THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National des Sciences Appliquées de Toulouse (INSA Toulouse)

Discipline ou spécialité :

Ingénieries Microbienne et Enzymatique

Présentée et soutenue par :

Letian SONG

le : jeudi 21 juillet 2011

Titre :

STUDY AND ENGINEERING OF A GH11 ENDO-BETA-XYLANASE, A BIOMASS-DEGRADING HEMICELLULASE

Etude et Ingénierie d'une Endo-beta-1,4-Xylanase de la Famille GH11, une Hémicellulase Dégradant la Biomasse Cellulosique

---

JURY

Mr Dion Michel-Pierre, Professeur, Université de Nantes
Mr Record Eric,   Chargé de recherche, INRA Marseille
Mr Giardina Thierry, Professeur, Université de Marseille
Mr Marty Alain,   Professeur, INSA-Toulouse
Mme Rémond Caroline, Professeur, URCA Reims

Ecole doctorale :
Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)
Unité de recherche :
Laboratoire d'Ingénierie des Système Biologiques et des Procédés (LISBP)
Directeur(s) de Thèse :
Mr O'Donohue Michael, Directeur de recherche INRA, Toulouse
Mme Bozonnet Sophie, Ingénieur de Recherches, INSA, Toulouse
Rapporteurs :
Mr Dion Michel-Pierre
Mr Record Eric

par

Letian SONG

STUDY AND ENGINEERING OF A GH11 ENDO-BETA-XYLANASE, A BIOMASS-DEGRADING HEMICELLULASE

Etude et Ingénierie d'une Endo-beta-1,4-Xylanase de la Famille GH11, une Hémicellulase Dégradant la Biomasse Cellulosique

Mr Dion Michel-Pierre	Professeur, Université de Nantes
Mr Record Eric	Chargé de recherche, INRA Marseille
Mr Giardina Thierry	Professeur, Université de Marseille
Mr Marty Alain	Professeur, INSA-Toulouse
Mme Rémond Caroline	Maître de conférence, URCA Reims
Mme Bozonnet Sophie	Ingénieur de Recherches, INSA, Toulouse
Mr O'Donohue Michael	Directeur de recherche INRA, Toulouse

NOM : SONG   **Prénom :** Letian

Titre : Etude et ingénierie d'une endo-beta-1,4-xylanase de la famille GH11, une hémicellulase dégradant la biomasse lignocellulosique

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

Année : 2011   **Lieu** : INSA de Toulouse

RESUME

La création de nouvelles enzymes pour l'hydrolyse de la biomasse est une stratégie clé pour le développement du bioraffinage. Dans ce contexte, les xylanases de la famille GH11 sont déjà déployées dans de nombreux procédés industriels et donc bien positionnées pour jouer un rôle important dans ces procédés. La cible de cette étude, la xylanase GH11 (Tx-Xyl) de la bactérie *Thermobacillus xylanilyticus*, est une enzyme thermostable et donc une bonne candidate pour des travaux d'ingénierie visant l'amélioration de son activité sur des substrats ligno-cellulosiques.

Dans cette étude, deux stratégies d'ingénierie des enzymes ont été employées afin d'obtenir de nouvelles informations portants sur les relations structure-fonction au sein de Tx-Xyl. La première stratégie a consisté en l'utilisation d'une approche de mutagenèse aléatoire, couplée à l'emploi de méthodes de recombinaison *in vitro*. Ces travaux avaient pour objectif d'améliorer la capacité hydrolytique de Tx-Xyl sur la paille de blé. La deuxième stratégie mise en œuvre s'est appuyée sur une approche semi-rationnelle visant la création d'une enzyme chimérique, qui bénéficierait d'une amélioration des interactions enzyme-substrat au niveau du sous-site -3.

Le premier résultat majeur de cette thèse concerne le développement d'une méthode de criblage qui permet l'analyse à haut débit de banques de mutants pour la détection de variants qui présentent une activité hydrolytique accrue directement sur paille de blé. A l'aide de ce crible, nous avons pu analyser plusieurs banques de mutants, représentant un total de six générations de mutants, et identifier une série de combinaisons de mutations différentes. D'un côté, un variant, comportant deux mutations silencieuses, permet une meilleure expression de Tx-Xyl, alors que d'autres enzymes mutées présentent des modifications intrinsèques de leurs aptitudes catalytiques. Comparés à l'enzyme parentale Tx-Xyl, certains mutants solubilisent davantage les arabinoxylanes de la paille et, lorsqu'ils sont déployés avec un cocktail de cellulases, participent à une réaction synergique qui permet un accroissement du rendement des pentoses et du glucose libérés.

A l'aide d'une approche semi-rationnelle, une séquence de 17 acides aminés en provenance d'une xylanase GH11 fongique a été ajoutée à l'extrémité N-terminale de Tx-Xyl, afin de créer de nouveaux brins β. L'enzyme chimérique a pu être exprimée avec succès et caractérisée. Néanmoins, l'analyse de ses propriétés catalytiques a révélé que celle-ci ne présente pas davantage d'interactions avec son substrat dans le sous-site -3, mais les résultats obtenus fournissent de nombreux renseignements sur les relations structure-fonction au sein de l'enzyme. De plus, ces travaux nous permettent de postuler que Tx-Xyl posséderait un site de fixation secondaire pour les xylanes, un élement jusqu'ici insoupçonné dans cette enzyme. Par ailleurs, l'analyse de nos résultats nous permet de proposer une explication rationnelle pour l'échec de notre stratégie initiale.

Mots clefs:

Xylanase; GH11*; Thermobacillus xylanilyticus*; biomasse lignocellulosique; paille de blé; bioraffinage de 2ème génération; criblage à haut débit ; ingénierie des enzymes.

Last Name: SONG   **First Name:** Letian

Title: Study and Engineering of a GH11 endo-beta-xylanase, a biomass-degrading hemicellulase.

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

Year : 2011   **Place** : INSA de Toulouse

SUMMARY

Engineering new and powerful enzymes for biomass hydrolysis is one area that will facilitate the future development of biorefining. In this respect, xylanases from family GH11 are already important industrial biocatalysts that can contribute to 2nd generation biorefining. The target of this study, the GH11 xylanase (Tx-Xyl) from *Thermobacillus xylanilyticus* is thermostable, and is thus an interesting target for enzyme engineering, aiming at increasing its specific activity on lignocellulosic biomass, such as wheat straw. Nevertheless, the action of xylanases on complex biomass is not yet well understood, and thus the use of a rational engineering approach is not really feasible.

In this doctoral study, to gain new insight into structure-function relationships, two enzyme engineering strategies have been deployed. The first concerns the development of a random mutagenesis and *in vitro* DNA shuffling approach, which was used in order to improve the hydrolytic potency of Tx-Xyl on wheat straw, while the second strategy consisted in the creation of a chimeric enzyme, with the aim of probing and improving -3 subsite binding, and ultimately improving hydrolytic activity.

The first key results that has been obtained is the development of a novel high-throughput screening method, which was devised in order to reliably pinpoint mutants that can better hydrolyze wheat straw. Using this screening method, several generations of mutant libraries have been analyzed and a series of improved enzyme variants have been identified. One mutant, bearing silent mutations, actually leads to higher gene expression, while others have intrinsically altered catalytic properties. Testing of mutants has shown that some of the enzyme variants can improve the solubilization of wheat straw arabinoxylans and can work in synergy with cellulose cocktails to release both pentose sugars and glucose.

Using a semi-rational approach, 17 amino acids have been added to the N-terminal of Tx-Xyl, with the aim of adding two extra β-strands coming from a GH11 fungal xylanase. A chimeric enzyme has been successfully expressed and purified and its catalytic properties have been investigated. Although this approach has failed to create increased -3 subsite binding, the data presented reveals important information on structure-function relationships and suggest that Tx-Xyl may possess a hitherto unknown secondary substrate binding site. Moreover, a rational explanation for the failure of the original strategy is proposed.

Keywords:

Xylanase; GH11; *Thermobacillus xylanilyticus*; lignocellulosic biomass; wheat straw; second generation biorefining; high-throughput screening; enzyme engineering   .

PUBLICATIONS

(i) Published and accepted

1. **Song, L.**, Laguerre, S., Dumon, C., Bozonnet, S., O'Donohue, M.J. A high-throughput screening system for the evaluation of biomass-hydrolyzing glycoside hydrolases. *Bioresource Technology*, 2010, 101, 8237-8243. (IF 4.253)

2. Dumon, C., **Song, L.**, Bozonnet, S., Fauré, R., O'Donohue, M.J., Progress and future prospects for pentose-specific biocatalysts in biorefining. *Process Biochemistry*, 2011. Accepted (IF 2.444)

(ii) To be submitted

1. **Song, L.**, Bozonnet, S., Siguier, B., Dumon, C., O'Donohue, M.J., Improved mutants of *Thermobacillus xylanilyticus* xylanase for the degradation of recalcitrant wheat straw. In preparation.

2. **Song, L.**, Dumon, C., Siguier, B., Kulminskayad A., André, I., Bozonnet, S., O'Donohue, M.J., Impact of N-terminal extension on the stability and activity of the GH11 xylanase from *Thermobacillus xylanilytics*. In preparation.

COMMUNICATIONS ORALES

1. **New High-Throughput Screening System for Hemicellulases based on Lignocellulosic Biomass**

   <u>Song L.</u>, Dumon C., Bozonnet S., O'Donohue M.J.

   First Workshop of COST Action CM0903 "Utilization of Biomass for Sustainable Fuels and Chemicals (UBIOCHEM)". 13-15 May 2010, Cordoue, Espagne

2. **Stratégies de criblage pour l'identification et l'ingénierie d'hémicellulases.**

   <u>Fauré R.</u>, **Song L.**, Ferreira F., Ragon M., Laguerre S., Bozonnet S., Dumon C., O'Donohue M.J.

   XXIIIèmes Journée du Groupe Français des Glucides (GFG), 17-20 may 2010, Cap Hornu - Baie de Somme, France.

3. **A New Lignocellulose-based Screen Methodology for the Optimization of GH11 Xylanase from** *Thermobacillus xylanilyticus*

   <u>Song L.</u>, O'Donohue M.J., Bozonnet S., Dumon C.

   Second Congress of Biomass Derived Pentoses: from biotechnology to fine chemistry. 14-16 Novembre 2010, Reims, France

COMMUNICATIONS PAR AFFICHES

1. **Development of a HTS method for the selection of xylanase variants displaying improved activity on wheat straw.**

   <u>Song L.</u>, Dumon C., Bozonnet S., O'Donohue M.J.

   Enzymes: Structure/Fonction/Catalyse/Ingénierie/Régulation, congrès annuel de la SFBBM. 27-29 august 2009, Nancy, France

2. **Development of a HTS method for the selection of xylanase variants displaying improved activity on wheat straw.**

   Song L., Dumon C., <u>Bozonnet S.</u>, O'Donohue M.J.

   Enzyme Engineering XX. 20-24 Septembre 2009, Groningen, Pays-Bas

CONTENTS

LIST OF FIGURES

ARTICLE I

ARTICLE II

ARTICLE III

GENERAL CONCLUSION

LIST OF TABLES

ABBREVIATIONS

ε	Molar coefficient variation
μ	Mean value
σ	Standard deviation
AX	Arabinoxylan
A/X ratio	Arabinose to xylose ratio
BWX	Birchwood xylan
CV	Coefficient variation
CFE	Cell-free extracts
DP	Degree of polymerization
Dpl-WS	Xylanase depleted wheat straw
DW	Dry weight
epPCR	Error-prone PCR
H bond	Hydrogen bond
In-WS	Intact (untreated) wheat straw
LVWAX	Low viscosity wheat arabinoxylan
Np-Xyl	*Neocallimastix patriciarum* xylanase
SBS	Secondary binding site
SDM	Site-dierected mutagenesis
StEP	Stagged extension process
Tx-Xyl	*Thermobacillus xylanilyticus* xylanase
Tx-Xyl-AF7	Tx-Xyl variant with mutations of T27C and C516T
Tx-Xyl-NTfus	Tx-Xyl variant with fused N-terminal
XOS	Xylo-oligosaccharides

RESUME EN FRANÇAIS

Etude et ingénierie d'une endo-beta-1,4-xylanase de la famille GH11, une hémicellulase dégradant la biomasse lignocellulosique

Introduction

La biomasse végétale s'impose à l'heure actuelle comme une ressource à privilégier pour la production d'énergies alternatives. A ce titre, de par leur abondance et leur disponibilité et puisqu'ils ne rentrent pas dans la chaine alimentaire, les coproduits lignocellulosiques issus de l'agriculture (pailles, rafles, etc…) sont ainsi considérés comme des matières premières idéales pour les futures bio-raffineries industrielles.

Parmi les 4 étapes unitaires qui interviennent dans tout procédé de bio-raffinerie (prétraitement, saccharification, fermentation et purification), le pré-traitement demeure à l'heure actuelle une étape critique. C'est là que la matière végétale subit une désorganisation massive de ces constituants (cellulose, hémicellulose et lignines), rendant les polysaccharides alors disponibles pour les enzymes hydrolytiques utilisées dans l'étape de saccharification. Le pré-traitement est une étape couteuse, mais indispensable, qui repose le plus souvent sur des procédés physico-chimiques couteux en énergie ou en réactif. Supprimer totalement cette étape est peu réaliste, mais réduire son cout en diminuant son impact est envisageable, à condition d'employer en aval, des enzymes adéquates, aptes à agir sur une matière première récalcitrante.

Depuis les 20 dernières années, l'ingénierie enzymatique a très largement profité des progrès faits dans le domaine de l'évolution moléculaire dirigée, qui permet de re-créér à l'échelle du laboratoire, les processus mis en place par la Nature pour améliorer une protéine. Très largement employée dans la R&D industrielle, ces techniques offrent des opportunités immenses, et présentent l'avantage majeur de ne pas nécessiter au préalable des connaissances approfondies des relations structure-fonction de la protéine à améliorer, mais simplement un crible puissant capable d'identifier parmi une population d'enzymes mutées des variants qui auraient acquis la propriété recherchée. De ce point de vue, les techniques d'évolution moléculaire dirigée sont particulièrement pertinentes pour

l'amélioration des propriétés d'enzymes destinées à une utilisation en bioraffinerie pour des opérations qui sont encore à caractériser finement.

