 4C and D) and indeed the addition of either wild type or mutated *Pm*25 made no real difference to core4 (Fig 5) and *Pm*21@core3 (data not shown) although *Pm*25 mutants slightly accelerates and increases the release of X2-X5 (Fig 5B-E), which agrees with the antisynergy of CBM and GH10 domain, observed in Fig2. The similar profile of X1-6 was also observed for reactions containing core2, *Pm*115@core3 and *Pm*21@core3 (data now shown).

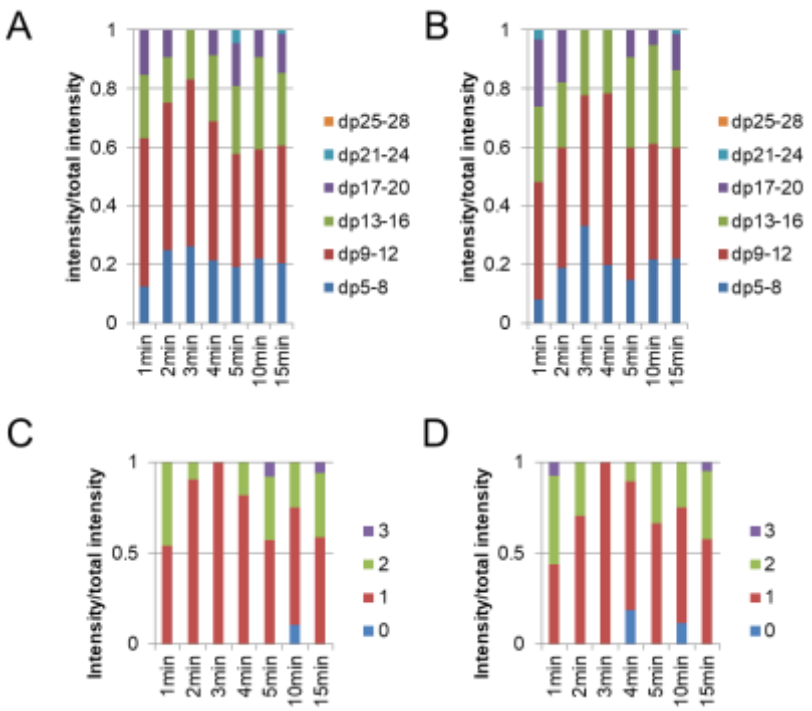


Fig 6. The product (DP from 5 to 28) pattern of *Pm*25 and *Pm*25ΔCBMs in combination with core2. The populations of medium-sized products after treatment of (A) core2+Pm25, (B) core2+Pm25ΔCBMs were represented as intensity ratio. The numbers of methylated glucuronic acids in the products were counted after the treatment of core2+Pm25 (C) and core2+Pm25ΔCBMs (D).

Regarding the longer XOS in the products, introduction of either *Pm*25 or *Pm*25ΔCBMs into core2 accelerated the appearance of diverse products (at 1 min) that were further hydrolyzed formining a stable product profile after 5 min, compared to 10 min when core2 alone was employed (Fig 4A, Fig 6A and B). Except for some minor fluctuations over the first 5 min, the product profile and substitution pattern of both *Pm*25 and *Pm*25ΔCBMs were almost identical from 10 to 15 min (Fig6), this also being the case for using other core enzymes, *Pm*115@core3, Pm21@core3 and core4 (data not shown).

CONCLUSION:

Characterization of the xylanases forming the putative *Pm*-derived PUL (Fig1) has provided the basis to define optimal conditions for testing of enzyme combinations. Overall, our results reveal that individually both *Pm*12 and *Pm*24 are more active xylanases than *Pm*25 (Table1), and indeed these enzymes figure among the most active xylanases reported (Pell et al., 2004). This disparity between the activities of the three enzymes suggests different functional roles within the PUL context. Basically, *Pm*12 and *Pm*24 appear to be the key

depolymerizing enzymes GH10 and GH11, while *Pm*21 plays important roles in xylose liberation and *Pm*115 provides sidechain-cleaving activities (Fig 3 and 4).

A load of researches proved that the CBM is not essential for hydrolytic activity of the enzyme towards soluble substrates (Ali et al., 2001; Black et al., 1996; Kellett et al., 1990; Valenzuela et al., 2012; Zhao et al., 2005). However these studies focused on only specific activity and product profile of short oligosaccharides, which is not compelling for the conclusion. Previously, the study of *Pm*25 revealed that the absence of its CBM4 domains did not radically alter its kinetic parameters or the product profile, when monitoring low DP (DP $\leqslant$ 6) products (*see* article I). The data obtained in this study confirmed this, showing that substrate depolymerization is not promoted by the presence of the CBM4 domains, which in fact produce a sort of antisynergy effect (Fig 2). Furthermore, other PUL enzymes, *Pm*12, *Pm*24, *Pm*21 and *Pm*115 were used to make core enzymes and synergistic effect was observed when *Pm*115 was added into *Pm*21 (Fig 3 and 4). Again, similar product (DP 1-28) and substitution pattern were observed between wild type and mutants of *Pm*25 when acting with other PUL enzymes (Fig 5 and 6).

All together these results suggest that, as one of the PUL compoments, the activity of *Pm*25 somehow complements other PUL components (Despres et al., 2016) but its contribution might be quite specific and/or subtle. However, further studies are required to confirm this by varying the substrates, concentration of enzymes/CBMs and substrates, and the order of addition, which may affect the cooperative action (Badino et al., 2017).

Reference:

Ali, M.K., Hayashi, H., Karita, S., Goto, M., Kimura, T., Sakka, K., and Ohmiya, K. (2001). Importance of the carbohydrate-binding module of Clostridium stercorarium Xyn10B to xylan hydrolysis. Biosci. Biotechnol. Biochem. 65, 41–47.

Arnal, G., Bastien, G., Monties, N., Abot, A., Leberre, V.A., Bozonnet, S., O'Donohue, M., and Dumon, C. (2014). Investigating the function of an Arabinan utilization locus isolated from a termite gut community. Appl. Environ. Microbiol. 81, 31–39.

Badino, S.F., Christensen, S.J., Kari, J., Windahl, M.S., Hvidt, S., Borch, K., and Westh, P. (2017). Exo-exo synergy between Cel6A and Cel7A from Hypocrea jecorina : role of carbohydrate binding module and the endo-lytic character of the enzymes. Biotechnol. Bioeng. 114, 1639–1647.

Bågenholm, V., Reddy, S.K., Bouraoui, H., Morrill, J., Kulcinskaja, E., Bahr, C.M., Aurelius, O., Rogers, T., Xiao, Y., Logan, D.T., et al. (2016). Galactomannan catabolism conferred by a polysaccharide utilisation locus of Bacteroides ovatus : enzyme synergy and crystal structure of a β-mannanase. J. Biol. Chem. 292, 229–243.

Bastien, G.G., Arnal, G.G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., Henrissat, B., Lefèvre, F., Robe, P., Bouchez, O., et al. (2013). Mining for hemicellulases in the fungus-growing termite Pseudacanthotermes militaris using functional metagenomics. Biotechnol. Biofuels 6, 78.

Biely, P., Vršanská, M., Tenkanen, M., and Kluepfel, D. (1997). Endo-β-1,4-xylanase families: Differences in catalytic properties. J. Biotechnol. 57, 151–166.

Black, G.W., Rixon, J.E., Clarke, J.H., Hazlewood, G.P., Theodorou, M.K., Morris, P., and Gilbert, H.J. (1996). Evidence that linker sequences and cellulose-binding domains enhance the activity of hemicellulases against complex substrates. Biochem. J. 319, 515–520.

Boraston, A.B., Bolam, D.N., Gilbert, H.J., Davies, G.J., Zverlov, V. V, Volkov, I.Y., Velikodvorskaya, G.A., and Schwarz, W.H. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. 382, 769–781.