Jusqu'ici les cellulases ont été la cible de la plupart des travaux de recherche dédiés à la création de nouvelles enzymes pour les procédés de bioraffinage, et aujourd'hui, la production de cocktails cellulolytiques performants devient une réalité, bien que la faible activité des cellulases sur la cellulose cristalline soit toujours une source de problèmes. En effet, malgré d'énormes efforts et 15 années de recherche, l'activité des cellulases n'a progressé que de 50%. Comparées aux cellulases, les xylanases ont fait l'objet de beaucoup moins d'études, alors qu'elles ont un rôle majeur dans les cocktails cellulolytiques, qui souffrent, entre autres, de l'inhibition des cellulases par les xylo-oligosaccharides. En conséquence, il est sans doute désormais pertinent de s'intéresser davantage aux xylanases et de réaliser des travaux d'ingénierie visant l'amélioration de leur action sur des substrats lignocellulosiques.

La xylanase de *Thermobacillus xylanilyticus* (Tx-Xyl) est une endo-β-(1→4) xylanase qui possède la capacité de dépolymériser les arabinoxylanes présents dans les parois secondaires de plantes d'origines botaniques diverses. Tx-Xyl est thermostable - son activité maximale est mesurée à 75°C- et fonctionne sur une plage de pH relativement large et, en particulier aux pHs alcalins. Le gène de la xylanase a été cloné et produit chez *Escherichia coli* et sa structure 3D a été résolue. Comme toutes les enzymes de la famille 11 des glycosides-hydrolases (GH11, base de données CAZy), Tx-Xyl possède un domaine catalytique en forme de β-jelly-roll. La topographie de ces enzymes est souvent assimilée à une main droite partiellement recroquevillée sur elle-même. La longue boucle entre les brins B7 et B8 -qui constitue le « pouce » de l'enzyme- a été notamment étudiée pour son rôle dans la fixation du substrat dans le site actif.

L'action de Tx-Xyl a été déjà caractérisée sur plusieurs substrats, dont la paille et le son de blé. Cette enzyme possède une forte activité sur xylane de bouleau, et solubilise 40 à 50% des arabinoxylanes du son de blé. Par contre, sur la paille de blé, Tx-Xyl s'avère nettement moins efficace, sans doute parce que cette biomasse présente une structure complexe, composée à la fois d'arabinoxylanes, de cellulose et de lignines. Cependant, les facteurs qui

gouvernent et limitent l'action de Tx-Xyl sur la paille de blé sont actuellement peu connus et il est donc impossible d'envisager des travaux d'ingénierie rationnelle des enzymes pour l'amélioration de telle ou telle propriété.

Cette étude doctorale porte donc sur les xylanases de la famille GH11 et leur utilisation comme outils de déstructuration de la biomasse. Elle a pour objectif majeur de contribuer significativement à leur optimisation en vue de leur futur emploi dans des procédés de bioraffinage et, par ailleurs, de fournir de nouvelles informations sur les relations structure-fonction qui régissent les différentes propriétés des enzymes de la famille GH11, et notamment leurs propriétés catalytiques.

Lors d'expériences d'évolution in vitro, si la méthode de criblage ne reflète pas fidèlement la réaction étudiée, il devient difficile d'isoler des mutants vraiment intéressants. Par conséquent, le premier défi de ce travail doctoral était de concevoir et mettre en œuvre une méthode de criblage à haut débit suffisamment robuste pour permettre la détection d'enzymes mutées présentant une activité hydrolytique accrue directement sur la paille de blé.

La deuxième partie de ce travail porte sur l'utilisation de la méthode de criblage pour l'analyse de plusieurs générations successives d'enzymes mutées. Dans cette phase, il a été nécessaire d'adapter à certains moments la procédure expérimentale, afin de tenir compte de nouvelles contraintes qui se sont révélées au fur et à mesure de la réalisation des travaux, sans pour autant perdre de vue l'objectif principal qui était l'isolement de xylanases performantes.

Enfin, le troisième volet de cette étude porte sur l'emploi d'une approche semi-rationnelle pour l'étude des relations structure-fonction impliquées dans l'interaction de la xylanase, Tx-Xyl, avec son substrat. Des travaux précédents ont permis l'hypothèse qu'un site actif plus étendu chez Tx-Xyl, comportant un sous site -3, conduirait à un accroissement de l'activité hydrolytique. En conséquence, l'ambition des travaux réalisés dans le cadre de cette thèse était d'étendre le site actif de Tx-Xyl dans sa région glycone et de caractériser l'enzyme chimérique produite.

A- Développement d'un criblage à haut-débit adapté à l'évaluation de glycoside-hydrolases sur substrats complexes

Avant de s'attaquer aux tests en microplaques, nous avons d'abord vérifié que l'absence d'agitation n'affecte pas l'hydrolyse de la biomasse insoluble. Toutefois, celle-ci reste dépendante, en début de réaction (jusqu'à 8 heures), de la quantité de catalyseur introduite au départ.

Pour pouvoir être utilisée en format microplaque, la paille de blé a d'abord été broyée finement afin d'obtenir une poudre fine (taille des particules < 0.5 mm). Après lavage, séchage et autoclavage, nous avons retenu l'utilisation d'un « multiscreen column loader » de la société Millipore pour repartir cette poudre de manière homogène et reproductible dans les 96 puits d'une microplaque (Fig. A).
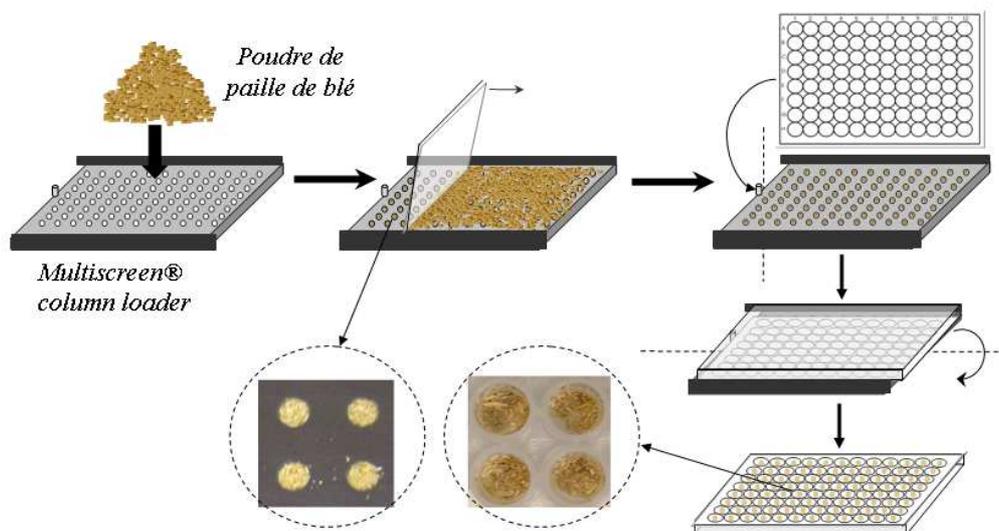
Fig. A : Distribution de la paille de blé broyée en microplaque 96 puits, à l'aide du Multiscreen column loader (Millipore).

L'homogénéité du transfert de matière a été confirmée par la faible dispersion mesurée entre les masses de microplaques chargées (2,27%) mais également par la mesure indirecte de l'activité d'une solution de xylanase uniformément répartie dans la microplaque, qui nous indique une répartition de biomasse variant d'environ 7 à 11% entre les puits.

Toutes les étapes du criblage ont ensuite été optimisées indépendamment : méthode de dosage des sucres réducteurs pour mesurer l'hydrolyse du substrat ; conditions d'incubation à 60°C ; solutions pour limiter l'évaporation sur une durée de 4 heures d'incubation. A chaque fois, notre souci était de minimiser l'erreur associée à chaque manipulation. Le mode opératoire optimal qui permet de mesurer de manière reproductible une activité xylanase produite est ainsi résumé dans la Fig. B.
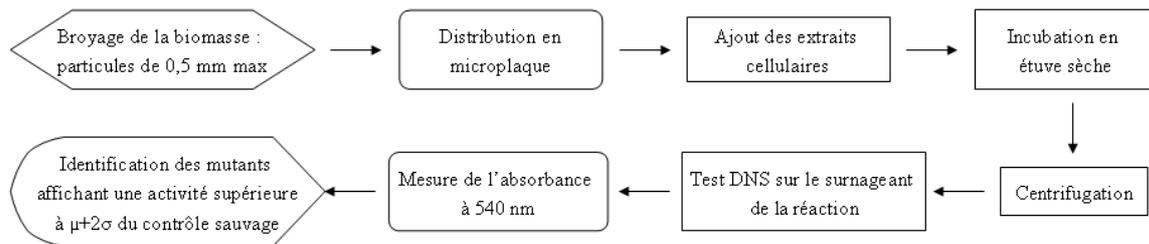
Fig. B : Diagramme des opérations intervenant dans le criblage à haut-débit sur paille de blé.

Après l'optimisation des variables physico-chimiques, ce sont les facteurs biologiques qui ont été étudiés un par un : souche de production, vecteur d'expression, conditions de culture (milieu, température et induction) ont été évalué afin de trouver les meilleures conditions de production d'enzymes en microplaques.

En cumulant les erreurs associées à chaque étape, cette méthode de criblage sur substrat complexe présente une déviation inférieure à 10% losqu'elle est testée sur la souche sauvage. Seulement 3% des mesures se retrouvent au-delà de la limite de $\mu+2\sigma$, en accord donc avec les principes dune distribution normale (5% attendus dans le cadre d'une loi normale).

Pour valider cette stratégie de criblage à haut-débit, une première série de 264 mutants obtenus après mutagénèse aléatoire (error-prone PCR, epPCR) a été soumise au crible. L'activité hydrolytique moyenne calculée sur chaque microplaque présente une variabilité faible (3,2%) confirmant la robustesse du crible. A l'issu du criblage sur paille de blé, les clones présentant une activité supérieure de 2 écarts-type à la moyenne des activités des clones contrôles ont été sélectionnés. L'étude du mutant Tx-Xyl-AF7 a montré que ce clone présentait une activité supérieure de 74.5 % à celle de la souche sauvage. Son séquençage a

indiqué la présence de 2 mutations silencieuses : cultivé en plus grand volume, ce mutant produit en fait environ 2 fois plus de xylanase que la souche sauvage. L'analyse attentive des mutations révèle que l'une d'entre elles (T27C) est située près du site de fixation du ribosome, une zone critique en termes d'efficacité de l'expression génique. C'est la réorganisation plus favorable de cette structure secondaire de l'ARNm correspondant qui serait à l'origine de l'augmentation de l'expression du gène muté.

B- Evolution moléculaire dirigée de la xylanase de *Thermobacillus xylanilyticus* pour l'obtention de mutants optimisés pour la dégradation de paille de blé récalcitrante

Le clone Tx-Xyl-AF7, isolé à l'issue de la validation de la méthode de criblage, présente une d'ores et déjà amélioration de son potentiel, du fait de sa meilleure capacité à produire l'enzyme recombinante. C'est donc ce gène qui a servi de matrice pour le second tour de mutagénèse aléatoire. La stratégie globale que nous avons suivi, s'est appuyée au total sur 3 tours successifs de mutagénèse aléatoire (PCR à erreurs, ou error-prone PCR, epPCR) suivis de différentes méthodes de recombinaison. La suite et le contenu de ces expériences sont résumés dans le Tableau. A.

Sur l'ensemble des 2 premiers tours, c'est un total de 4597 clones qui a été criblé sur paille de blé broyée, permettant d'obtenir une sélection de 5 mono-mutants présentant une activité hydrolytique significativement supérieure par rapport à l'enzyme sauvage.

Ce lot de mutants a ensuite été re-soumis à un dernier tour de mutagénèse à erreur, mais cette fois, le criblage a été adapté pour être effectué sur paille de blé déplétée, c'est-à-dire ayant été soumis à une complète hydrolyse par la xylanase Tx-Xyl. La biomasse résultante ne contient plus que la part d'arabinoxylanes inaccessibles ou résistants à l'action de l'enzyme sauvage.

Au terme de ces 3 tours de mutagénèse, le taux de mutation moyen dans les clones de la dernière banque tourne autour de 7,2 substitution par kilobase, avec près de 2 % des clones présentant une activité supérieure de plus de 4 écart-types de la moyenne du clone parental.

Tableau. A : Succession des différents tours de mutagénèse aléatoire et de recombinaison.

Génération	Type de banque	Substrat de criblage	N° de mutants criblés	CV of wild-type control †	% de clones présentant une activité supérieure à la moyenne du sauvage de :					N° de hits sélectionnés
					>4CV	>5CV	>6CV	>7CV	>8CV	
1ere	epPCR sur gène sauvage	In-WS	264	11.1±1.3%	0.4%	0.4%	-	-	-	1
2nde	epPCR sur mutant Tx-Xyl-AF7	In-WS	4333	18.1±5.4%	0.1%	-	-	-	-	4
3ème	Mutagenèse dirigée	-	-	-						11
4ème	epPCR sur 11 gènes sélectionnés	Dpl-WS	4300	10.9±2.2%	1.2%	0.6%	-	-	-	30
5ème	Recombinaison	Dpl-WS	3840	8.1±0.6%	1.4%	6.0%	2.1%	0.8%	0.1%	7
6ème	Recombinaison	Dpl-WS	864	10.2%	9.3%	2.8%	0.9%	0.2%	-	8
7ème	Recombinaison	Dpl-WS	864	11.3%	19.5%	7.5%	2.4%	0.5%	0.2%	7

† CV = coefficient de variation. Pour le premier tour, le contrôle est la Tx-Xyl, alors que pour les tours suivants, nous avons utilisé le mutant Tx-Xyl-AF7.

Afin de valoriser plus efficacement les combinaisons fonctionnelles obtenues par mutagénèse, les tours suivant d'évolution se sont appuyés sur la recombinaison de mutations obtenues dans les meilleurs clones issus du tour précédent. Ainsi, la technique StEP (pour Staggered Extension Process) a été utilisée du 5ème au 7ème tour d'évolution. Plus de 5560 mutants ont été soumis à la méthode de criblage sur paille de blé déplétée, et à chaque tour, comme illustré sur la Fig. C, les recombinaisons ont permis d'optimiser graduellement l'enzyme pour le substrat considéré.

Le séquençage des hits sélectionnés au terme du 7ème tour d'évolution a montré que la majorité des mutations mises en évidence étaient déjà des mutations observées lors des tours précédents, indiquant que le processus d'évolution de Tx-Xyl sur paille déplétée atteignait son terme.