Coughlan, M., and Hazlewood, G. (1993). beta-1,4-D-xylan-degrading enzyme systems: biochemistry, molecular biology and applications. Biotechnol. Appl. Biochem. 17, 259–289.

Despres, J., Forano, E., Lepercq, P., Comtet-Marre, S., Jubelin, G., Chambon, C., Yeoman, C.J., Berg Miller, M.E., Fields, C.J., Martens, E., et al. (2016). Xylan degradation by the human gut Bacteroides xylanisolvens XB1AT involves two distinct gene clusters that are linked at the transcriptional level. BMC Genomics 17, 326.

Din, N., Damude, H.G., Gilkes, N.R., Miller, R.C., Warren, R. a, and Kilburn, D.G. (1994). C1-Cx revisited: intramolecular synergism in a cellulase. Proc. Natl. Acad. Sci. U. S. A. 91, 11383–11387.

Dodd, D., Mackie, R.I., and Cann, I.K.O. (2011). Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. Mol. Microbiol. 79, 292–304.

Dondelinger, E., Aubry, N., Chaabane, F. Ben, Cohen, C., Tayeb, J., and Rémond, C. (2016). Contrasted enzymatic cocktails reveal the importance of cellulases and hemicellulases activity ratios for the hydrolysis of cellulose in presence of xylans. AMB Express 2–9.

Filho, E.X.F., G.Tuohy, M., Plus, J., and Coughlan, M.P. (1991). The xylan-degrading enzyme systems of Penicillium capsulatum and Talaromyces emersonii. Biochem Soc Trans 19, 25S.

Flint, H.J., Whitehead, T.R., Martin, J.C., and Gasparic, A. (1997). Interrupted catalytic domain structures in xylanases from two distantly elated strains of Prevotella ruminicola. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1337, 161–165.

Gao, D., Uppugundla, N., Chundawat, S.P., Yu, X., Hermanson, S., Gowda, K., Brumm, P., Mead, D., Balan, V., and Dale, B.E. (2011). Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. Biotechnol. Biofuels 4, 5.

Gebler, J., Gilkes, N.R., Claeyssens, M., Wilson, D.B., Beguin, P., Wakarchuk, W.W., Kilburn, D.G., Miller, R.C., Warren, R.A.J., and Withers, S.G. (1992). Stereoselective hydrolysis catalyzed by related β-1,4-glucanases and β- 1,4-xylanases. J. Biol. Chem. 267, 12559–12561.

Gilbert, M., Yaguchi, M., Watson, D.C., Wong, K.K.Y., Breuil, C., and Saddler, J.N. (1993). A comparison of two xylanases from the thermophilic fungi Thielavia terrestris and Thermoascus crustaceus. Appl. Microbiol. Biotechnol. 40, 508–514.

Grondin, J.M., Tamura, K., Déjean, G., Abbott, D.W., and Brumer, H. (2017). Polysaccharide Utilization Loci: Fuelling microbial communities. J. Bacteriol. 199, e00860-16.

Hu, J., and Saddler, J.N. (2018). Why does GH10 xylanase have better performance than GH11 xylanase for the deconstruction of pretreated biomass? Biomass and Bioenergy 110, 13–16.

Katapodis, P., Vršanská, M., Kekos, D., Nerinckx, W., Biely, P., Claeyssens, M., Macris, B.J., and Christakopoulos, P. (2003). Biochemical and catalytic properties of an endoxylanase purified from the culture filtrate of Sporotrichum thermophile. Carbohydr. Res. 338, 1881–1890.

Kellett, L.E., Poole, D.M., Ferreira, L.M., Durrant, a J., Hazlewood, G.P., and Gilbert, H.J. (1990). Xylanase B and an arabinofuranosidase from Pseudomonas fluorescens subsp. cellulosa contain identical cellulose-binding domains and are encoded by adjacent genes. Biochem. J. 272, 369–376.

Kim, I.J., Lee, H.J., Choi, I.G., and Kim, K.H. (2014). Synergistic proteins for the enhanced enzymatic hydrolysis of cellulose by cellulase. Appl. Microbiol. Biotechnol. 98, 8469–8480.

Kolenová, K., Vršanská, M., and Biely, P. (2006). Mode of action of endo-β-1,4-xylanases of families 10 and 11 on acidic xylooligosaccharides. J. Biotechnol. 121, 338–345.

Kormelink, F.J.M., and Voragen, A.G.J. (1993). Degradation of different [(glucurono)arabino] xylans by a combination of purified xylan-degrading enzymes. Appl. Microbiol. Biotechnol. 38, 688–695.

Koropatkin, N.M., and Smith, T.J. (2010). SusG: A Unique Cell-Membrane-Associated alpha-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. Structure 18, 200–215.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, 490–495.

Mackenzie, C.R., and Bilous, D. (1988). Ferulic Acid Esterase Activity from Schizophyllum commune. Appl. Environ. Microbiol. 54, 1170–1173.

Matsuzawa, T., Kaneko, S., Kishine, N., Fujimoto, Z., and Yaoi, K. (2017). Crystal structure of metagenomic β-xylosidase/ α-l-arabinofuranosidase activated by calcium. J. Biochem. 162, 173–181.

Miller, G.L. (1959). Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal. Chem. 31, 426–428.

NR.Gilkes, B.Henrissat, DG.kilburn, R.C.Miller, JR., and R.AJ.Warren (1991). Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. 55, 303–315.

Pell, G., Taylor, E.J., M.Gloster, T., Turkenburg, J.P., M.G.A.Fontes, C., M.A.Ferreira, L., Nagy, T., J.Clark, S., J.Davies, G., J.Gilbert, H., et al. (2004). The Mechanisms by Which Family 10 Glycoside Hydrolases Bind Decorated Substrates. J. Biol. Chem. 279, 9597–9605.

Pollet, A., Delcour, J. a, and Courtin, C.M. (2010). Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. Crit. Rev. Biotechnol. 30, 176–191.

Poutanen, K., and Sundberg, M. (1988). An acetyl esterase of Trichoderma reesei and its role in the hydrolysis of acetyl xylans. Appl. Microbiol. Biotechnol. 28, 419–424.

Puls, J., Schmidt, O., and Granzow, C. (1987). α-Glucuronidase in two microbial xylanolytic systems. Enzyme Microb. Technol. 9, 83–88.

Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Baslé, A., Morland, C., Day, A.M., Zheng, H., et al. (2015). Glycan complexity dictates microbial resource allocation in the large intestine. Nat. Commun. 6, 1–15.

Sørensen, H.R., Pedersen, S., Jørgensen, C.T., and Meyer, A.S. (2007). Enzymatic hydrolysis of wheat arabinoxylan by a recombinant “minimal” enzyme cocktail containing β-xylosidase and novel endo-1,4-β-xylanase and α-L-arabinofuranosidase activities. Biotechnol. Prog. 23, 100–107.

Suh, J.-H., Cho, S.-G., and Choi, Y.-J. (1996). Synergic effects among endo-xylanase, β-xylosidase, and α-L-arabinofuranosidase from Bacillus stearothermophilus. J. Microbiol. Biotechnol. 6, 179–183.
Sunna, A., and Antranikian, G. (1997). Xylanolytic enzymes from fungi and bacteria. Crit. Rev. Biotechnol. 17, 39–67.
Tabka, M.G., Herpoël-Gimbert, I., Monod, F., Asther, M., and Sigoillot, J.C. (2006). Enzymatic saccharification of wheat straw for bioethanol production by a combined cellulase xylanase and feruloyl esterase treatment. Enzyme Microb. Technol. 39, 897–902.

Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Leclerc, M., et al. (2010). Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome Res. 20, 1605–1612.
Tauzin, A.S., Laville, E., Xiao, Y., Nouaille, S., Le Bourgeois, P., Heux, S., Portais, J.C., Monsan, P., Martens, E.C., Potocki-Veronese, G., et al. (2016). Functional characterization of a gene locus from an uncultured gut Bacteroides conferring xylo-oligosaccharides utilization to Escherichia coli. Mol. Microbiol. 102, 579–592.
Valenzuela, S.V., Diaz, P., and Javier Pastor, F.I. (2012). Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module. Appl. Environ. Microbiol. 78, 3923–3931.
Vardakou, M., Katapodis, P., Topakas, E., Kekos, D., Macris, B.J., and Christakopoulos, P. (2004). Synergy between enzymes involved in the degradation of insoluble wheat flour arabinoxylan. Innov. Food Sci. Emerg. Technol. 5, 107–112.
Vincentelli, R., and Romier, C. (2013). Expression in Escherichia coli: Becoming faster and more complex. Curr. Opin. Struct. Biol. 23, 326–334.
de Vries, R.P., Kester, H.C.M., Poulsen, C.H., Benen, J. a E., and Visser, J. (2000). Synergy between accessory enzymes from Aspergillus in the degradation of plant cell wall polysaccharides. Carbohydr. Res. 327, 401–410.
Wang, K., Pereira, G. V., Cavalcante, J.J. V., Zhang, M., Mackie, R., and Cann, I. (2016a). Bacteroides intestinalis DSM 17393, a member of the human colonic microbiome, upregulates multiple endoxylanases during growth on xylan. Sci. Rep. 6, 1–11.
Wang, W., Yan, R., Nocek, B.P., Vuong, T. V., Leo, R. Di, Xu, X., Cui, H., Gatenholm, P., Toriz, G., Tenkanen, M., et al. (2016b). Biochemical and structural characterization of a five-domain GH115 α-Glucuronidase from the marine bacterium saccharophagus degradans 2-40T. J. Biol. Chem. 291, 14120–14133.

Weaver, J., Whitehead, T.R., Cotta, M.A., Valentine, P.C., and Salyers, A.A. (1992). Genetic analysis of a locus on the Bacteroides ovatus chromosome which contains xylan utilization genes. Appl. Environ. Microbiol. 58, 2764–2770.

White, B. a, Lamed, R., Bayer, E. a, and Flint, H.J. (2014). Biomass Utilization by Gut Microbiomes. Annu. Rev. Microbiol. 68, 279–296.

Withers, S.G., Dombroski, D., Berven, L.A., Kilburn, D.G., Miller, R.C., J. Warren, R.A., and Gilkes, N.R. (1986). Direct 1H N.M.R. determination of the stereochemical course of hydrolyses catalysed by glucanase components of the cellulase complex. Biochem. Biophys. Res. Commun. 139, 487–494.

Zhang, M., Chekan, J.R., Dodd, D., Hong, P.-Y.P.-Y.P.-Y.P.-Y., Radlinski, L., Revindran, V., Nair, S.K., Mackie, R.I., and Cann, I. (2014). Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. Proc. Natl. Acad. Sci. U. S. A. 111, E3708-17.

Zhao, G., Ali, E., Araki, R., Sakka, M., Kimura, T., and Sakka, K. (2005). Function of the family-9 and family-22 carbohydrate-binding modules in a modular β-1,3-1,4-glucanase/xylanase derived from Clostridium stercorarium Xyn10B. Biosci. Biotechnol. Biochem. 69, 1562–1567.

SUPPLEMENTARY MATERIAL:

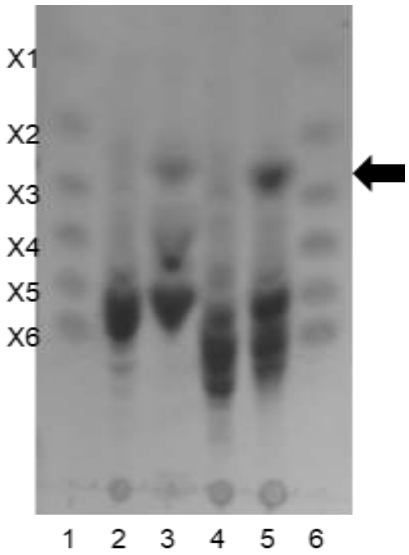


Fig S1. Hydrolysis products of UXXR and XUXXR (Megazyme) by *Pm*115. The reaction was performed with 5 µM of *Pm*115, 10 mM of UXXR/XUXXR in citric phosphate buffer (pH7) at room temperature for 114h. Lane 1 and lane 6 are mixed standard sugars; lane 2, UXXR; lane3, UXXR treated with *Pm*115; lane 4, XUXXR; lane5, XUXXR treated with *Pm*115. The arrow indicated the hydrolytic product of methylated glucuronic acid.

ARTICLE III

Quantifying CBM carbohydrate interactions using Microscale Thermophoresis

Haiyang Wu[1], Cédric Y. Montanier[1] and Claire Dumon[1]*

[1]LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

*To whom correspondence should be addressed: *LISBP-INSA Toulouse, 135 Avenue de, Rangueil, 31077 Toulouse, France. Phone*: + 33 (0)5 61 55 94 93; Fax: +33 (0)5 61 55 94 00; Email: claire.dumon@insa-toulouse.fr

Key words: Microscale thermophoresis (MST), binding studies, fluorescence quenching, fluorescent label, K_d (dissociation constant).

Running Head: Microscale Thermophoresis

ABSTRACT

MicroScale Thermophoresis (MST) is an emerging technology for studying a broad range of biomolecular interactions with a high sensitivity. The affinity constant can be obtained for a wide range of molecules within minutes based on reactions in microliters. Here we describe the application of MST in quantifying two CBM-carbohydrate interactions, a CBM3a towards cellulose nanocrystals and a CBM4 (CBM4-1 in Article I) against xylohexaose.

1 INTRODUCTION

Protein-carbohydrate interactions are involved in a myriad of important biological functions. Therefore many techniques are used in order to quantify and characterize those interactions, such as Isothermal Titration Calorimetry (ITC) which allows characterization of thermodynamics during interaction process, Affinity Gel Electrophoresis (AGE) using soluble polysaccharides, solid state depletion assay using insoluble polysaccharides, UV difference and fluorescence spectroscopy (Abbott and Boraston, 2012).

Thermophoresis, or thermodiffusion / Soret effect, was first described by Carl Ludwig in 1856. This technique uncovered the different responses of particles in a mobile phase towards a gradient of temperature. The miniaturized thermophoresis, namely MST was developed in the last decade to monitor the interactions and calculate the thermodynamic constants among biomolecules ranging in size from 10^1 to 10^7 Dalton in only few microliters. The description of thermophoresis in a frame of a local thermodynamic equilibrium allows one to quantify the thermophoretic behavior based on the charge, size and hydration shell of biomolecules (Duhr and Braun, 2006a, 2006b). Basically, the MST instrument contains an infra-red (IR) laser which induces a microscopic temperature gradient (e.g. 1°K/10 µm) in a glass capillary, a high-power LED (light-emitting diode) as source of light excitation, a fluorescence detector, and a sample tray of 16 capillaries (Baaske et al., 2010; Jerabek-Willemsen et al., 2014). Briefly, the labeled molecule is mixed with a decreasing concentration of non-fluorescent ligand and each mixture is loaded into one of the 16 glass capillaries. Initial fluorescence is measured as a base line in each capillary and the IR laser is switched on. The IR laser beam is focused into the mid-point of capillary and strongly absorbed by aqueous solution resulting in a temperature gradient from the laser spot to the outer region. Simultaneously, a dichroic mirror allows fluorescence from labeled molecule to be continuously monitored, while complexes migrate in the capillaries. The fluorescence in the laser spot area decreases or increases during thermophoresis, at a velocity related to the concentration of ligand, due to change in size, charge or hydration shell of the complex. Each capillary is scanned in minutes one after another. The constant of dissociation K_d is deduced from the binding curve derived by plotting the normalized fluorescence (F_{Norm}) at a given time against each ligand concentration (L) (Baaske et al., 2010; Duhr et al., 2006b).