L'analyse fine des mutations observées dans les meilleurs clones a permis d'identifier plusieurs changements, dont la fréquence était importante : ainsi, la mutation Y111H se retrouve dans tous les hits sélectionnés, et la fréquence de répétition des mutations Y6H et S27T a été augmentée entre le 5ème et le 7ème tour de mutagénèse.
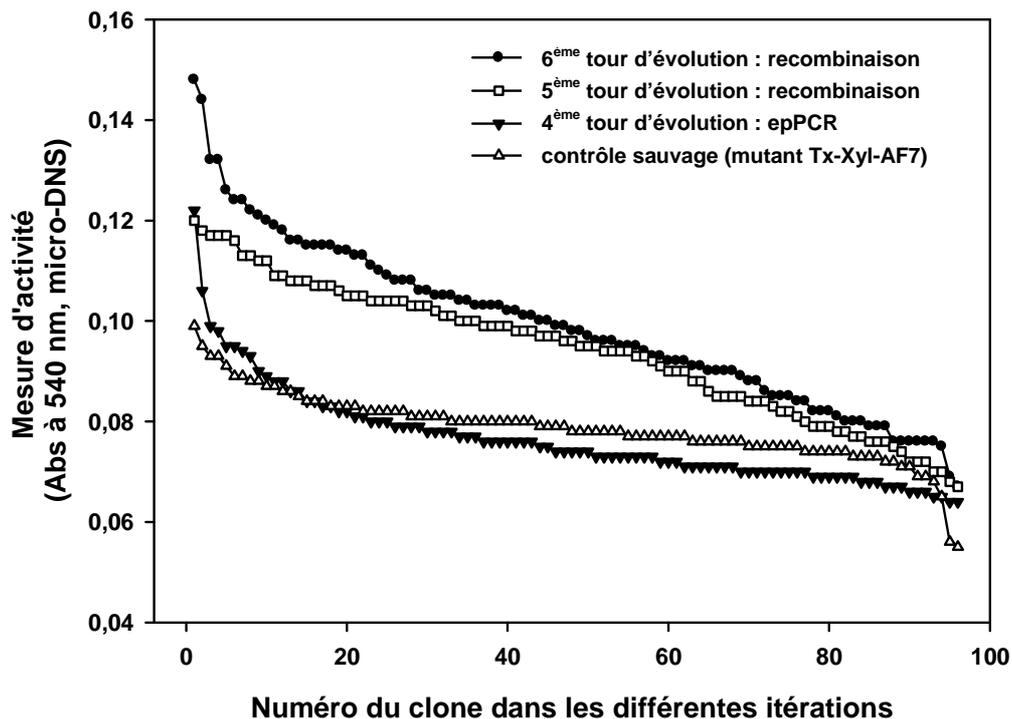
Fig. C : Accélération de l'évolution de Tx-Xyl soumise à des tours successifs de mutagénèse et de recombinaison, suivie de criblages sur paille de blé déplétée. Le même lot de paille de blé déplétée a été utilisé pour les 4 expériences.

Lors du 2ème cycle de mutagénèse aléatoire, 2 clones ont été sélectionnés, qui avaient bénéficié d'une mutation ponctuelle du résidu Tyrosine en position 3. Cette position ainsi que la Tyrosine 111 –systématiquement mutée dans les meilleurs clones- ont donc été soumises à de la mutagénèse à saturation. Les 3 meilleurs clones provenant de la librairie Y3N présentent tous la même mutation –Y3W- alors que le criblage de la banque Y111N a permis d'isoler 3 mutants présentant un acide aminé hydroxylé, sous la forme de 2 mutants serine et d'un mutant thréonine.

L'analyse détaillée de l'activité des meilleurs mutants a été menée à la fois sur substrats solubles, xylane de bouleau (non ramifié) et arabino-xylane de blé (arabinose/xylose = 0.54), pour lesquels les paramètres cinétiques d'hydrolyse ont été déterminés, mais également sur les substrats complexes utilisés lors du criblage, seuls ou en combinaison avec un cocktail enzymatique commercial, l'Accelerase 1500.

Sur xylane de bouleau et arabinoxylane de blé, le mutant S27T affiche une augmentation de son efficacité catalytique de 55 et 20% respectivement, et hydrolyse 2,3 fois plus de xylane dans la paille de blé déplétée que l'enzyme sauvage. Les effets de la mutation Y111T sont plus modestes, mais plus discriminants entre les substrats testés. Ainsi, sa constante catalytique sur xylane de bouleau est augmentée de 48%, mais reste identique à la valeur du sauvage pour les arabinoxylanes de blé. Tout comme S27T, ce mutant arrive à hydrolyser 2 fois plus de substrats dans la paille déplétée que Tx-Xyl.

Si l'effet synergétique attendu entre la xylanase sauvage et le cocktail commercial a bien été vérifié, la substitution de Tx-Xyl par différents mutants (Y111S et Y111T) a permis d'améliorer significativement cette performance.

C- Impact of an N-terminal extension on the stability and activity of the GH11 xylanase from *Thermobacillus xylanilyticus*

La xylanase produite par *Neocallimastix patriciarum* (Np-Xyl) appartient tout comme TxXyl à la famille 11 des glycosides-hydrolases, et présente une activité plus élevée que la plupart des xylanases apparentées. Ceci a été en partie attribué à la présence d'un site catalytique étendu, possédant 6 sous-sites (-3 à +3). L'existence d'un sous-site supplémentaire (-3) par rapport à Tx-Xyl, localisé au niveau d'un feuillet β additionnel situé en N-term, pourrait justifier ces résultats, d'autant plus que l'on sait que cette extrémité joue un rôle important dans la liaison au substrat et potentiellement dans la thermostabilité de l'enzyme. Ainsi, dans la famille GH11, alors que Np-Xyl se distingue par une extrémité N-terminale allongée, Tx-Xyl tout au contraire se caractérise par une extrémité plus courte (Fig. D). Par conséquent, sur la base de ces observations et d'une forte homologie entre les 2 enzymes, nous avons

donc décidé de fusionner les 2 brins β qui constitue ce feuillet additionnel sur l'extrémité N-terminale de Tx-Xyl.
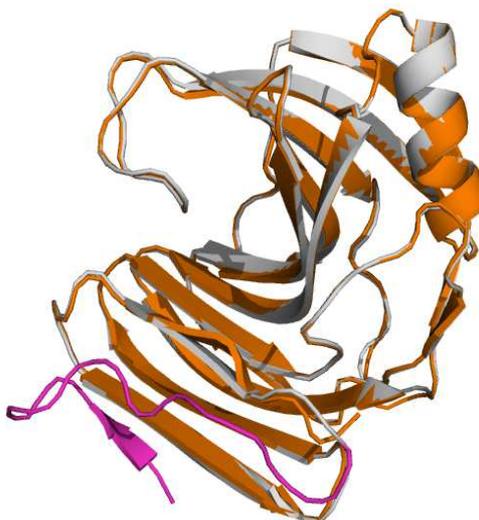
Fig. D : Superposition des structures de Tx-Xyl (orange) et Np-Xyl (gris). Les 2 feuillets β supplémentaires de Np-Xyl sont indiqués en rose.

C'est par PCR, en utilisant une amorce en 5' longue de 81 pb et contenant la séquence codante des 17 acides aminés supplémentaires à introduire, qu'a été obtenu le gène de la protéine hybride, nommée Tx-Xyl-NTfus. Après clonage et surexpression dans *E. coli*, cette nouvelle protéine a été produite sous forme soluble, et purifiée selon le même protocole employé pour Tx-Xyl.

En utilisant du xylane de bouleau comme substrat de référence, l'activité spécifique de Tx-Xyl-NTfus correspond à environ 80% de celle de Tx-Xyl, et son activité est stable sur une gamme de pH large (5,0 à 7,5), tout comme sa version raccourcie (Tableau. B).

Tableau. B : Caractéristiques enzymatiques des protéines Tx-Xyl et Tx-Xyl-NTfus.

Enzymes	ε $(M^{-1} cm^{-1})$	*Mw* (kDa)	AS $(U\ mg^{-1})$	pH_{opt}	T_{opt} (°C)	T_m (°C)	60°C $t_{1/2}$ (heure)	70°C $t_{1/2}$ (heure)
Tx-Xyl	102790	20,65	1450	5,8 – 6,0	~ 67	75,9	5,4	0,32
Tx-Xyl-NTfus	102790	22,30	1127	~ 6,2	~ 67	70,9	4,1	0,16

ε: coefficient d'extinction molaire; AS: activité spécific at 60°C; T_m: température de melting; $t_{1/2}$: demi-vie

Même si la protéine de fusion conserve une température optimale d'activité relativement élevée (67°C), son profil d'inactivation au delà de 70°C est plus marqué que pour Tx-Xyl. C'est au niveau de sa thermostabilité que la protéine hybride se distingue le plus de son parent le plus proche : même si les 2 protéines présentent une bonne stabilité à 50°C, conservant près de 100% de leur activité originelle au bout de 6h d'incubation, à 60 et 70°C, Tx-Xyl-NTfus affiche une demi-vie réduite de 25 à 50% respectivement par rapport à Tx-Xyl (Tableau. B).

En utilisant des xylanes solubles différant par leur taux de ramification par des résidus arabinose, les paramètres cinétiques de la protéine hybride ont pu être comparés à ceux des ses parents. Alors que sur substrat non ramifié (xylane de bouleau), Tx-Xyl et Tx-Xyl-NTfus présentent des paramètres cinétiques comparables, la constante catalytique de la protéine de fusion sur arabino-xylane de blé –qui compte un résidu arabinose pour 2 résidus xyloses- est augmentée d'environ 14,4 %.

Afin de vérifier si un sous-site supplémentaire a bien été créé lors de la fusion des 2 brins beta additionnels, des cinétiques d'hydrolyse de xylo-oligosaccharides synthétiques –oNP-X$_2$ et oNP-X$_3$ ont été réalisées. En plus de nous donner accès aux paramètres cinétiques des enzymes sur ces substrats, la méthodologie employée nous permet également de calculer l'énergie de liaison propre aux sous-site -3. Ainsi, même si l'activité de Tx-Xyl-NTfus est légèrement plus faible sur les 2 oligosaccharides comparé à Tx-Xyl, les calculs révèlent, dans les 2 protéines, la présence d'un sous-site -3, de force équivalente et relativement faible.

Lors des tests d'activité sur paille de blé brute, les enzymes testées ont été utilisées sur la biomasse broyée intacte mais également déplétée, seules ou en combinaison avec un cocktail cellulolytique commercial, l'Accelerase 1500. Les analyses montrent que dans tous les cas, Tx-Xyl-NTfus solubilise plus de xylose que Tx-Xyl, avec une augmentation pouvant atteindre 30% de xylose supplémentaire produit sur paille de blé déplétée.

Conclusions

La découverte et l'ingénierie d'enzymes qui dégradent la biomasse végétale, à l'aide de technologies telles que l'approche métagénomique ou l'évolution moléculaire in vitro, ne se font pas sans difficultés, car la biomasse est un substrat d'une complexité inouïe qui est en conséquence extrêmement difficile à employer dans des méthodes de criblage traditionnelles. Néanmoins, parce que ce type de stratégie obéit obligatoirement à la maxime « on ne trouve que ce que l'on recherche », l'emploi de la biomasse dans le criblage parait inévitable si l'on vise la détection efficace de mutants pertinents par rapport aux objectifs fixés. Les travaux développés durant cette thèse auront donc permis dans un premier temps de mettre au point une méthode robuste et efficace de criblage à haut-débit, spécifiquement adaptée à l'hydrolyse de la matière lignocellulosique brute. L'optimisation rigoureuse de cette méthode a permis d'obtenir un taux d'erreur à la fois faible (environ 15%) et maitrisée, et d'atteindre un débit d'environ 2000 clones par semaine. Peu de travaux similaires ont été publiés, et notre approche se caractérise par la simplicité des outils mis en œuvre, notamment au niveau du système de distribution de la biomasse dans les microplaques. **En outre, notre méthode est totalement versatile** et pourra être adaptée à la fois à d'autres biomasse -dès lors que celle-ci se présentera sous la forme de poudre- mais également pour le criblage d'autres cellulases, hémicellulases et cocktails enzymatiques performants.

Pour développer des xylanases capables de solubiliser une plus grande partie des arabinoxylanes de la paille de blé, nous avons adopté une approche combinatoire qui associe la mutagenèse aléatoire, l'évolution moléculaire in vitro et la mutagenèse à saturation. Pour l'exécution de la phase initiale de ce projet, nous avons employé comme substrat une poudre de paille de blé intacte (In-WS), puis, à partir du 4ème tour d'évolution, nous avons utilisé une paille blé préalablement traitée avec le Tx-Xyl, désignée Dpl-WS, afin de favoriser la détection d'enzymes qui dépassent le point final d'hydrolyse.

L'analyse des données obtenues à chaque phase du criblage a révélé que **la performance globale des banques de mutants augmentait avec chaque génération successive**, ce qui est cohérent avec l'effet escompté de l'évolution moléculaire in vitro. A l'issu de six générations de mutagenèse, puis de recombinaison homologue, neuf mutants impliquant des substitutions des acides aminés 3 (Y3W), 6 (Y6H), 27 (S27T) and 111 (Y111H, Y111S, Y111T)

ont été retenus pour des études approfondies (Fig. E). Comparés à Tx-Xyl, toutes ces enzymes mutées présentent une activité hydrolytique sur paille de blé (Dpl-WS) accrue, ce qui valide le mode de criblage. Nous pouvons donc conclure que la stratégie adoptée a bien été adaptée à l'objectif recherché, même si **l'augmentation de la performance des meilleurs mutants reste modeste en termes absolus, de l'ordre de 20 % maximum pour le meilleur variant.**
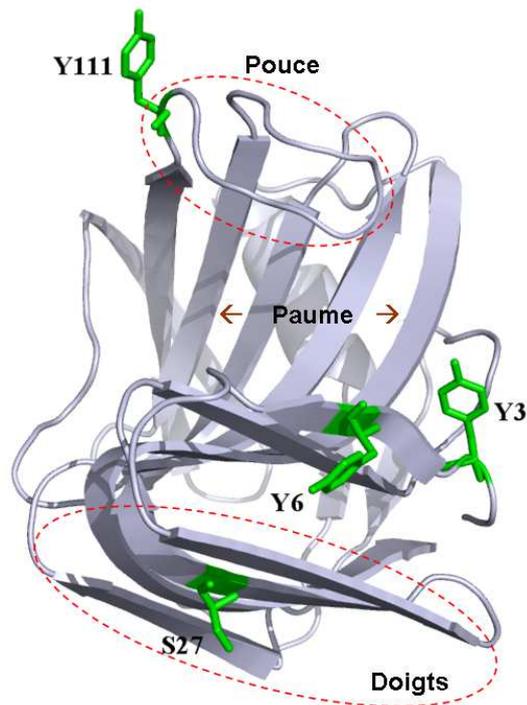
Fig. E : Positions des 4 acides-aminés les plus fréquemment mutés dans la structure tertiaire de Tx-Xyl : Y3, Y6, S27 and Y111. Tx-Xyl présente un repliement typique an β-jelly roll, souvent comparé à une main droite partiellement repliée.

Les raisons de ces résultats modestes sont sans doute liées à la complexité du substrat. Face à la paille de blé, Tx-Xyl est confrontée à une multitude de défis : la complexité structurale de la biomasse, l'accessibilité du substrat, la complexité chimique et structurale des arabinoxylanes eux-même, ainsi que la présence de multiples sources d'inhibition. La limitation liée à l'accessibilité est particulièrement bien mise en évidence lorsque Tx-Xyl ou l'une des enzymes mutées, est déployée avec un cocktail de cellulases. Dans ce cas, l'action des xylanases se trouve augmentée, grâce à l'action des cellulases et vice versa. Ce résultat

indique que la dégradation de la biomasse consiste en l'épluchage progressif des couches d'hémicelluloses et de cellulose, qui alternent et s'entremêlent au sein des parois végétales.

Les modifications d'acides aminés observées lors des différents cycles d'ingénierie combinatoire nous permettent néanmoins **d'obtenir de nouvelles informations sur les relations structure-fonction de la xylanase Tx-Xyl.**

Ainsi, alors que plusieurs études sur des xylanases apparentées indiquent la présence d'un site secondaire de fixation du xylane (secondary binding site, SBS), situé en surface de la protéine, aucune preuve expérimentale de la présence d'un tel site n'existait pour Tx-Xyl, malgré une identité globale élevée entre les différentes séquences (>70%), et même renforcée au niveau des déterminants du SBS (81%). Dans ce travail, nous avons généré le mutant S27T dont l'activité hydrolytique renforcée constituerait **peut être la première preuve expérimentale de l'existence d'un SBS chez Tx-Xyl.** En principe, l'acide aminé 27 se trouverait dans la trajectoire du SBS, et le changement d'une sérine en thréonine aurait pour effet d'augmenter l'hydrophobicité locale, ce qui pourrait être bénéfique pour la fixation de xylanes.

Concernant la thermostabilité de Tx-Xyl, les résultats obtenus dans cette étude apportent **de nouveaux éléments qui tendent à confirmer une théorie qui relie la thermostabilité de Tx-Xyl à la présence de plusieurs résidus aromatiques exposés à la surface de l'enzyme**. Selon cette hypothèse, un certain nombre de ces résidus seraient impliqués dans des interactions intermoléculaires et seraient à l'origine de la formation d'oligomères, qui constitueraient la forme stable de Tx-Xyl. Et parmi ces résidus, figurent les acides aminés 6 et 111, qui ont fait l'objet de mutations dans cette étude. Conformément aux attentes, les mutations Y6H et Y111H ont conduit chacune à une baisse de la thermostabilité.

De manière intéressante, le **mutant S27T** s'est montré plus thermostable que l'enzyme parentale, ce qui est en parfaite adéquation avec une théorie selon laquelle **la thermostabilité s'améliore avec l'augmentation du ratio Thr/Ser à la surface des enzymes.** C'est cette même corrélation qui pourrait également expliquer les observations faites à l'égard du mutant Y111T, qui présente une thermostabilité similaire à celle de l'enzyme

parentale, en dépit du probable effet négatif associé à la perte d'une interaction intermoléculaire hydrophobe de surface.

La mutation du résidu Tyr111 est particulièrement intéressante, car ce résidu est situé sur le pouce, un élément mobile qui est impliqué dans la fixation du substrat et dans le déroulement du cycle catalytique. Localisé à la base du pouce, le résidu 111 déterminerait en partie la mobilité de cette boucle, même si le mécanisme du mouvement de cet élément n'est pas totalement élucidé. Dans cette étude, les mutations introduites à la position 111 (Y111H, Y111T, Y111S) réduisent toutes l'occupation spatiale locale et conduisent à une augmentation de l'activité hydrolytique, probablement liée un mouvement plus fluide du pouce et, donc, à une cadence catalytique plus élevée. Il est aussi intéressant de noter que la perte d'une chaine latérale aromatique à cette position pourrait diminuer l'inhibition de Tx-Xyl par des inhibiteurs hydrophobes, tels que les acides phénoliques (présents dans les parois végétales) et les lignines,

Dans le but d'investiguer et/ou de créer un sous-site -3 chez Tx-Xyl, nous avons prolongé l'extrémité N-terminale de l'enzyme, en ajoutant une séquence de 17 acides aminés en provenance d'une xylanase GH11 fongique (Np-Xyl de *N. patriciarum*), ce qui représente l'addition de deux nouveaux brins β. Etonnamment, nous avons rapidement réussi à exprimé et purifié l'enzyme chimère sous une forme soluble et stable. Néanmoins, l'analyse de l'enzyme chimère (Tx-Xyl-NTfus) et sa comparaison avec Tx-Xyl, a montré que cette enzyme ne présente pas davantage d'interactions avec le substrat dans le sous-site -3. Nos analyses on tout de même procuré **la première preuve expérimentale de l'existence au sein de Tx-Xyl d'un sous-site -3, même si celui-ci est caractérisé par une interaction faible**. A la lumière de nos résultats expérimentaux, nous avons réalisé a posteriori une étude de modélisation qui révèle que parmi les 3 acides aminés impliqués dans le sous-site -3 chez Np-Xyl -Gln11, Ile151 et Tyr193- deux résidus (Gln11 et Ile132) sont correctement positionnés pour former un sous-site -3 dans notre protéine de fusion. Seule la Tyr193 n'est pas conservée : l'acide aminé correspondant -Ser178- pourrait donc être muté afin reproduire la configuration constatée chez Np-Xyl, et finir de reconstituer l'environnement nécessaire à l'établissement d'un sous-site -3 additionnel.

GENERAL INTRODUCTION

Research Background

To move our oil-based economy towards a more sustainable bio-based one, considerable efforts are now being deployed to harness biomass as both a source of energy and carbon. In this respect, lignocellulosic co-products of agriculture (e.g. straws, cobs etc.) are considered to be ideal feedstocks for the future biorefinery industry, because these are abundant, available and do not constitute a threat for the food chain.

Biorefinery processes that convert lignocellulosic biomass into biofuels and biochemicals are often termed second generation processes, to distinguish them from first generation biorefining, which relies on sugar and starch-based feedstocks. A typical second generation biorefinery is composed of four major operational units, which are pretreatment, saccharification, fermentation and product separation/purification. According to the current state of the art, all of these operations are necessary, but pretreatment is absolutely critical, because it largely determines the efficiency of the cell wall-degrading enzymes (i.e. cellulases and hemicellulases) that are used in the saccharification step. Indeed, in the second step, partially disrupted lignocellulosic biomass is subjected to enzyme treatment with the goal of procuring fermentable sugars.

According to Charles Wyman, a prominent researcher in the field of biorefining, the only biorefinery concept that would be more expensive than one with a pretreatment would be one without pretreatment. In other words, pretreatment is costly, but currently indispensable (Wyman, 2007). This is because state of the art enzymes are unable to efficiently deconstruct raw lignocellulosic biomass, which is a chemically and structurally complex substrate. Nevertheless, if it were possible to dispense with pretreatment, or even use less energy and chemically intensive pretreatments, it would be possible to reduce cost and thus accelerate the development of second generation biorefining.

Over the last 15 to 20 years enzyme engineering has entered a new phase, with the introduction of powerful techniques such as *in vitro* enzyme evolution. However, despite major progress, enzyme engineering is still in its infancy, and there remains considerable

scope for progress, since the ultimate goal of being able to predict changes, perform rational engineering and create *de novo* enzyme activities is still a long way off. Nevertheless, the possibilities are enormous and industrial areas such as biorefining will almost certainly be prime targets for future developments.

So far, cellulases have been the major focus of R&D, because the reduction of production costs is a priority. Today, large-scale commercial production of cellulases is becoming a reality (Clarke, 2010; Wilson, 2009), although the limited activity of cellulases on crystalline cellulose is still a cause for concern. This is illustrated by the fact that cellulase engineering has so far only managed to increase activity by 1.5-fold despite 15 years of research efforts (Wilson, 2009).

Compared to cellulases, hemicellulases have been the focus of much less attention, despite the fact that hemicellulases are important in biorefining. Indeed, their vital role in cellulase cocktails has recently been highlighted in several publications (Gao et al., 2010; Kumar and Wyman, 2009b; Prior and Day, 2008) that have shown that xylanases and xylosidases are necessary to alleviate inhibition of cellulases by xylo-oligosaccharides, increase cellulase activity via synergistic interactions and thus reduce process costs (Merino and Cherry, 2007; Murashima et al., 2003). Therefore, it appears timely to address the improvement of hemicellulases, in particular xylanases, using state of the art enzyme engineering and experience gained from cellulase engineering.

Key concepts and Research Strategy

The key contextual elements presented above are at the heart of the research that has been tackled in the doctoral studies described in this dissertation. These are: (i) second generation biorefining is today's challenge and tomorrow's industry; (ii) new biorefinery concepts that use less or no pretreatment will accelerate the implementation of biorefining; (iii) hemicelluloses and hemicellulases are key elements in the biorefining story; and (iv) it is important and timely to engineer new xylanases for biorefining. In addition to these grand challenges, another goal of the doctoral project was to open new avenues of investigation, notably in terms of structure-function relationships, and especially with regard to the

complex processes that govern the action of cell wall-degrading enzymes when faced with structurally and chemically complex biomass.

Designing an experimental plan to tackle the various aims of this project was not an easy task, since the goals themselves were rather ambitious. Nevertheless, it was clear that one way forward was to employ some of the latest techniques in enzyme engineering, including *in vitro* random mutagenesis and DNA shuffling, which have already been shown to be powerful for the improvement of enzyme properties (Johannes and Zhao, 2006). However, previous experience drawn from work performed on cellulases indicates that engineering enzymes for biomass deconstruction is hazardous, especially when model substrates are employed. Therefore, the first challenge in this doctoral work was the conception, development and testing of a high-throughput screening procedure that employs raw biomass and that is sufficiently reliable to identify true hits in vast libraries.

The second part of the work described here addresses the implementation of the screening procedure and the generation of a series of mutants. In this phase of the work, it was necessary to adapt experimental procedure to new constraints and to maintain an overall aim, which was to exploit the power of *in vitro* random mutagenesis and DNA shuffling, by performing several rounds of these techniques.

Finally, in the third part of the research performed in this study, a semi-rational approach was adopted, whose context was the investigation of structure-function relationships via the creation of a chimeric enzyme. Based on previous observations made by workers in Prof. Harry Gilbert's laboratory (University of Newcastle, UK), the primary goal of this study was to create a more extensive active site in a xylanase and, if successful, examine whether this modification would be translated into higher enzyme activity.

All of the work performed in this study, and described here, was performed on a xylanase from *Thermobacillus xylanilyticus* that belongs to family 11 of the glycoside hydrolase classification CAZy. This enzyme has been one of the preferred study objects of Dr. Michael O'Donohue, the supervisor of this work, for many years, and thus a considerable amount of prior knowledge is available and importantly, its 3D structure has been solved. Moreover, as

a member of family GH11, this xylanase belongs to the group of xylanases that has been most extensively implemented in industrial processes, and as a thermostable protein, it figures among some of the best candidates for further development. Finally, unlike many xylanases, the action of the GH11 xylanase from *Thermobacillus xylanilyticus* has already been studied in detail on industrially pertinent biomass, such as wheat bran and wheat straw.

In the chapters that follow, the reader is invited to discover an extensive bibliographic introduction to the key concepts and state of the art knowledge that form the basis of the study, the description of experimental work, which has been organized into three standalone manuscripts (the first of which is published), followed by a discussion of general conclusions and suggestions for future work. An annexe to this thesis supplies a complementary view of state of the art knowledge in the area of hemicelluloses-specific biocatalysis.

Clarke, N.D., 2010. Protein engineering for bioenergy and biomass-based chemicals. *Current Opinion in Structural Biology*, 20, 527-532.

Gao, D., Chundawat, S.P.S., Krishnan, C., Balan, V., Dale, B.E., 2010. Mixture optimization of six core glycosyl hydrolases for maximizing saccharification of ammonia fiber expansion (AFEX) pretreated corn stover. *Bioresource Technology*, 101, 2770-2781.

Johannes, T.W., Zhao, H., 2006. Directed evolution of enzymes and biosynthetic pathways. *Curr Opin Microbiol*, 9, 261-7.

Kumar, R., Wyman, C.E., 2009b. Effect of xylanase supplementation of cellulase on digestion of corn stover solids prepared by leading pretreatment technologies. *Bioresour Technol*, 100, 4203-13.

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

Murashima, K., Kosugi, A., Doi, R.H., 2003. Synergistic effects of cellulosomal xylanase and cellulases from Clostridium cellulovorans on plant cell wall degradation. *J Bacteriol*, 185, 1518-24.

Prior, B.A., Day, D.F., 2008. Hydrolysis of ammonia-pretreated sugar cane bagasse with cellulase, beta-glucosidase, and hemicellulase preparations. *Applied Biochemistry And Biotechnology*, 146, 151-164.

Wilson, D.B., 2009. Cellulases and biofuels. *Curr Opin Biotechnol*, 20, 295-9.

Wyman, C.E., 2007. What is (and is not) vital to advancing cellulosic ethanol. *Trends in Biotechnology*, 25, 153.

BIBLIOGRAPHY

In this chapter, the scientific literature, relevant to the research work presented, is organized into four sections. In part A, we describe the nature of lignocellulosic biomass and present the configurations of state-of-the-art biorefining processes, highlighting the importance of enzyme-mediated operations. In part B, the general biological and biochemical properties of xylanases are exposed and current knowledge of structure-function relationships of GH11 xylanases is described. Part C is a historical review of studies that have been performed on Tx-Xyl xylanase, which is the target enzyme in this thesis study. Finally, in part D, the state of the art in random engineering of proteins is described and the application of these techniques to xylanases is summarized.

Part A Biorefining of lignocellulosic biomass

A.1 What is lignocellulosic biomass?

The term lignocellulosic biomass designates parts of plants that are generally inedible and composed of composite material, consisting of various plant polysaccharides (including cellulose and hemiucelluloses) and lignin (Himmel et al., 2007). This organic matter, generated from carbon dioxide fixed by plants during photosynthesis and stored in plant cell walls, plays a wide variety of functions in plant physiological processes. Unlike the use of fossil resources (oil, coal and natural gas) that constitute natural, long-term repositories of "ancient CO_2", stocked by plants millions of years ago, the use of lignocellulosic biomass does not inevitably lead to increases in atmospheric greenhouse gases. Regarding the production of lignocellulosic biomass, global annual growth represents between $10 - 50$ billion tonnes (Sticklen, 2006), although most of this is available for use. Since the use of excess biomass (i.e. the amount of biomass produced once permanent biomass, such as primary forests, have been substracted) represents a renewable, sustainable source of carbon and energy, it constitutes a viable alternative to unsustainable fossil resources.