Depending on the molecules investigated, the source of fluorescence could be intrinsic (*e.g.* indole side chain of tryptophan), coming from attached dyes (*e.g.* Fluorescein,

Alexa546) or from fluorescent protein (*e.g.* GFP, YFP). The fluorescence excitation/detection spectrum range would lead to choose appropriate apparatus. Several labeling strategies could be employed to target proteins by attaching for instance specific fluorophore molecule to the ε-amine group of lysine, thiol group of cysteine, fusion protein (such as GFP), or even unnatural amino acids.

Here we present the use of the MST for the quantification of CBM/ligand interactions. The CBMs used in our protocol are the CBM3a from *Clostridium Thermocellum* (Tormo et al., 1996) and a newly identified CBM4 from termite gut microbiota (Bastien et al., 2013). They display affinity for insoluble crystalline cellulose and soluble xylan respectively. Both CBMs were labeled to primary amine group of lysine using an adapted protocol from the protein labeling kit provided by NanoTemper and then MST experiments were performed using Monolith NT. 115 instrument.

2 MATERIALS

All solutions were prepared using deionized water dH_2O and analytical grade reagents. Buffers were prepared and stored at room temperature. Protein samples were kept on ice until use.

2.1 Chemicals and kits

1. Deionized water (dH_2O)
2. Protein Labeling Kit RED-NHS (Cat# MO-L001 RED-NHS, NanoTemper Technologies GmbH).
3. DMSO.
4. Regents for SDS-PAGE (*see* **Note 1**).
5. Xylohexaose.
6. Cellulose nanocrystals.
7. Sodium dihydrogen phosphate dehydrate ($NaH_2PO_4 \cdot 2H_2O$).
8. Disodium hydrogen phosphate dodecahydrate ($Na_2HPO_4 \cdot 12H_2O$).
9. Imidazole.
10. 5% Pluronic® F-127 (Nanotemper).
11. Sodium dodecyl sulfate (SDS, optional for SD test).
12. Dithiothreitol (DTT, optional for SD test).

2.2 Buffers and Solutions

1. Labeling buffer: 50 mM sodium phosphate buffer pH 7.0. To prepare 0.5 M NaH_2PO_4 (monobasic), weigh 78 g of $NaH_2PO_4 \cdot 2H_2O$ and transfer to a 1 L glass beaker containing about 800 mL of dH_2O. Dissolve by magnetic stirring. Adjust the volume up to 1 L with dH_2O and filter at 0.22 µm. To prepare 0.5 M Na_2HPO_4 (dibasic), weigh 179 g of $Na_2HPO_4 \cdot 12H_2O$ in 1 L and prepare the solution as described above. Mix 250 mL of the monobasic stock solution and 700 mL of the dibasic stock solution to obtain 0.5 M sodium phosphate buffer pH 7 stock solution. Check the pH. Transfer 10 mL of the 0.5 M stock solution to 90 mL of dH_2O in a measuring cylinder to obtain 100 mL of 50 mM sodium phosphate buffer pH 7.0 (*see* **Note 2**).
2. Elution buffer: 50 mM sodium phosphate, 200 mM imidazole buffer pH 7.0. Dissolve 1.36 g of imidazole in 100 mL of labeling buffer. Check the pH and adjust with phosphoric acid if needed.
3. Xylohexaose (300 mM). Weigh 0.015 g of xylohexaose in a 1.5 mL micro centrifuge tube. Add 62 µl of dH_2O and vortex to solubilize. Store at 4°C (*see* **Note 3**).
4. Cellulose nanocrystals at 20 g/L (Martinez et al., 2015).
5. SD mix (optional for SD test) : 4% SDS, 40mM DTT. Weigh 0.4 g of SDS, 0.0617 g of DTT and dissolve in 10 mL of dH_2O.

2.3 Biological material

Proteins: CBM3a (Blake et al., 2006; Tormo et al., 1996) and CBM4 (Bastien et al., 2013), were produced in house as His-tag fusion proteins.

2.4 Sample preparation materials

1. Vivapure Metal Chelate Mini spin columns (Cat# VS-MC01MC12) (*see* **Note 4**).
2. TALON® Metal Affinity Resin (Cat# 635504, Clontech) (*see* **Note 5**).
3. MO-K005 Monolith™ NT.115 MST Premium Coated Capillaries (*see* **Note 6**).
4. 0.2 mL individual PCR tubes (*see* **Note 7**).
5. 1.5 mL micro centrifuge tubes.

2.5 Instruments

1. Monolith® NT.115 Blue/Red (Cat # G008, NanoTemper Technologies GmbH) (*see* **Note 8)**.

2. Bench centrifuge (*see* **Note 9**).
3. Magnetic stirrer.

2.6 Computer software

1. MO.Control software.
2. MO.Affinity Analysis software.

3 METHODS

3.1 Labeling and purifying proteins.

Protein Labeling Kit RED-NHS protocol was used to label our CBMs (http://www.helsinki.fi/biosciences/corefacilities/microscalethermophoresis/Protein_Labeling_Manual_V012-RED-NHS.pdf). Procedure for purification of labeled CBMs was modified as our CBMs displayed affinity for the resin of the desalting column provided in the kit (*see* **Note 4**).

1. As our CBMs were already in 50 mM phosphate buffer pH 7.0, the buffer exchange step was not necessary (*see* **Note 2**).
2. Protein concentration was adjusted to 20 µM in 100 µL in the labeling buffer (*see* **Note 10**). The original concentration of protein is determined by measuring the absorbance at 280 nm and calculated according to the Beer-Lambert Law using theoretical extinction coefficient determined by protparam tool on expasy website (http://web.expasy.org/protparam/) (Gasteiger et al., 2005).
3. According to the manufacturer protocol, solid fluorescent dye concentration was set to 470 µM by adding 30 µl of DMSO into the solid dye, and vigorously vortex until complete dissolution (*see* **Note 11**).
4. 15 µl of dye was added to 85 µl of labeling buffer to reach 3.5 fold concentration of the protein (*see* **Note 12**).
5. Mix 100 µl of protein to 100 µl of dye and incubate 1h at room temperature, protected from light by aluminum foil. Then, place labeled protein on ice.

   We changed protocol from Step C of Labeling Kit protocol.
6. Add 100 µl of TALON® Metal Affinity Resins into the Vivapure Metal Chelate Mini spin column, centrifuge at 1500 rcf 1 min at 10 °C to remove storage buffer.
7. Wash the resin with 500 µl of ice cold dH_2O, leave caps open (*see* **Note 13**), centrifuge as described in step 6, and discard the flow through.

8. Equilibrate the resin with 400 µl of ice cold labeling buffer, centrifuge as described in step 6, and discard the flow through. Repeat this step three times.
9. Load 200 µl of labeled protein (*see* step 5) to the center of the column. Gently mix the protein-dye mixture with resin by shaking and incubate on ice for 15 min. Centrifuge as described in step 6 to remove excess of free dye.
10. Wash with 400 µl of ice cold labeling buffer, centrifuge as described in step 6, and discard the flow through. Repeat this step two times.
11. Add 200 µl of ice cold elution buffer, resuspend gently the resin by shaking and incubate on ice for 5 min. Centrifuge at 1,500 rcf 3 min at 10 °C; collect the flow through as the labeled protein solution #1. Repeat one time and collect the flow through as labeled protein solution #2.
12. Assess the concentration of the labeled protein by SDS-PAGE (*see* **Note 1**).

3.2 Assay optimization

Before setting up an assay, determining the optimized condition is required. Regarding this, the starting guide NT115 was followed: (http://www.helsinki.fi/biosciences/corefacilities/microscalethermophoresis/StartingGuide_NT115.pdf). Briefly, the concentration of labeled protein is adjusted to reach an initial fluorescence counts between 500 and 1500. Different types of capillary and additives such as BSA, Tween 20 or Pluronic have to be tested to avoid sample aggregation or adsorption to the inner surface of capillary. Using Tween 20 will also result in an increasing of the sample fluorescence. Buffer composition may also influence fluorescence. Proper experimental condition will result in a symmetrical fluorescence peak (see Note 14).