A.1.1 Structure and composition of lignified plant cell wall

In the plants, polysaccharides, highly glucosylated proteins and lignin are the principal components of cell walls (Somerville et al., 2004). In this section, the microstructure and the biochemical composition of the plant cell wall, an important object in this doctoral thesis, will be described

A.1.1.1 Structural organisation of plant cell wall

The plant cell wall is the outer layer of cells that surrounds the plasma membrane and provides structural support and protection to the cell contents. Its shape and components vary between different cell types. Plant cell walls can be divided into two types: primary and secondary walls (Figure A-1). Primary walls can be found in all living plant tissues, and are located outside the secondary wall when there is one. The secondary wall is formed after the cell is fully grown and is not found in all cells (Brett and Waldron, 1996; Somerville et al., 2004).

Generally, the primary wall is a thin and extensible layer generated during cell growth and division. The predominant components are cellulose, hemicellulose and pectin. The framework of primary wall is mainly built up by cellulose microfibrils embedded in hemicellulose and pectin chains, with only a small amounts of lignin being (sometimes) present.
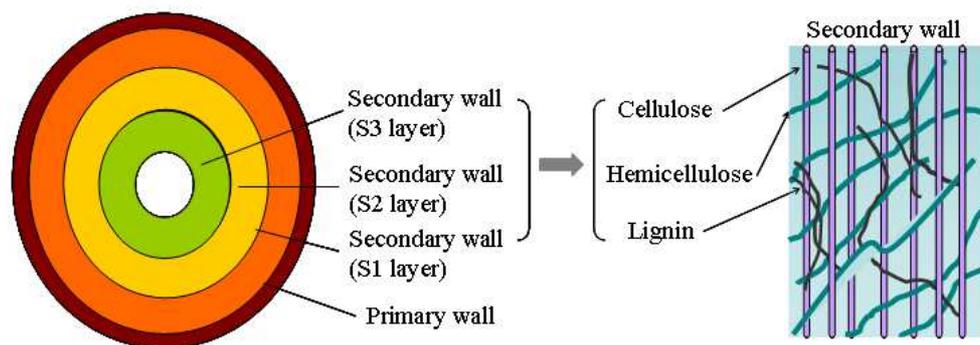
Figure A-1. Schematic representation of the plant cell wall cross-section and the polymeric complexes in the secondary wall. The primary wall encloses the secondary wall, and the secondary wall is composed of S3, S2 and S1 layers from inside to outside. The framework of the secondary wall is formed by three kinds of cross-linked macromolecules, which are cellulose, hemicellulose and lignin. Adapted from (Sticklen, 2008) .

Unlike the primary wall which is very flexible, the secondary wall is stiffer, its principal role being to strengthen and protect the cell from phytopathogenic aggression, dehydration and other environmental risks. The secondary wall is decomposed into S1, S2 and S3 layers (Figure A-1), where S1 and S3 are very thin, contrary to the S2 layer that constitutes 70-90% of the overall thickness of the secondary wall (Sticklen, 2008). The secondary wall is mainly made of cellulose, hemicelluloses and lignins, with the latter being intimately linked to the polysaccharide chains through both covalent and non-covalent bonds. As a consequence, cells possessing a secondary wall are also known as lignified plant cells.

A.1.1.2 Structure and interactions of wall polymers in lignified plant cell wall

As mentioned above, the main chemical components of lignified cell walls are cellulose, hemicellulose and lignin. Their structure, chemical and biological properties are depicted as follows.

(1) Cellulose

Cellulose is primary terrestrial repository of renewable carbon, and accounts for 35-50% of the dry content of lignocellulose. The structure of cellulose *in vitro* consists of a long linear chain of β-1,4 linked glycosyl groups, with a degree of polymerization ranging from 300 to 10 000 units. *In vivo*, the cellulose chains assemble together side-by-side, interacting through multiple hydrogen bonds involving the hydroxyl groups on the glucose residues and thus form multichain structures, called microfibrils (Figure A-2) (Bayer et al., 1998; Nishiyama et al., 2002). Microfibrils are highly stable, and high temperature and pressure are required to solubilise their very resistant crystalline organisation (over 320°C and 25 bar) (Shigeru Deguchi et al., 2006).
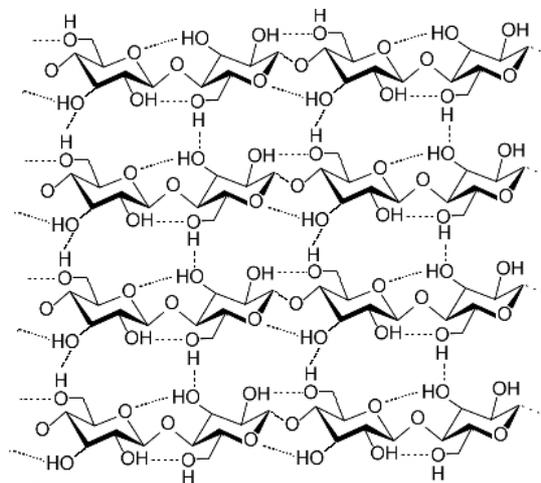
Figure A-2. Mesh structure of cellulose microfibrils.

(2) Hemicelluloses

Hemicelluloses are branched heteropolysaccharide chains, composed of pentoses (D-xylose and L-arabinose), hexoses (D-mannose, D-galactose, and D-glucose) and uronic acids (D-glucuronic acid and D-galacturonic acid). Hemicellulose constitutes 20-40% of the dry weight content of lignocellulosic biomass, and is the second most abundant source of renewable carbon and the first source of pentose sugars (Ebringerová et al., 2005; Fraser-Reid et al., 2008; Saha, 2003; Scheller and Ulvskov, 2010). Depending on the backbone residues, hemicelluloses are grouped into xylans, glucomannans, xyloglucans, mannans, and β-(1→3, 1→4)-glucans. The content of various hemicellulose types is quite different in herbage, softwoods and hardwoods. Predominant hemicelluloses are the (glucurono-

)arabinoxylans (including arabinoxylan and glucurono-arabinoxylan) and galacto-glucomannans that belong to xylans and mannans respectively , while the other types are present in lesser proportions (Fraser-Reid et al., 2008). The details are summarized in Table A-1.

(Glucurono)arabinoxylans are the major hemicellulose found in the secondary cell walls of grasses (20-40%, w/w of dry mass), and are also present in softwoods (5-15%), whereas glucuronoxylan is the major hemicellulose of hardwoods (15-30%) (Scheller and Ulvskov, 2010). Consequently, heteroxylans are the second most abundant polysaccharide family after cellulose (Haltrich et al., 1996; Prade, 1996).
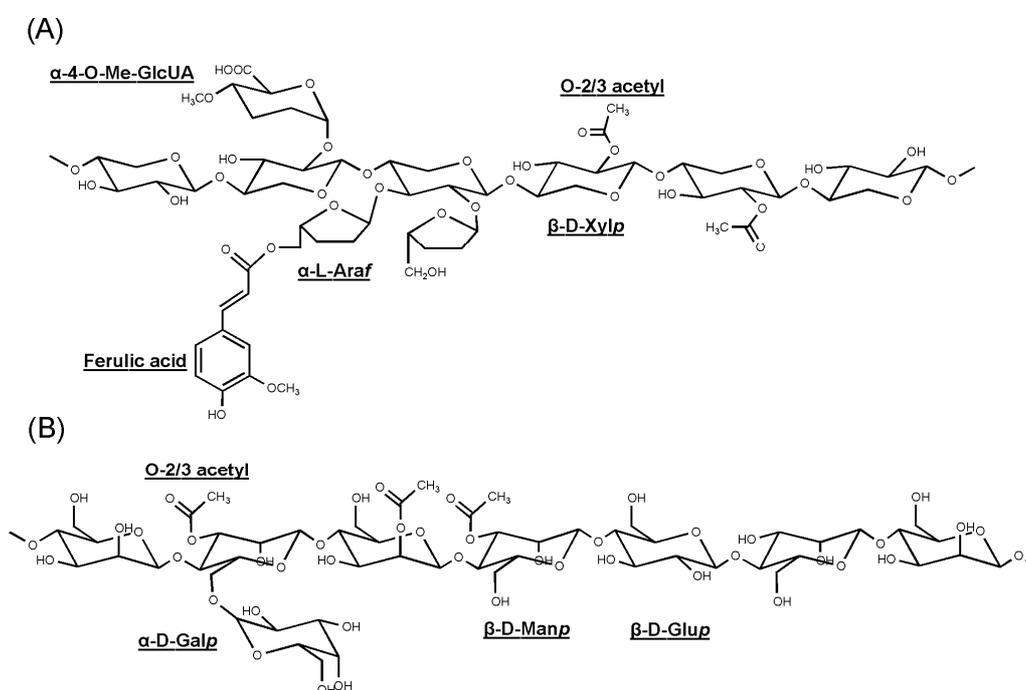
Figure A-3. Schematic structures of glucuronoarabinoxylan (A) and galactoglucomannan (B).

Xylans are composed of a backbone made up of 1,4-linked β-D-xylosyl residues, which can be substituted with L-arabinofuranosyl (L-ara*f*), *O*-acetyl, α-4-*O*-methyl-D-glucuronic acid (MeGlcUA), feruloyl and *p*-coumaroyl residues (Figure A-3.A and Table A-1). More specifically, side chains of glucurono/arabino-xylans are principally α-1,2 or 1,3-linked L-ara*f*, which can be present simultaneously on one main chain xylosyl unit, MeGlcUA, and sometimes feruloyl or *p*-coumaroyl moieties, that are linked to the C-4 position of the arabinosyl residue via an ester linkage (Berrin and Juge, 2008; Zimmermann, 1991). Glucuronoxylans are substituted by α-1,2-linked D-glucuronic acids that are frequently methylated on their *O*-4 position, and

nearly 70% of xylosyl residues are acetylated at their C-2 or C-3 positions (Kulkarni et al., 1999; Numan and Bhosle, 2006). Side chains of xylans are important, since they influence many polymer properties such as solubility, physical conformation, interactions with cellulose and lignins.

Galacto-glucomannans are the second most abundant hemicelluloses, mainly present in softwood species. The backbone of galacto-glucomannans is a β-1,4-linked copolymer chain made up of glucose and mannose units in the ratio 4:1. Single galactose side chain are occasionally present as C6-substitutents of mannosyl residues (Figure A-3.B and Table A-1).

(3) Lignin

Lignin is a complex phenolic macromolecule that accounts for 10-25% dry weight content of lignocelullosic biomass. The structure of lignin is derived mainly from three hydroxycinnamyl alcohol monomers: *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol (Figure A-4) (Boerjan et al., 2003; Lapierre et al., 1986). Three mono-lignols are further transformed into *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) phenylpropanoid units and form an amorphous three-dimensional lignin polymer. Inside the plant cell wall, lignin never exists as an independent polymer but covalently binds to hemicelluloses via linkages between α-carbons and C-4 of benzene rings (Buranov and Mazza, 2008; Ralph et al., 1995). Due to its high heterogeneity and complex linkage, it is impossible to isolate pure lignin polymer and define its precise primary structure.
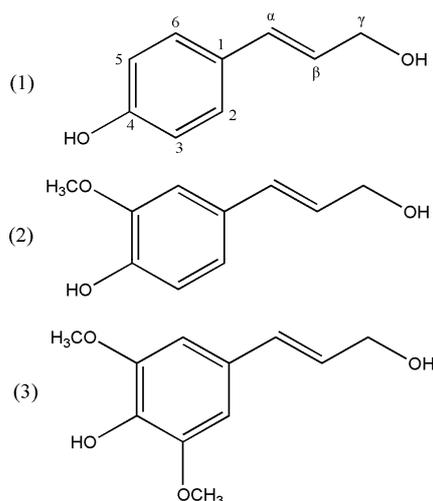
Figure A-4. The three lignin monomers: *p*-coumaryl alcohol (1), coniferyl alcohol (2), and sinapyl alcohol (3). Adapted from (Buranov and Mazza, 2008)

Table A-1. **Main types of hemicelluloses in plant cell wall.** Information are from (Ebringerová et al., 2005; Fraser-Reid et al., 2008; Girio et al., 2010; Scheller and Ulvskov, 2010)

Hemicelllulose Type	Amount in plant cell wall (% w/w, dry mass)						Backbone		Side chains	
	Herbage		Softwood		Hardwood		Unit	Linkage	Unit	Linkage
	Primary	Secondary	Primary	Secondary	Primary	Secondary				
(Glucurono-)arabinoxylan*	15-30	20-40	Minor	2-15	0-5	-	D-Xyl*p*	β-(1→4)	L-Ara*f* 5-*O*-Feruloyl-L-Ara*f* 4–*O*-Me-D-GlcA*p* Acetyl	α-(1→2/3) α-(1→2/3) α-(1→2/3) *O*-2/3
Glucuronoxylan	-	-	-	-	-	15-30	D-Xyl*p*	β-(1→4)	4–*O*-Me-D-GlcA*p* Acetyl	α-(1→2) *O*-2/3
Galacto-glucomannan	-	-	Minor	10-30	-	0-3	D-Man*p* and D-Glu*p* in ratio 4:1	β-(1→4)	D-Gal*p* † Acetyl	α-(1→6) *O*-2/3
Glucomannan	0-2	0-5	-	-	2-5	2-5	D-Man*p* and D-Glu*p*	β-(1→4)		
Xyloglucan	2-5	-	0-10	-	20-25	Minor	D-Glc*p*	β-(1→4)	D-Xyl*p* β-D-Gal*p*-(1→2)-D-Xyl*p* α-L-Fucp-(1→2)-β-D-Galp-(1→2)-D-Xylp α-L-Ara*f* -(1→2)-D-Xyl*p* β-L-Ara*f* -(1→3)-α-L-Ara*f* -(1→2)-D-Xyl*p* Acetyl	α-(1→6) α-(1→6) α-(1→6) α-(1→6) α-(1→6) *O*-2/3 of Xyl*p*
β-(1→3, 1→4)-glucan	2-15	Minor	-	-	-	-	D-Glc*p*	β-(1→4):β-(1→3) = 3:4		

* (Glucurono-)arabinoxylan include glucurono-arabinoxylan and arabinoxylan

† Substitution of D-Gal*p* only happens on mannose backbone residues

(4) Overall spatial organisation

In secondary walls, the three polymer families described above are spatially dispersed, and attached to one another, making up a highly cross-linked network (Lin and Tanaka, 2006; Sticklen, 2008; Takashi, 1989). Here, hemicelluloses work as molecular glue, coating the surface of cellulose microfibrils, mainly via hydrogen bonds, and associating with lignin via covalent bonds (Grabber et al., 2004; Somerville et al., 2004). In addition, a portion of lignin directly links to cellulose (Choi et al., 2007; Tenkanen et al., 1999).

In woody biomass, lignin forms a complex with hemicelluloses via benzyl ester, benzyl ether, acetal and glycosidic bonds (Koshijima and Watanabe, 2003; Takashi, 1989). In herbaceous crops, lignin can be covalently linked to hemicelluloses via so-called phenolic bridges (Figure A-5). These phenolic bridges (involving feruloyl or *p*-coumaroyl moieties) which are formed via the establishment of ester links between two phenolic acids or by various types of ether or C-C links between a phenolic acid and lignin, provide the means to physically link two hemicellulose molecules or a hemicellulose molecule to lignin (Grabber et al., 2004; Ralph et al., 1998). Both alkali and acidic reagents can cleave the lignin-carbohydrate complex, acting on ester- or ether- bonds respectively, and acid treatment can cut the ether bond between sugar residues (Figure A-5) (Buranov and Mazza, 2008; Sun et al., 2002).
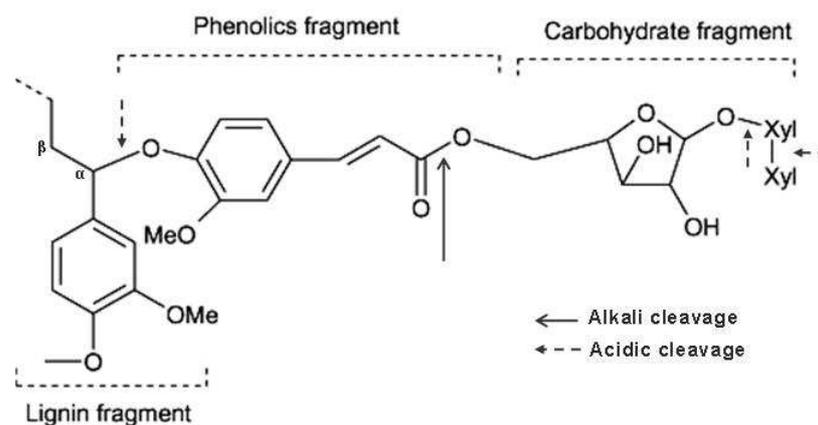
Figure A-5. Lignin-arabinoxylan complex in wheat straw. Adapted from (Buranov and Mazza, 2008).

Since lignin is a hydrophobic and aromatic macromolecule, the association of carbohydrates and lignin results in the cell wall becoming more hydrophobic, impermeable and difficult to degrade, thus lignification can be considered to be the erection of a defensive barrier

(Buranov and Mazza, 2008). From the point of view of biorefining, the existence of lignin-carbohydrate complexes constitutes a major physical and spatial obstacle in the conversion of lignocellulosic biomass into useful products.

In conclusion, lignocellulosic biomass holds considerable promise in terms of renewable carbon and energy, and as such is a potential feedstock for biofuels and biochemicals. However, lignified cell walls are naturally designed to resist biological or chemical aggression. Therefore one of the greatest challenges for biorefining is to find how to efficiently breakdown the cell wall barrier, without destroying the intrinsic value of the molecular constituents.

A.1.2 Availability of lignocellulosic feedstocks for biorefining

In terms of quantity, the primary sources of lignocellulosic biomass are wood and the coproducts of major cereal crops. Another important source of lignocellulosic biomass that is often neglected is domestic and industrial waste materials, which can also form part of virtuous recycling strategies. Therefore, it is possible to classify lignocellulosic biomass into two groups: one is natural resource and the other is derived from human activities (Table A-2).

Table A-2. Available lignocellulosic feedstocks

Category	Sub-category	Examples
Natural source	Woody plants	Forest residues Short rotation plantations (e.g. willow, poplar)
	Herbaceous plants	Agricultural crops (e.g. wheat straw, corn stover) Grasses (e.g. Napier grass, reed canarygrass)
Industrial & municipal sources	Industrial by-products and wastes	Fibrous waste and lignin from paper industry; Industrial waste wood (e.g. sawdust from sawmills)
	Municipal wastes	Waste papers, demolition wood, building rubble

Theoretically, all lignocellulosic biomass types are potential feedstocks for biorefining. However, accounting for economic and logistical constraints such as purchase price, available volumes, feasibility of harvesting and transportation, cereal crop by-products are suitable sources for countries that have a strong agricultural industry (e.g. west Europe, eastern and south-eastern Asia and USA), and wood is a potentially very abundant feedstock

in forest-dense areas (e.g. north Europe and north America) (Ash, 2004; Möller et al., 2007) (Somerville et al., 2010). In addition to forestry and agri-food industries, an alternative source of biomass are so-called energy crops, which are plants that are grown in the primary aim of supplying biorefineries. Ideally, these crops, which include *Miscanthus* sp., switchgrass and short rotation coppice such as willow, poplar etc., should be grown on marginal or polluted land that is not used for food production (Borland et al., 2009; Campbell et al., 2008; Somerville et al., 2010; Yuan et al., 2008).

A.1.2.1 Wheat straw as an important biomass feedstock

Wheat (*Triticum spp.*) is the second most popular cereal crop in the world and is widely cultivated between 30-50° N and 25-40° S latitudes (U.S.DOE, 2006). Wheat cultivation worldwide uses about 2.2 billion hectares, and 681 million tons of wheat grain was harvested in 2009 (FAO, June 2009). The ratio of wheat straw to grain yields is reported to be in the range 1.25-1.67 (Engel et al., 2005). Therefore annual global wheat straw production is estimated to be at least to 850 million tons. Taking into account the fact that the European Union is the world's primary producer of wheat, it is logical that wheat straw is the most abundant crop residue in Europe.

Table A-3. Chemical composition of wheat straw (w/w, % content of dry mass)

Country	Cellulose	Hemicellulose (Xylan)	Klason lignin	Ash	Reference
Mexico	49.2	31.4	10.8	-	(Salmones et al., 2005)
India	36.6	24.8 (21.6)	14.5	9.6	(Chandrakant and Bisaria, 1998)
USA	36.7	34.7	18	7.4	(Mckean and Jacobs, 2007)
USA	39.8	27.3 (24.5)	22.6	4.2	(Kristensen et al., 2008)
Denmark	49.8	22.9 (19.2)	16.9	1.9	(Pedersen and Meyer, 2009)
France	41.2	- (26.1)	19.1	-	(Rémond et al., 2010)

The composition of wheat straw can be influenced by both the genotype and pedoclimatic conditions. Table A-3 lists the chemical contents of wheat straw from different countries. In general terms, wheat straw contains a high level of carbohydrates that account for 70.1±5.9% of total weight, and xylan represents the majority of the hemicellulose. Lignin content varies significantly, but represents about 17.0±3.7% (coefficient variation = 22%), which may result from different methods of biomass decomposition and measurement. In the molecular structure of wheat straw hemicellulose, *p*-coumaric and ferulic acids are

involved in cross-linkages between polysaccharide and lignin chains, so that these two hydroxycinnamic acids exist as esterified or etherified compounds *in vivo* (Buranov and Mazza, 2008). Sun et al. have observed that esterified and etherified *p*-coumaric acids account for 3.78% and 1.72% of total components, while esterified and etherified ferulic acids are in the amount of 1.02% and 2.20% respectively, indicating a much less amount of esterified *p*-coumaric acid than in corn stovers (6.7%) (Sun and Tomkinson, 2002).

Current industrial applications of wheat straw mainly focus on animal feed, papermaking, and combustible fuel. In a long-term view, wheat straw constitutes an attractive feedstock for biorefinery industry due to its advantages of low cost (35-55 € per dry tonne, quote online), regional abundance and homogeneous production (Banowetz et al., 2008; Buranov and Mazza, 2008; Kerstetter and Lyons, 2001). In addition, excellent knowledge of wheat genetics and breeding techniques holds promises to optimize cell wall characteristics and to generate dedicated bioenergy wheat varieties (Möller et al., 2007). It is believed that expansion of wheat straw as biorefinery feedstock will offer a new market for agricultural crop residues and will benefit rural areas.

A.2 The biorefinery concept

This section will consider the state-of-the-art processing options used in the biorefinery of lignocelluloses, and highlights the actions of microorganisms/enzymes in the process.

A.2.1 General introduction

By definition, biorefining concerns the production of energy, chemicals and materials from biomass, associated with an overall favourable environmental footprint that is consistent with 'green' strategies. To achieve this, biorefineries will rely to some extent on the use of biotechnology tools that will contribute to a zero waste approach and to sustainable production (Clark and Dewarte, 2009; McKendry, 2002). Biorefining has been identified as a key technology for the necessary transition towards a post-petroleum economy, not only because it provides renewable energy, especially liquid fuels (bioethanol and biodiesel), but also because its predominant contribution is to produce chemicals and materials that other renewable energy sources (winds, solar and hydro powers, etc) can not furnish.

In contrast to 1st generation biorefinery processes, which use cereal grains (corn, wheat) and sucrose-rich plants (sugar cane, sugar beet) as raw material and lead to potential food competition, lignocellulose-based biorefining uses non-food feedstocks and is thus classified as 2nd generation biorefining. However, the conversion of lignocellulose is more difficult than the conversion of starch grain or sucrose, as the structural and chemical composition of lignocelluosic biomass is much more complex. Currently, hemicellulose conversion is divided into two distinct itineraries, which are the thermochemical and biotechnological routes (Goyal et al., 2008; Hubbard et al., 2007; McKendry, 2002).

Thermochemical conversion relies on heat and chemical reactions, as means of extracting and creating products and energy (Hubbard et al., 2007). The main processes include combustion, gasification, pyrolysis and liquefaction (Goyal et al., 2008). Combustion directly burns biomass to produce heat and electricity; gasification occurs at 800-900°C to decompose lignocelluloses into synthesis gases, mainly comprised of hydrogen and carbon monoxide; pyrolysis is performed in anaerobic conditions within a temperature range of 350-550°C, and converts biomass into energy, liquid oil, gases and char; liquefaction liquefies biomass under conditions of high pressure and low temperature, together with hydrogen and a catalyst (Goyal et al., 2008; Saxena et al., 2008). In the case of lignin conversion into added value products, themochemical methods, especially pyrolysis, appears to offer more potential than biotechnological strategies (Lynd et al., 2005; Yang et al., 2007) (Goyal et al., 2008; McKendry, 2002).

Biotechnological conversion can be subdivided into two main processes: anaerobic digestion and fermentation (McKendry, 2002).

(1) Anaerobic digestion converts lignocellulosic biomass into biogases using consortia of methanogenic bacteria operating in an anaerobic environment. The biogas product is mainly composed of methane and carbon dioxide. This process is widely used for biomass wastes with high moisture content, and the most used microbial species include *Methanobacterium sp.*, *Methanococcus sp.*, *Methanosarcina sp.* and *Methanopyrus sp.* (Arshadi and Sellstedt, 2009). The purified methane can be utilized as fuel or as a chemical raw material. The overall efficiency from biomass to electricity through gas combustion is about 10-16% (McKendry, 2002).

(2) Unlike anaerobic digestion, fermentation employs pure (axenic) microbial cultures involving a bacterium or yeast strain. The process is usually oxidative and monosaccharides (predominantly glucose and xylose extracted from lignocelluloses) are transformed into fuels or desired chemicals.

With the development of strain engineering, especially for the use of pentose sugars, fermentation is becoming an extremely powerful industrial option (Hendriks and Zeeman, 2009). Glucose can be selectively fermented into various chemicals, among which the two main representatives are lactic acid and ethanol (Ohara, 2003). Lactic acid can be polymerized to make plastics, solvents and other chemicals through the processing of esterification, dimerization and polymerization (Dodds and Gross, 2007; Ragauskas et al., 2006). Ethanol mainly serves as a transportation fuel, and can also be esterified, forming ethyl lactate which is used as a biodegradable solvent, or dehydrogenated to form ethylene which then can be used to form polyvinyl chloride (Clark and Dewarte, 2009; Ohara, 2003). Xylose is mostly fermented into ethanol or xylitol. The latter is one of the top 12 value-added chemicals evaluated by the US Department of Energy, and its derivatives include xylaric acid, hydroxyl-furans, polyesters, and propylene glycol (Koutinas et al., 2009). The residues of fermentation are usually rich in phosphates and potash, both of which can be utilized as fertilizers.

A.2.2 State of the art of process configurations

To date, although biorefining of lignocellulosic biomass only exists at laboratory or pilot-scale stage (Huang et al., 2008), extensive studies are progressing in Europe (e.g. Biorefinery Euroview, BIOPOL, BIOCORE projects), USA and other countries (Clark and Dewarte, 2009). It is expected that 2nd generation biorefining should be highly flexible, using a wide variety of lignocelluloses feedstocks and producing an array of end-products.

To achieve these ambitious goals, it is recognized that an integrated platform should basically include the following four modules, also shown in Figure A-6.

(1) Module 1 – physical separation and pretreatment: involve milling or grinding of biomass into small particles, and further extraction of lignin and hemicelluloses from cellulose through chemical, physical and/or biological pretreatment methods.

(2) Module 2 – enzymatic hydrolysis (saccharification): aimed at fractionating cellulose and hemicelluloses into monomeric sugars independently or simultaneously by catalytic cocktails made of various microorganisms and/or enzymes. Enzymatic hydrolysis is thought to be superior to chemical methods, such as acid hydrolysis, because enzymes are more specific, thus avoiding the generation of degradation products and reducing the need for depollution of effluents.

(3) Module 3 – biological and/or chemical catalysis (including fermentation): tends to use multiple processes to convert sugar syrups (glucose and pentoses) into a range of bio-products.