1. The ligands, buffers and all the additives used in the assay are scanned with LED power 60% to make sure there is no background fluorescence.
2. Premium capillaries were used for both CBM3a and CBM4, and Pluronic® F-127 was used at a final concentration of 0.05% in the dilution buffer to avoid severe adsorption to the surface of the capillaries (*see* **Note 14**).
3. To obtain an initial fluorescence counts between 500 and 1500 while fixing LED power to 60%, MST power to 20%, CBM3a and CBM4 were diluted to 4.3 10^{-4} g/L and 0.057 µM respectively with dilution buffer containing Pluronic® F-127 (0.05% final)(*see* **Note 15**).
4. The highest concentration of the ligand in the reaction should be 20-fold above the expected K_d, but not below 100 nM (the lowest K_d which can be detected by NT.115 is 1 nM). For CBM3a, the highest concentration of cellulose nanocrystals was 10 g/L. For

CBM4, the highest concentration of xylohexaose was 150 mM (*see* **Note 16**). Both ligands were diluted using the dilution buffer.

3.3 Assay setup

A titration series of 16 dilutions of the ligand is prepared in 0.2 mL PCR tubes. Transfer 20 µl of the highest ligand concentration into vial 1 (*see* **Note 16**), then proceed to cascade dilution as follow: add 10 µl of dilution buffer into vial 2 to vial 16. Transfer 10 µl from vial 1 to vial 2, mix by pipetting up and down three times (*see* **Note 17)**. Then transfer 10 µl from vial 2 to 3. Repeat this step successively for all vials. Remove 10 µl from vial 16 after mixing to end up with 10 µL final**.**

1. Prepare 2 X stock solution of labeled protein (*see* **Note 18)**. CBM3a and CBM4 were diluted to 8.6 10^{-4} g/L and 0.114 µM respectively with dilution buffer and Pluronic® F-127 (0.1% final). Add 10 µl of protein stock to each vial of the dilution series (*see* **Note 19**). Mix by pipetting up and down three times.
2. Fill capillaries by sticking the capillary horizontally into the reaction tube to aspirate the sample (*see* **Note 20**). The capillary containing the highest ligand concentration is placed in the front of the sample tray.
3. Place the sample tray in the instrument and perform the scan with the MO.Control software (*see* **Note 21**).

3.4 Data Setting

3.4.1 CBM3a

1. After performing the capillary scan and MST experiment, the initial fluorescence intensity was checked (Fig. 1a). The variation of fluorescence is less than ± 10% and the MST signal could be analyzed (*see* **Note 22**).
2. Select the tab "MST Analysis Set" in the MO.Affinity Analysis software. The fluorescence ratio before and after switching on the laser is calculated and plotted against the ligand concentration. A sigmoidal dose-response curve is obtained (*see* **Note 23**). Choose K_d model, enter the concentration of CBM3a in the same unit used as for the ligand.
3. Choose the ΔF_{Norm} as y-axis (Fig. 1b). The sigmoidal curve needs to have (1) a clear baseline and a clear saturation (at least 3 points respectively), (2) an amplitude/noise > 3 (*see* **Note 24***).*

4. Then the K_d value is determined for CBM3a. MO.Affinity Analysis software could also evaluate the consistency among repeated experiments by simply including all the repeated experiments that have been done in the exact same conditions into the same analysis set, and errors bars will automatically appeared in the dose response curve.
5. The K_d of CBM3a towards cellulose nanocrystals deduced from the binding curve is *0.24* ± 0.05 g/L. Regarding the complexity of cellulose nanocrystals, K_d would be considered as an apparent K_d.

3.4.2 CBM4

1. The initial fluorescence shown in Fig. 2a displays a typical quenching sigmoidal curve not compatible with proper MST experiment. In such case it is recommended to perform a SDS-denaturation Test or SD-test to investigate if the quenching comes from a binding event or non-specific interactions with the ligand and the fluorophore, nonspecific adsorption to tube walls, protein aggregation, or just because the fluorophore is in close vicinity to the binding site (*see* **Note 25**). To perform a SD-Test, the samples 1-3 and 14-16 from the 16 reaction series are centrifuged 10 min at maximum speed in order to avoid sample adsorption to the tube walls. 10 µl of SD mix is added to each sample and well mixed by pipetting up and down three times. The mixture is incubated for 5 minutes at 95 °C and centrifuge briefly in order to denature the protein. The fluorescence intensity of each sample is measured after filling capillaries (*see* **Note 26**). In our case, protein denaturation leads to restore uniform fluorescence intensity in all capillaries, meaning that the variation of fluorescence was induced by a binding event. Select the tab "Initial Fluorescence Analysis Set" in the MO.Affinity Analysis software, select the raw data and then move to "Dose Response Fit". Choose K_d Model, fix the concentration of CBM4, a sigmoid curve is derived and well fitted to the data.
2. Choose ΔRaw fluorescence [counts] as y-axis, and then the dose response curve appears showing that the fluorescence intensity decreases with increasing concentrations of the ligand (Fig. 2b). The sigmoidal curve needs to have (1) a clear baseline and a clear saturation (at least 3 points respectively), (2*)* an amplitude/noise > 3(*see* **Note 24**).
3. The K_d of CBM4 towards xylohexaose derived from the binding curve was $2.5 \pm 0.3\ 10^3$ µM.

The two proteins were labeled on the lysine. However unlike CBM3a binding site (Fig. 1c) quenching fluorescence within CBM4 could be explained by the presence of a putative labelled lysine close to the binding site (Fig. 2c).

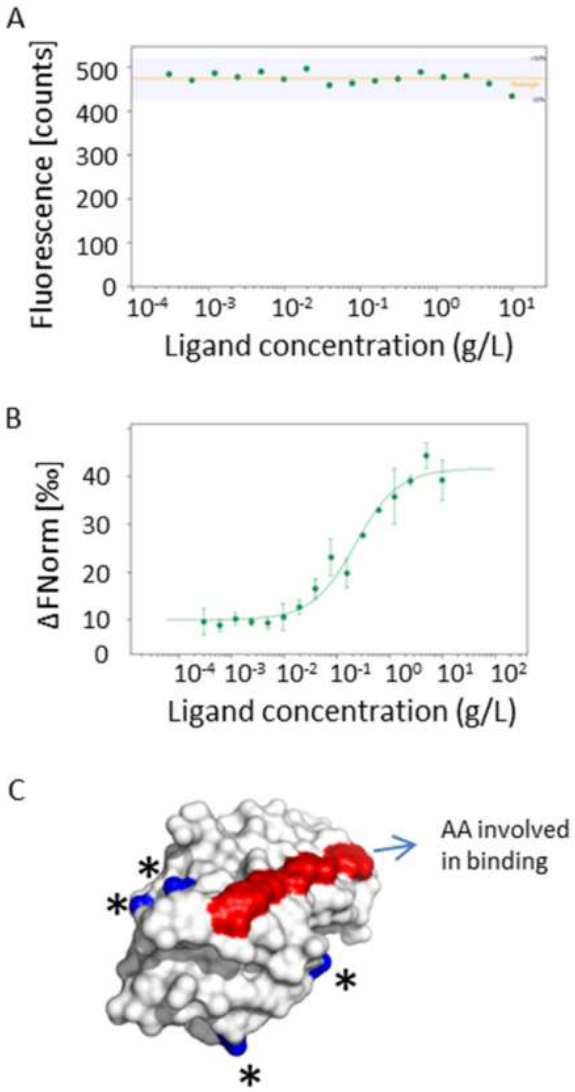


Fig. 1 CBM3a. (**A**) Initial fluorescence of CBM3a towards different concentrations of cellulose nanocristals. The data showed a homogenous initial fluorescence, which indicated that MST experiment could be performed. (**B**) Measurement of CBM3a binding to cellulose nanocristals. K_d apparent = 0.24 ± 0.05 g/L. (**C**) Surface representation of CBM3a (pdb 1NBC). Arrow indicates residues involved in cellulose binding. * Indicate lysine residues potentially labeled with NHS-dye. None of these residues are in closed vicinity to the flat surface of the binding site.