(4) Module 4 – product recovery/separation and purification: these are general grouped under the term downstream processes.
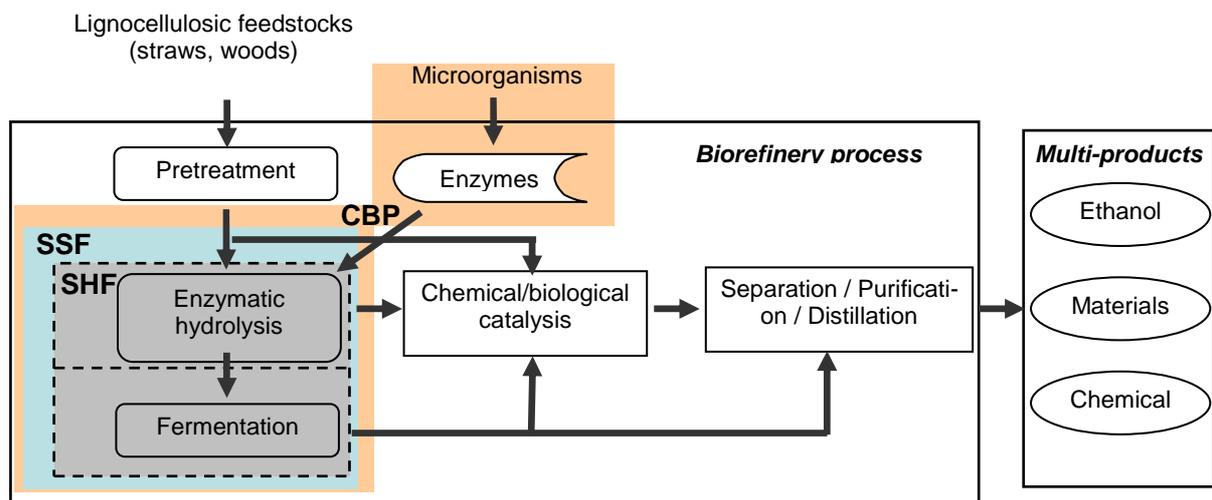
Figure A-6. Biotechnological biorefinery strategies for lignocellulosic biomass. Enzymatic saccharification and fermentation are performed independantly in the separate hydrolysis and fermentation concept (SHF) (grey background with dashed frame), but are performed simultaneously in the simultaneous saccharification and fermentation concept (SSF) (blue background). Consolidated bioprocessing (CBP) (pink background) is an integrated concept in which enzyme production, saccharification and fermentation occur in a single reactor. The figure is adapted from (Clark and Dewarte, 2009) and (Girio et al., 2010).

With the exception of downstream processes, which generally constitute independent process operations, the other three modules can be conducted either separately, or through the combination of two or three steps. Currently two process strategies exist: separate hydrolysis and fermentation (SHF) and simultaneous saccharification and fermentation (SSF)

(Figure A-6). As indicated by their names, the two strategies differ in that hydrolytic and fermentative steps are either discrete or simultaneous. More specifically, when soluble pentose is fermented together with hexose, the above two processes are designated as separate hydrolysis and co-fermentation (SHCF) and simultaneous saccharification and co-fermentation (SSCF).

In the concept of future biorefining, <u>consolidated bioprocessing</u> (CBP) is a favourable alternative that was first proposed by Lynd in 1996. Basically, CBP combines enzyme production, polysaccharide hydrolysis and fermentation in a single step, using microorganisms that possess both (hemi)cellulolytic and fermentative properties (Figure A-6) (Lynd, 1996). Compared to other transformation itineraries, CBP has important advantages, especially in cost reduction and efficiency improvement (French, 2009; Lynd et al., 2002). As a comparison, in 2005, it was estimated that the CBP process could decrease the theoretical cost of bio-ethanol from 0.036 €/l (SSCF process) to 0.008 €/l, while shortening reaction time from 7 days to 1.5 days (Lynd et al., 2005).

A.2.3 Pretreatment of lignocellulosic biomass

As lignocellulosic biomass is hard to fractionate, the aim of pretreatment is to break down the plant cell wall matrix, extract the three main components (cellulose, lignocelluloses and lignin) and render cellulose more accessible to further enzymatic hydrolysis (Ragauskas et al., 2006). Consequently, pretreatment critically influences the feasibility and efficiency of subsequent steps (Wyman, 2007). Existing pre-treatment methods are classified into chemical, physical and/or biological families according to the process and reagents employed (Figure A-7).

Mechanical milling or grinding is generally carried out for all raw biomasses, to increase biomass density, enlarge accessible surface area, and reduce cellulose crystallinity. Regarding chemical pretreatment, the intrinsic pH of this step determines to a large extent the composition of the residual, solid lignocellulosic fraction. Acidic (including hydrothermal) treatments selectively target hemicellulose removal through the generation and action of hydronium ions and/or high temperature, and result in liquid fractions rich in hemicellulose oligomers and cellulose/lignin-enriched solids (except in the case of concentrated acid

treatment, where monosaccharides are generated) (Martin et al., 2008; Ruiz et al., 2008). In contrast, alkaline reagents cleave ester bonds between lignin and polysaccharides, and thus remove lignin, but only partly solubilise hemicelluloses (Girio et al., 2010). In addition, alkaline pretreatment removes acetate substitutions (Carvalheiro et al., 2008). Of the various methods, hydrothermal pretreatment using liquid hot water is considered the best strategy to remove hemicellulose from cellulose/lignin-enriched solids, because of the mild pH involved, the reduced amounts of polluted effluents, and low level of inhibitory by-products (Girio et al., 2010; Merino and Cherry, 2007; U.S.DOE, 2006). However, to maximize lignin extraction, leaving a residue of hemicelluloses and cellulose, ammonia fiber explosion appears to be a better choice (Carvalheiro et al., 2008; Chundawat et al., 2007).
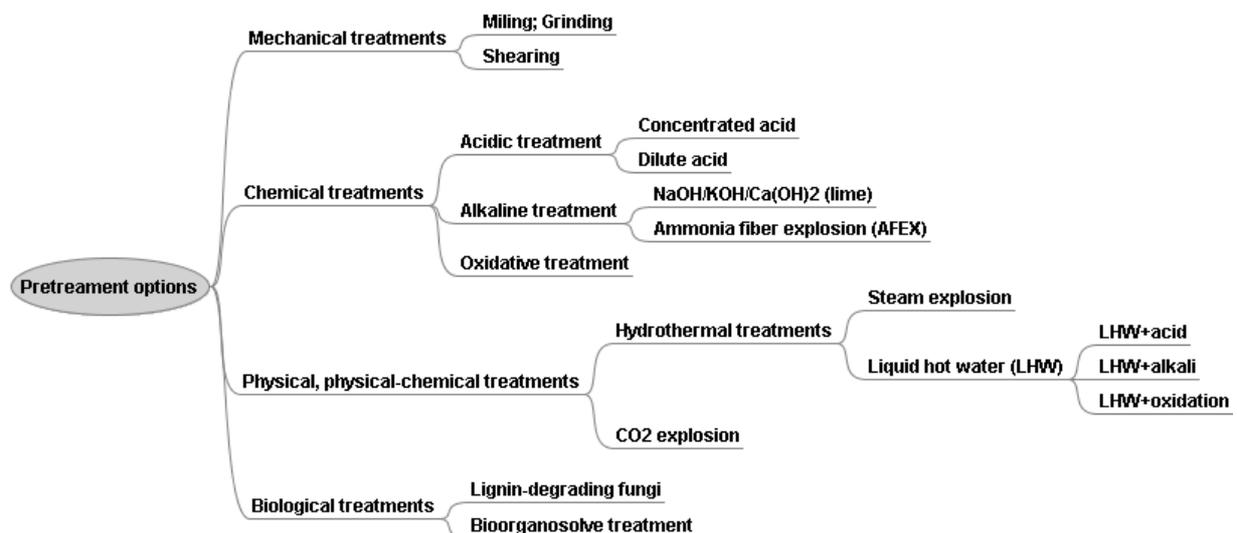
Figure A-7. Various pretreatments tailored for lignocellulosic biomass.

Biological pretreatment employs microorganisms (e.g. white-rot fungi) that directly act on raw lignocellulosic biomass, by secreting one or more lignin and/or hemicellulose-degrading enzymes. Compared to other pre-treatments, the biological approach is characterized by mild reaction conditions, high specificity, and no generation of fermentation inhibitors, and is potentially highly environmentally friendly. However, current biological pretreatments are slow and thus require long reaction times, which are incompatible with industrial demands

(Sánchez and Cardona, 2008). Therefore, it is necessary to develop new strains and enzymes with improved efficiency.

A.2.4 Saccharification and fermentation in various process configurations

Regarding enzymes for SH(C)F process, complex cocktails of containing cellulases and hemicellulases are employed. For example, a cocktail of CMC-ase cellulase, β-glucosidase xylanase, β-xylosidase and α-arabinofuranosidase successfully solubilised 96.7% of total sugars held within alkaline peroxide-pretreated wheat straw (Saha and Cotta, 2006). Monomeric hexoses and pentoses produced by enzymatic hydrolysis can be fermented, either using yeast (e.g. *Saccharomyces cerevisiae, Kluyveromyces marxianus, Pichia stiptitis*) or bacteria (e.g. *E. coli, Z. mobilis*). The yeast *S. cerevisiae* is the most commonly used microorganism to produce ethanol, which uses hexoses as substrate. Many years of metabolic engineering on this yeast have recently allowed the creation of new strains that are also able to ferment pentoses (Fonseca et al., 2007; Hahn-Hagerdal et al., 2007a; Hahn-Hagerdal et al., 2007b)

In SS(C)F process, enzymes and microorganisms are added into one reactor, and monosaccharides generated by enzyme action on polysaccharides are immediately consumed by the microorganism, thus avoiding side reactions and end-product inhibition (Olofsson et al., 2008). In many cases, fermentation can be carried out prior to the completion of saccharification. As a matter of fact, the extent and speed of ethanol production can increase in SS(C)F compared to SHF (Arshadi and Sellstedt, 2009; Merino and Cherry, 2007; Olofsson et al., 2008; Stenberg et al., 2000). Nevertheless, to maintain correct rheological conditions, very large amounts of culture medium are used, resulting in diluted products. Moreover, because the SS(C)F process requires a compromise between hydrolysis and fermentation conditions (in terms of pH and temperature optima), enzyme loading is higher than in the SH(C)F process in order to ensure adequate hydrolysis (Sánchez and Cardona, 2008). Consequently, decreasing the cost of saccharolytic enzymes is important for the SS(C)F strategy.

Identifying naturally occurring microorganisms that satisfy all of the requirements for CBP is difficult, although a number of candidates that exhibit some of the requirements are known

(e.g. thermophilic bacteria that display good utilization of pentoses and hexoses or yeast that are characterized by high productivity and product yields). Therefore, strain engineering is almost certainly a prerequisite for the production of efficient CBP microorganisms and in this respect synthetic biology holds much promise (French, 2009; Lynd et al., 2005). In recent studies, cellulases, hemicellulases or even mini-cellulosomes have been successfully expressed on the surface of recombined *S. cerevisiae* to complement its natural fermentation ability (Lilly et al., 2009; van Zyl et al., 2007), constituting a definite breakthrough for the CBP configuration.

A.3 Enzymes in biorefining

The enzymes involved in biorefining are categorized into two groups: cellulases and hemicellulases. Cellulases refer to a series of enzymes participating in cellulose hydrolysis, but which can also be active on hemicelluloses. Similarly, hemicellulases correspond to a multi-component enzyme system active towards the degradation of hemicelluloses. These enzymes can be either added into the reactor in the form of a crude or purified enzymatic preparation, or directly produced in the reactor by (hemi)cellulolytic microorganisms. In this paragraph, we will focus on the different elements of each enzyme group and on how they participate to substrate hydrolysis.

A.3.1 Cellulases

A.3.1.1 Type of cellulases and mode of action

In order to break down cellulose chains into glucose monomers, a minimum of three major cellulose activities are required, which are endoglucanases (EC 3.2.1.74), exoglucanases (EC 3.2.1.91) and β-glucosidases (EC 3.2.1.21) (Chang, 2007; Lynd et al., 2002). Concerning their enzymatic activities (Figure A-8), endoglucanases cleave internal β-1,4 glycosidic linkages of amorphous cellulose to form oligosaccharides of various lengths. Exoglucanases cover two subtypes: CBHI and CBHII, which cleave off cellobiosyl residues (and to some extent glucose and other cello-oligosaccharides) from the reducing (CBHI) or the non-reducing ends (CBHII) of poly/oligo-cellulose chains. In addition, exoglucanases can also slowly break down microfibril structures and decrease the degree of polymerization of cellulose. Finally, β-glucosidases act upon the cellobiose and cellodextrins and turn them into glucose monomers (Himmel et al., 2007; Lynd et al., 2002; U.S.DOE, 2006).

Most fungal cellulases are composed of a catalytic module and a carbohydrate binding module (CBM) which are connected by a flexible linker (Lynd et al., 2002). The CBM only affects the ability of the enzyme to bind to cellulose, and presumably helps the cellulase to remain in close contact with its insoluble substrate (Lynd et al., 2002).

Due to synergetic interactions between all the components, the efficiency of a cellulolytic system is always higher than the sum of the individual activities of each enzyme. This cooperation is mainly linked to the following four aspects (Lynd et al., 2002):

(1) synergistic interactions between the catalytic domain and CBMs for substrate binding;

(2) endo-exo synergism between endoglucanases and exoglucanases in degrading cellulose into oligomers;

(3) exo-exo synergism between CBHI and CBHII exoglucanases in cleaving cellulose at both reducing and non-reducing ends;

(4) synergistic interactions between exoglucananases and β-glucosidases in converting cellobiose into glucose.
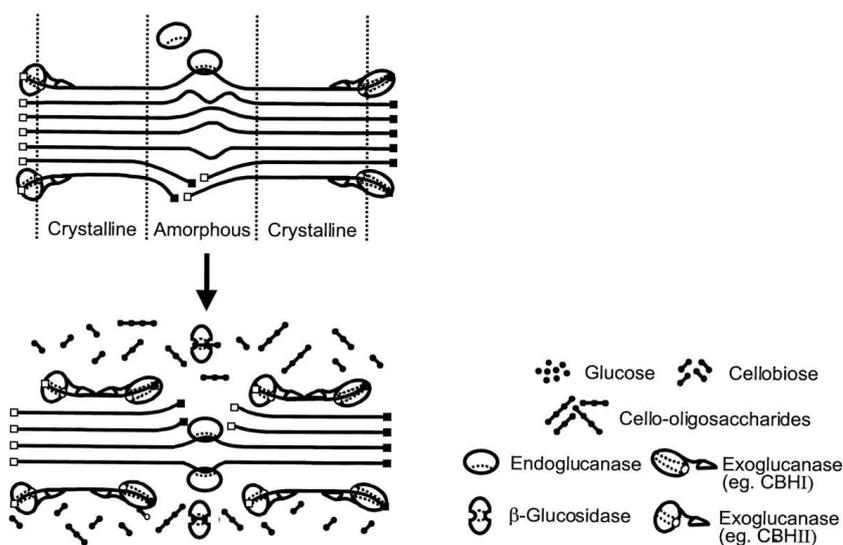
Figure A-8. Functional representation of a free cellulase system in the hydrolysis of cellulose (Lynd et al., 2002).