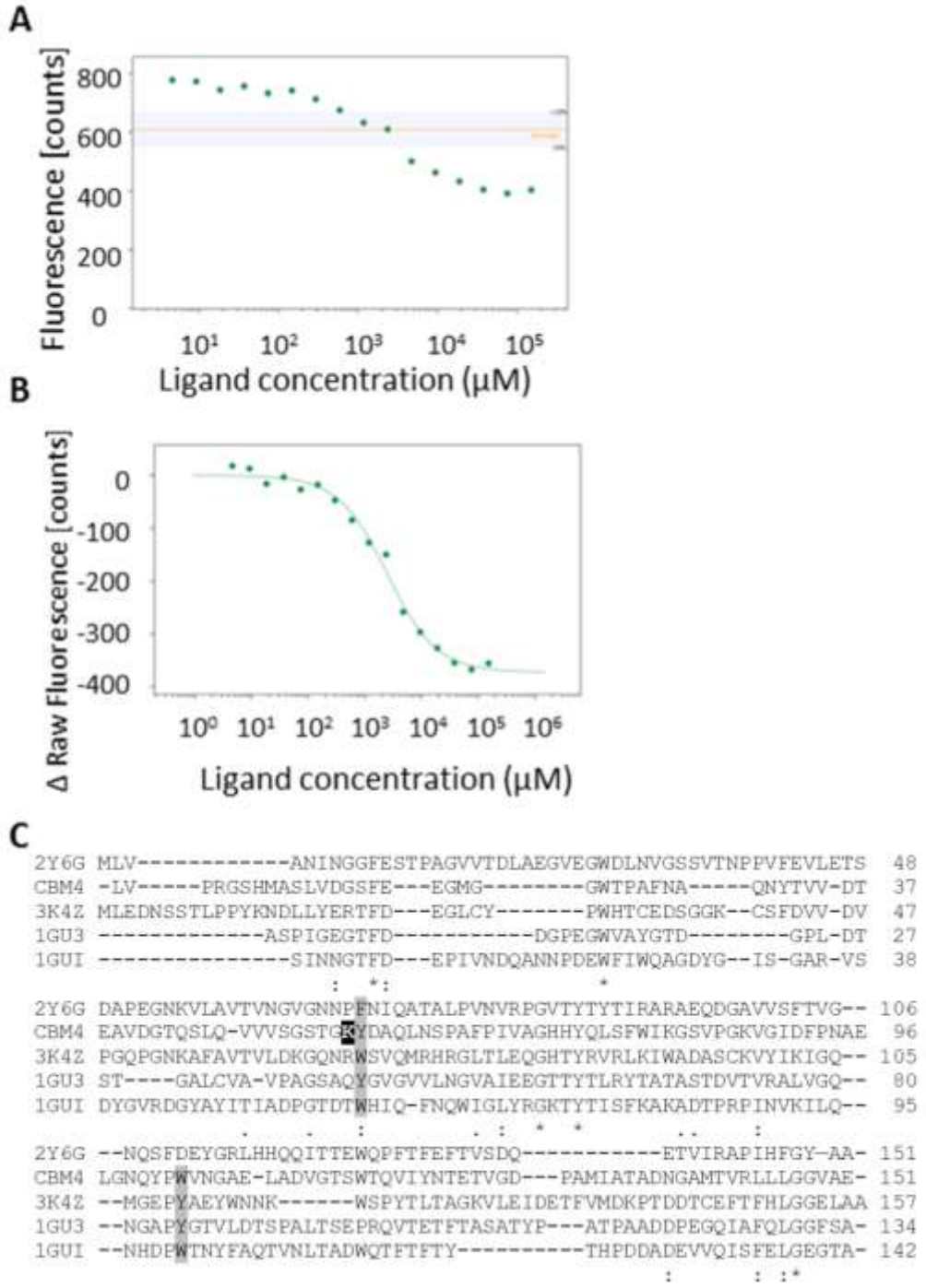


Fig. 2 CBM4. (**A**) Initial fluorescence of CBM4 towards different concentrations of xylohexaose. Quenching of fluorescence indicated that MST experiment could not be performed. (**B**) Measurement of CBM4 binding to xylohexaose. $K_d = 2.5 \pm 0.3\ 10^3\ \mu M$. (**C**) Multiple protein sequence alignment of CBM4 with family 4 CBMs of known crystallographic structure. In grey, mains residues involved in ligand recognition. In black, CBM4 lysine residue potentially labeled with NHS-dye. Close proximity between Lys56 and Tyr57 in the binding site may explain the quenching of fluorescence during titration.

4 NOTES

1. After purifying the protein by IMAC, determining the protein concentration was not possible by UV absorption because of imidazole. In this case, to avoid dialysis dilution, we performed sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with labeled protein and a serial diluted protein as standards, and then quantify the labeled protein by the quantifying tool in the software of Image Lab 5.2.1.

2. The buffer of the labeling kit is an amine free buffer at pH 8. Our protein storage buffer was 50 mM phosphate buffer pH 7.0. It is also an amine free buffer. Labeling efficiency using NHS-ester group is optimal at pH 8 but could also be performed at pH 7, not below. The $Na_2HPO_4 \cdot 12H_2O$ can be dissolved faster if the dH_2O is warmed up to about 50°C. If the binding event is calcium dependent, HEPES buffer is recommended instead of the phosphate buffer.

3. Dealing with costly ligand such as xylohexaose, it is better to start with a rather high concentration (in our case 300 mM) for the very first trial before knowing the K_d to make sure the CBM can be saturated.

4. The Gravity Flow Column B for purification provided in Monolith NT™ Protein Labeling Kit is a dextran-based gel filtration column. CBM3a and CBM4 displayed affinity for dextran, and were retained by the desalting column. Therefore, vivapure metal chelate mini spin columns were selected because they are made of cellulose free based membrane. Furthermore it allows the use of small volume of IMAC resins in the purification step.

5. The TALON Metal Affinity Resin is made with Sepharose CL-6B, an agarose based polysaccharide polymer material, which has no affinity with the CBMs studied.

6. There are three different types of capillaries for NT.115, standard treated capillaries, hydrophobic capillaries and premium coated capillaries. The premium coated one is the best to tackle with samples that have significant adsorption to surfaces, although it is the most expensive. We recommend trying with the standard treated capillaries first using buffer supplemented with additives such as BSA or detergent if the protein sticks to the inner surfaces of the capillary. If the adsorption is still a problem, then try with other capillaries.

7. 20 µl of each sample is usually prepared for MST experiment, thus a 0.2 mL individual tube is enough. Bigger reaction tubes are not recommended since the large surface area to volume ratio can increase adsorption to the tube wall and loss of protein. Besides, the 0.2 mL

strip tubes are also recommended for their convenient manipulation especially in the serial dilution step.

8. A red channel is always recommended as it allows to measure in complex bioliquids (cell lysate, serum) and no autofluorescence of a molecule in this channel has been observed.

9. A refrigerated and temperature regulated centrifuge is recommended for the centrifuge of protein samples to maintain their stability.

10. We recommend preparing a little bit more than 100 µl (e.g. 120 µl) for accurate pipetting afterwards.

11. It is possible to aliquot dye in DMSO and store at -20°C.

12. In our case, as the pH of the labeling buffer is not optimal we used a higher concentration of dye to improve the labeling efficiency which was 3.5 fold of the protein concentration.

13. The maximum volume of the column is 600 µl, and adding 500 µl of dH_2O into the column does not leave space for the air. Therefore, leaving the cap open is recommended to ensure that the wash buffer is completely removed.

14. Buffers containing detergent (0.05% pluronic acid, 0.05% tween-20), BSA (0.05 – 0.5 mg/mL), casein and others can be tested. A modified pH or ionic strength can also improve the performance.

15. Concentration of the CBM3a was expressed in g/L to be in accordance to the unit of its substrate concentration, cellulose nanocrystals.

16. The highest concentration of xylohexaose in the reaction is 150 mM therefore 300 mM stock solution is required as the ligand will be diluted twice when adding the same volume of protein. We recommend to mix the solution by pipetting up and down carefully and not to vortex in order to avoid bubbles.

17. Vortex will result in a possible adsorption of protein to the inner wall of the tube and considering the very low concentration of protein, even small adsorption would result in a significant decreasing of protein concentration and so to a non-optimal assay.

18. The protein and buffer additives will be diluted twice after mixing with the same volume of ligand keep this in mind when designing the experiment.

19. Stock solution of both protein and ligand required to be centrifuged at full speed 5 min before mixing to avoid aggregates in capillaries.

20. Avoid touching the capillary in the middle where the optical measurement will be performed because it could modify the fluorescence.

21. Detailed introduction of the manipulation of the instrument and the software can be downloaded elsewhere http://www.helsinki.fi/biosciences/corefacilities/microscalethermophoresis/Manual_NT115.pdf.

22. It is not recommended to use the data to determine a binding affinity if the deviation is more than ±10%.

23. If a sigmoidal dose-response curve is not obtained with 20% nor 40% MST power, use 60% or 80% MST power.

24. The amplitude is the difference between bound and unbound state, the noise is the variations of fluorescence in the baseline. I*f a high noise level is observed, experimental conditions need to be improved by using detergents for instance.*

25. The SD-test allows verifying that the fluorescence quenching is due to the binding event. SD-test cannot be performed with samples containing potassium (200 mM or more) because the SDS will precipitate.

26. The SD-test is performed only with sample 1 to 3 and 14 to 16 which are the highest and lowest ligand concentrations respectively. In case of binding event, these contrasted samples should have no difference in fluorescence. See also (http://openwetware.org/images/5/59/FAQ_Fluorescence_Changes_V11-2.pdf).

Acknowledgement:

We acknowledge the Fédération de Recherche Agrobiosciences, Interactions et Biodiversité (FR 3450), CNRS, Université de Toulouse, UPS, Castanet-Tolosan, France, and the IDEX "UNITI" Université de Toulouse (GO-AHEAD project) for the Nanotemper Monolith NT.115 facilities.

References:

Abbott, D.W., and Boraston, A.B. (2012). Quantitative approaches to the analysis of carbohydrate-binding module function (Elsevier Inc.).

Baaske, P., Wienken, C.J., Reineck, P., Duhr, S., and Braun, D. (2010). Optical thermophoresis for quantifying the buffer dependence of aptamer binding. Angew. Chemie - Int. Ed. 49, 2238–2241.

Bastien, G.G., Arnal, G.G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., Henrissat, B., Lefèvre, F., Robe, P., Bouchez, O., et al. (2013). Mining for hemicellulases in the fungus-growing termite Pseudacanthotermes militaris using functional metagenomics. Biotechnol. Biofuels 6, 78.

Blake, A.W., McCartney, L., Flint, J.E., Bolam, D.N., Boraston, A.B., Gilbert, H.J., Knox, J.P., Flint, E., Bolam, D.N., Alisdair, B., et al. (2006). Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. J. Biol. Chem. 281, 29321–29329.

Duhr, S., and Braun, D. (2006a). Thermophoretic depletion follows boltzmann distribution. Phys. Rev. Lett. 96, 1–4.

Duhr, S., and Braun, D. (2006b). Why molecules move along a temperature gradient. Proc. Natl. Acad. Sci. U. S. A. 103, 19678–19682.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, (Totowa, NJ: Humana Press), pp. 571–607.

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P., and Breitsprecher, D. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. J. Mol. Struct. 1077, 101–113.

Martinez, T., Texier, H., Nahoum, V., Lafitte, C., Cioci, G., Heux, L., Dumas, B., O'Donohue, M., Gaulin, E., and Dumon, C. (2015). Probing the functions of carbohydrate binding modules in the cbel protein from the oomycete phytophthora parasitica. PLoS One 10, 1–14.

Tormo, J., Lamed, R., Chirino, a J., Morag, E., Bayer, E. a, Shoham, Y., and Steitz, T. a (1996). Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. EMBO J. 15, 5739–5751.

GENERAL CONCLUSION AND FUTURE WORK

The plant cell wall polysaccharides are of industrial relevance since they represent the most abundant carbon source on earth. PCW polysacharide valorization is an important target of the bioeconomy which implies developing biocatalysts in order to degrade this highly recalcitrant material in milde conditions and high yields. After cellulases, xylanases are important key player in this field targeting the hemicellulose part of the PCW enhancing the cellulose hydrolysis by reducing the inhibitors and realizing pentose sugars which can be further valorized. Because PCW polysaccharides are also an important part of our diet, xylanase are used in a wide range of industrial applications such as the pulp and paper industry, detergents, in food and feed or bakery. Idealy, enzyme in the deconstruction of plant cell wall shoud be highly effective, thermostable, alkaline/acid tolerable, synergistically work with other enzymes to result in low enzyme loading depending on the process. This doctoral thesis mainly focuses on the characterization of a GH10 xylanase, *Pm*25 which was identified from a xylan utilization system XUS in the metagenome of the termite gut *Pseudacantotermes militaris*.

*Pm*25 is a typical GH10 encountered only in XUS displaying a peculiar sequence organization where the GH10 catalytic module is inserted by two CBM4. The role of the enzyme in the XUS is still not clear and the objective of this thesis was to characterize the biochemical and catalytic properties of *Pm*25 and to interrogate the role of the CBM4 in the function.

How has this thesis contributed to the enzymology in terms of biorefinery?

A detailed biochemical characterization of *Pm*25 was performed. The physiochemical parameters showed *Pm*25 could act in a broad pH range (pH 5 to 9), which is of industrial interests.

Despite the pH range, compared with other GH10 xylanases in literature, *Pm*25 and its homologues are over tens and hundreds times lower in catalytic efficiency both with polysaccharides and oligosaccharides (Despres et al., 2016).

In the XUS fragment, *Pm*25 is surrounded by other xylanases, therefore the activity of *Pm*25 in the context of the XUS was investigated. The results showed that the other GH10 xylanase, *Pm*12, was among one of the most active GH10 xylanases (Mirande et al., 2010; Pell et al., 2004). Although both *Pm*25 and *Pm*12 are located in the same gene cluster, their putative

cellular location is different, the former is anchored onto the cell-membrane and the latter is located in the periplasm based on the model for XUS in xylanolytic *Bacteroidetes* (Dodd et al., 2011). It was proposed that the *Pm*25 homologues are responsible for "limited" degradation of glycan at the cell surface before importing into the periplasm to avoid diffusing the XOS products that may be uptaken by other bacteria. Thus, *Pm*25 and its homologues seem to be "designed" as an inefficient enzyme in producing the final products constituting a "selfish strategy" for the bacteria in the ecosystem. Combing the experimental data and the putative model for xylan utilization system, we suggest that the periplasmic enzymes are more favorable for biorefinery than the suface enzymes from the PUL of *Bacteroidetes*.

Regarding the domain organization, *Pm*25 and its homologues are not the only ones showing insertional architecture. In fact, 9% of domain combinations are discontinuous, with one domain being inserted into another one (Aroul-Selvam et al. 2004). Inspired by naturally-occurring domain insertion, researchers have created chimeric enzymes (Kaneko et al. 2004; Q. Wang and Xia 2007). To achieve this, the insertion site needs to be carefully selected in order to obtain correct folding of the chimeric protein (Saadat 2017). We suggest that *Pm*25 could be a convenient template for enzyme engineering work. Using the β3/α3 loop as a docking site, insertional constructions could be created combining GH10 xylanase with other catalytic and/or substrate binding functions. More generally, considering that the β3/α3 loop constitutes a docking site for insertions in GH10 (i.e. the *Pm*25-like enzymes) and also in other TIM barrel enzymes (Nagano, Orengo, and Thornton 2002) it is likely that this loop is an insertional hot spot that has been used several times throughout evolution. This could lead to the developpement of industrially relevant enzymes and would certainly required ingeneering on the GH10 part to improve its catalytic efficiency.

Like expansins and auxiliary activity family 9 proteins, CBMs are recognized as synergistic proteins and enhanced the catalysis by the proximity effect, targeting and disruptive function (Kim et al., 2014). In a recent study, CBM in one cellobiohydrolase was suggested to mediate the synergy with another cellobiohydrolase by its targeting role (Badino et al., 2017). Therefore, it is reasonable to hypothesize that the internal CBMs in *Pm*25 may play a role in orchestrating the synergistic action of PUL enzymes. To better understand this, the product profile of wild type and the CBM domain abrogated *Pm*25 in combination with other PUL enzymes were analysed using beechwood xylan as a substrate. However, under the tested conditions, no striking difference could be observed. In plant biomass based biorefinery, the

substrates are much more complex than beechwood xylan used in this study, which may highlight the potential role of CBMs in mediating synergistic reactions another explanation is that they recognized a very specific motif in the xylan that has not yet been discovered.

How has this thesis contributed to the glycan forgaging study in gut microbiota?

The starch utilization system (SUS) along with polysaccharide utilization loci (PUL) is a paradigm for complex glycan catabolism in Bacteroidetes, one of the dominant gut-associated bacterial phyla (Martens et al., 2009). Bioinformatic study of *Pm*25 showed that its homologes were mostly distributed in putative PULs. Xyn10C (*Pm*25 homologue) is proved to be the one of the most induced/expressed proteins when *Bacteroides* species are growing on xylan (Despres et al., 2016; Zhang et al., 2014), which indicated its important role in xylan degradation. Xyn10C in xylan utilization system is equivalent to SusG as they are the only out membrane glycoside hydrolases in each system (Fig B-6, Bibliography). The combined *in vivo* and *in vitro* experiments have revealed that SusG is important for starch degrading activity and starch binding affinity (Reeves et al., 1997; Shipman et al., 1999, 2000). In terms of activity, SusG is essential for bacterial growth on starch but has very low activity (compared to SusA); regarding the binding affinity, SusG is not important for starch binding but may prevent the degraded starch molecule from diffusing in the environment until imported in the periplasm. Therefore, the primary role of SusG is to degrade the starch on the cell surface. The *in vitro* biochemical characterization of *Pm*25 and its mutants achieved in this thesis strengthened the similarity between Xyn10C and SusG, such as low activity and independent domain function (Koropatkin and Smith, 2010), which helped to deduce the comparable role of Xyn10C *in vivo*, degrading xylan at cell surface and producing XOS that can be used for feeding target species of Bacteroidetes. In this case, Xyn10C, and all other SusG like proteins have potential values in generating useful oligosaccharides, such as prebiotics. Overall, enzyme systems from gut microbiota degrading xylan arouses great interest in nutritionist as knowledge in this area could better understand the gut ecology and personalize human gut health.

What can be done in the future to further address this topic?

The function of the CBMs in *Pm*25 is probably relative to insoluble substrate/complex biomass. It would be interesting to compare the wild type and mutant *Pm*25 in degrading complex biomass with/without other PUL enzymes. In this respect, new techniques are required to compare the medium-sized products as the HPAEC/MALDI-TOF are not practical.

In the PUL, there is an unknown protein (Figure 1A, Article I) (Dodd et al., 2011), whose gene locus is always between SusD protein and *Pm*25. It was predicted to be SusE-like protein. In our PUL, we produced the SusE-like protein and showed that it can bind to beechwood xylan. It would be interesting to further characterize this protein and better understand its role in the PUL.

Until now, no crystal structure of Xyn10C protein is published. Its 3D structure would be very informative to understand its domain organization, pH tolerance, and in designing chimeric proteins. Although several attempt to obtain the crystal structure was unsuccessfull (E. Ioannou's thesis). And it would be interesting to compare the crytal structure of wild type *Pm*25 and the CBM-truncated mutant thereof.

The deletion of the PUL where there is Xyn10C affected the bacterial growth on simple xylan (Rogowski et al., 2015). To further understand the function of Xyn10C, the xyn10C-knock out *Bacteroidale* strain could be constructed and see how it affect the growth on xylan with/without the rescue experiments, the different mutant of *Pm*25 could be used to restore the function and could may be help to understand if the CBM have a critical role in the xylan degradation.

Reference:

Badino, S.F., Christensen, S.J., Kari, J., Windahl, M.S., Hvidt, S., Borch, K., and Westh, P. (2017). Exo-exo synergy between Cel6A and Cel7A from Hypocrea jecorina : role of carbohydrate binding module and the endo-lytic character of the enzymes. Biotechnol. Bioeng. 114, 1639–1647.

Despres, J., Forano, E., Lepercq, P., Comtet-Marre, S., Jubelin, G., Chambon, C., Yeoman, C.J., Berg Miller, M.E., Fields, C.J., Martens, E., et al. (2016). Xylan degradation by the human gut Bacteroides xylanisolvens XB1AT involves two distinct gene clusters that are linked at the transcriptional level. BMC Genomics 17, 326.

Dodd, D., Mackie, R.I., and Cann, I.K.O. (2011). Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. Mol. Microbiol. 79, 292–304.

Kim, I.J., Lee, H.J., Choi, I.G., and Kim, K.H. (2014). Synergistic proteins for the enhanced enzymatic hydrolysis of cellulose by cellulase. Appl. Microbiol. Biotechnol. 98, 8469–8480.

Koropatkin, N.M., and Smith, T.J. (2010). SusG: A Unique Cell-Membrane-Associated alpha-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. Structure 18, 200–215.

Martens, E.C., Koropatkin, N.M., Smith, T.J., and Gordon, J.I. (2009). Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. J. Biol. Chem. 284, 24673–24677.

Mirande, C., Mosoni, P., Béra-Maillet, C., Bernalier-Donadille, A., and Forano, E. (2010). Characterization of Xyn10A, a highly active xylanase from the human gut bacterium Bacteroides xylanisolvens XB1A. Appl. Microbiol. Biotechnol. 87, 2097–2105.

Pell, G., Taylor, E.J., M.Gloster, T., Turkenburg, J.P., M.G.A.Fontes, C., M.A.Ferreira, L., Nagy, T., J.Clark, S., J.Davies, G., J.Gilbert, H., et al. (2004). The Mechanisms by Which Family 10 Glycoside Hydrolases Bind Decorated Substrates. J. Biol. Chem. 279, 9597–9605.

Reeves, A.R., Wang, G.R., and Salyers, A.A. (1997). Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. J. Bacteriol. 179, 643–649.

Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Baslé, A., Morland, C., Day, A.M., Zheng, H., et al. (2015). Glycan complexity dictates microbial resource allocation in the large intestine. Nat. Commun. 6, 1–15.

Shipman, J.A., Berleman, J.E., and Salyers, A.A. (2000). Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. J. Bacteriol. 182, 5365–5372.

Shipman, J. a, Cho, K.H., Siegel, H. a, and Salyers, A. a (1999). Physiological Characterization of SusG , an Outer Membrane Protein Essential for Starch Utilization by Bacteroides thetaiotaomicron. J. Bacteriol. 181, 7206–7211.

Zhang, M., Chekan, J.R., Dodd, D., Hong, P.-Y.P.-Y.P.-Y.P.-Y., Radlinski, L., Revindran, V., Nair, S.K., Mackie, R.I., and Cann, I. (2014). Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. Proc. Natl. Acad. Sci. U. S. A. 111, E3708-17.