A.3.1.2 Sources of cellulolytic enzymes and cellulosome

Cellulolytic microorganisms are classified into two categories. The first category, known as the free cellulase system, includes filamentous fungi and actinomycetes bacteria, which secrete cellulases into the surrounding medium (Lynd et al., 2002). Such secretomes can be recovered and used as so-called cocktails for biorefining. The majority of current industrial cellulose cocktails are of fungal origin, especially from *Trichoderma sp.*, *Penicillium sp.* or *Aspergillus sp.* (Lynd et al., 2002; Nieves et al., 1998).

In contrast, cell-bound cellulosomes (2-100 MDa) are found in anaerobic cellulolytic bacteria and fungi (e.g. *Clostridium* and *Bacteroides*) (Lynd et al., 2002). Cellulosomes are remarkably large enzymatic complexes, which not only include cellulolytic modules, but also possess modules of other glycoside hydrolases, polysaccharide lyases and carboxyl esterases. These catalytic domains are linked together by a large noncatalytic scaffoldin protein, consisting of a cluster of cohesins, connected to other functional units such as enzymes and CBMs via dockerins (Gilbert, 2007; Lynd et al., 2002). The cellulosome thus formed is attached to the cell wall surface through an anchoring protein.

The fundamental mechanisms that describe the hydrolysis of cellulose by cellulosomes are identical to those that characterize free cellulase systems, except that the overall activity is much more efficient thanks to powerful synergistic effects (Lynd et al., 2002). The first observation is that the cellulosome itself is a combination of various cellulases, so it provides critical co-localisation of the vital cellulolytic components. Secondly, other enzymatic activities also present on the cellulosome, such as hemicellulases and/or esterases, will enhance the impact of the cellulases by helping to overcome physical hindrances of lignocelluloses (Demain et al., 2005; Gilbert, 2007; Lynd et al., 2002).

A.3.2 Hemicellulases

In order to maximize the use of lignocellulosic biomass and thus to minimize waste and reduce the cost of products, it is important to account for hemicellulose valorization (Bhat, 2000; Wyman, 2007). Nevertheless, the complexity of hemicellulose structures requires a wide diversity of hemicellulases, which can be decomposed into three types according to linkage specificity:

(1) Depolymerizing hemicellulases, such as endo-1,4-β-xylanase and endo-1,4-β-mannanase that cleave the backbone of hemicellulose – acting on heteroxylans and heteromannans respectively- and release short chain oligosaccharides.

(2) Accessory hemicellulases that act on mainchain substitutions and thus reduce spatial impediment and facilitate the action of depolymerizing hemicellulases. Accessory hemicellulases include α-glucuronidase, esterases, β-glucosidase and α-galactosidase.

(3) Oligosaccharide-cleaving hemicellulases convert oligomers into monosaccharides : they include β-xylosidase, β-mannosidase, and α-L-arabinofuranosidase producing xylose, mannose and arabinose respectively.

Hemicellulases act synergistically both between themselves to achieve hemicelllulose hydrolysis, but also with cellulases to better degrade cellulose. Indeed, because hemicellulose and cellulose are intimately associated in biomass, the degradation of one polymer increases the accessibility of the other and vice versa (Kumar and Wyman, 2009b; Murashima et al., 2003; Shallom and Shoham, 2003).

A.3.2.1 Hemicellulases for heteroxylan hydrolysis

The key enzymes required for heteroxylan hydrolysis are endo-1,4-β-xylanase (EC3.2.1.8), α-L-arabinofuranosidase (EC 3.2.1.55), and β-xylosidase (EC 3.2.1.37) (Figure A-9).
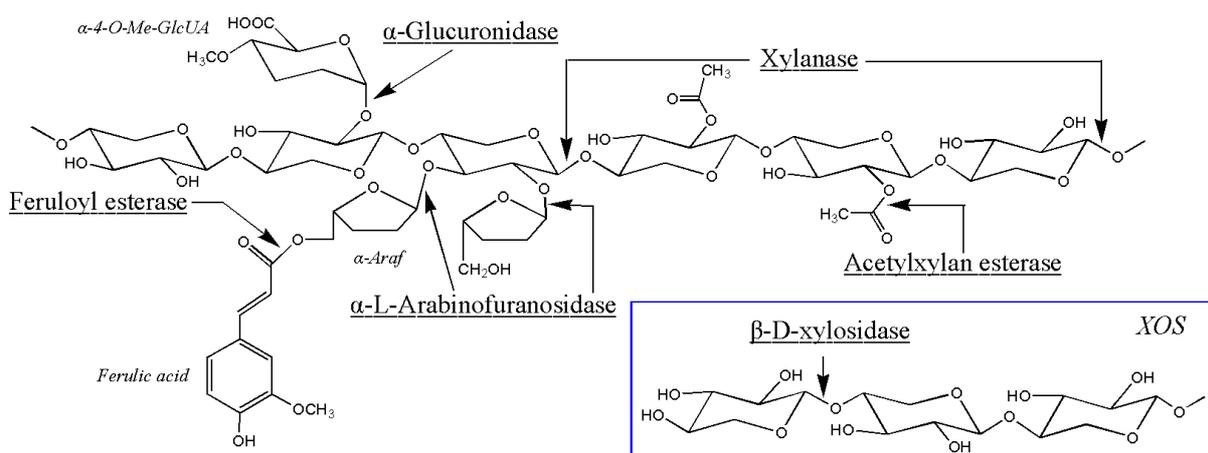
Figure A-9. **Representation of hemicellulase action on the hydrolysis of heteroxylan.** All involved enzymes are underlined. The figure is adapted from (Dodd and Cann, 2009).

Endo-1,4-β-xylanase selectively hydrolyse β-1,4-linkages between xylosyl units that form low to moderately substituted regions in xylan. These enzymes initially liberate short xylo-oligosaccharides as products. Upon prolonged incubation, the main final products of endo-1,4-β-xylanase are xylobiose and xylotriose. Xylanase properties will be detailed in Part B.

α-L-arabinofuranosidases are exo-acting enzymes that release α-(1→2) and/or α-(1→3) linked arabinofuranose residues from soluble products of xylanases (Numan and Bhosle, 2006; Paës et al., 2008; Sørensen et al., 2007). Finally, β-xylosidase attacks the non-reducing ends of xylooligosaccharides and liberates terminal xylosyl residues, thus helping to decrease the inhibition of cellulases by xylo-oligosaccharides (Kumar and Wyman, 2009b).

Other accessory hemicellulases commonly involved in heteroxylan hydrolysis include α-glucuronidase (EC 3.2.1.139) that removes α-1,2-linked glucuronic acid residues in glucuronoxylan and glucurono-arabinoxylan, acetyl-xylan esterase (EC 3.1.1.72) that cleaves acetyl groups substituted at C-2 and C-3 positions of backbone xylosyl residues, and ferulic acid/*p*-coumaric acid esterases (EC 3.1.1.73) that removes phenolic acids linked to arabinosyl residues (Figure A-9) (Beg et al., 2001; Carvalheiro et al., 2008).

In Nature, xylanolytic microorganisms are widespread among bacteria and fungi (Sunna and Antranikian, 1997). However, so far *Bacillus* sp*, Aspergillus niger, Thermomonospora fusca,* and *Trichoderma reesei* have been used as the primary sources of industrial xylanases (Polizeli et al., 2005). Also, genes encoding various hemicellulases have been cloned into *S. cerevisiae* in order to create a recombinant strain suitable for SSF applications (Gorgens et al., 2004; La Grange et al., 2001; Lee et al., 2009).

Like the cellulosome, the xylanosome – a complex xylanolytic system – has also been found in some anaerobic bacteria. An extracellular xylanosome from *Butyvibrio fibrisolvens* H17c has been reported to possess eleven enzyme subunits displaying xylanase activity and three subunits displaying endoglucanase activity (Lin and Thomson, 1991). Likewise, Jiang *et al* found a xylanosome in *Streptomyces olivaceoviridis* that was composed of at least four xylanase units, one CMCase unit and one xylan binding module (Jiang et al., 2006).

A.3.2.2 Hemicellulases for heteromannan hydrolysis

Mannanases also have potential applications for biorefining of softwoods, in which galacto-glucomannans are the predominant hemicelluloses (see section I.1.1 for details). The complete hydrolysis of heteromannans first requires endo-1,4-β-mannanase (EC 3.2.1.78) to depolymerize the backbone, as well as branch-removing enzymes, including α-galactosidase (EC 3.2.1.22), acetylmannan esterase and β-glucosidase. Finally, mannose liberation is accomplished by β-mannosidase (Ademark et al., 1998; Wyman, 2003). Important species for mannanase production include *Bacillus sp.*, *Streptomyces sp.* and *Caldibacillus sp.* (Girio et al., 2010).

A.3.3 Current and future R&D targets and challenges

Reducing the cost of enzymes is a key issue in enzymatic deconstruction of lignocellulosic biomass (Merino and Cherry, 2007; U.S.DOE, 2006), and improving enzymatic efficiency and increasing yield and productivity constitute the two major strategies to achieve cost reduction. Over the past 15-20 years, cellulases have benefited from considerable research focus, with large-scale commercial production now being a reality.

Table A-4. Commercially available lignocellulosic enzyme cocktails on the market.

Product	Manufacturer	State	Production organism	Components
Celluclast 1,5L	Novozymes	Liquid	*Trichoderma reesei*	Cellulases
Cellic Ctec 1/2	Novozymes	Liquid	*unknown*	Cellulases
Cellic Htec 1/2	Novozymes	Liquid	*unknown*	Xylanases
Novozym188	Novozymes	Liquid	*Aspergillus niger*	β-glucosidases
Accellerase 1000/1500	Genencor	Liquid	*Trichoderma reesei*	Cellulases
Accellerase XC	Genencor	Liquid	*Penicillium funiculosum*	Cellulases and xylanases
Accellerase XY	Genencor	Liquid	*Trichoderma reesei*	Xylanases
Accelerase BG	Genencor	Liquid	*Trichoderma reesei*	β-glucosidases
CellulaseA"Amano"3	Amano Enzyme	Liquid	*Aspergillus niger*	Cellulases
Meicelase	Meiji Seika Kaisha	Powder	*Trichoderma viride*	Cellulases

In contrast, studies on hemicellulases have lagged behind, despite the fact that these are also important and that it is necessary to properly valorise hemicelluloses. Therefore, one of today's challenges is to further develop hemicellulases, selecting robustness, efficiency and productivity, and aiming for enzymes that will better penetrate and hydrolyze complex biomass and work in synergy with cellulases.

When one considers today's commercial biomass-degrading cocktails, it is apparent that *Trichoderma reesei* is the main industrial source of these (Table A-4), with the secretome of this fungus being composed of at least two exoglucanases, five endoglucanases and two β-glucosidases (Lynd et al., 2002).

Regarding future improvements of such cocktails, it will be necessary to further increase catalytic turnover of the component enzymes, reduce absorption of enzymes onto, or inhibition, by lignins and phenolic compounds, harmonize pH and temperature optima of the enzyme components, and reduce sensitivity to high temperatures and/or other denaturants and inhibitors. To achieve all of the targets, sophisticated enzyme engineering technologies will be required. These will include both rational engineering and random approaches, which both can be very powerful when the necessary prerequisite conditions are present. For rational enzyme engineering it is vital to possess detailed knowledge of 3D structure, structure-function relationships, fundamental mechanisms and the factors that define enzyme action on biomass. Although, increasingly, methods now exist to provide such information, the role of dynamic protein movements is still under study and the understanding of complex interactions that characterize the action of enzymes on biomass and the synergies between biomass-degrading enzymes is still at an immature stage. Comparatively, random engineering approaches require much less input information. However, the key issue in random engineering is the availability of an appropriate high throughput screening method that is sufficiently well built to provide pertinent output from a given mutant library. This is a considerable challenge because very simple screens are likely to procure oversimplified answers, while highly complex screens can provide a multitude of different solutions, which are then difficult to interpret.

Regarding enzyme production, future advances in this area will have to address issues such as the development of a wider range of industrial host expression systems and/or the improvement of robust gene expression in existing ones, focussing among other things on regulation factors. The development of new production modes is also an attractive target, for example, the co-cultivation of cellulolytic organisms (microbial consortia) producing complementary cellulolytic and hemicellulolytic components (Kumar et al., 2008). Enzyme

production based on solid-state fermentation, using moist agricultural wastes as the carbon source is also an interesting avenue of research. In recent studies, this technique has provided high yields of both cellulases and xylanases (Beg et al., 2001; Dogaris et al., 2009). Finally, further development of autonomous or CBP microorganisms that can both produce their own enzymes and perform bioconversion of biomass intermediates into useful products is to be expected.

Part B Endo- β-1,4-xylanase

B.1 Source and classification

Endo-1,4-β-xylanases (EC 3.2.1.8), commonly known as xylanases, randomly attack and depolymerize β-1,4 linked heteroxylans, producing short xylo-oligosaccharides (XOS) and, in some cases, xylose as main products (Berrin and Juge, 2008; Zimmermann, 1991). Only endo-1,3-β-xylanases (EC 3.2.1.32), rarer than endo-1,4-β-xylanases, hydrolyse β-1,3-linked xylan or β-1,3/β-1,4-mixed linkage xylan, which are mainly found in marine organisms (Deniaud et al., 2003; Frei and Preston, 1964; Okazaki et al., 2002; Turvey and Williams, 1970).

Endo-1,4-β-xylanases are widespread among bacteria and fungi, with the genera *Bacillus, Aspergillus, Penicillium, and Streptomyces* being particularly associated with xylanase production. Usually, fungal xylanolytic organisms prefer acidic growth conditions compared to bacterial ones (Kulkarni et al., 1999) and the production of xylanases in natural hosts is usually induced by the presence of xylan or xylan derivatives in the culture medium (Beg et al., 2001; Prade, 1996). Some fungal species are able to produce multiple xylanases. For example, *Aspergillus niger* and *Trichoderma viride* secrete fifteen and thirteen extracellular xylanases respectively (Collins et al., 2005). Post-translational glycosylation of xylanases is frequently observed among eukaryotic endoxylanases and even some prokaryotic xylanases, from *Bacillus sp.* and *Clostridium stercorarium*, have been reported to be non-covalently associated with sugar chains (Kulkarni et al., 1999; Wong et al., 1988). In all cases, it is hypothesized that the associate carbohydrate groups contribute to the stability of xylanases, offering protection in extreme environments (Kulkarni et al., 1999).

At present time, the most common classification system for glycoside hydrolases (GH) is the CAZy database (http://www.cazy.org/) that organizes enzymes into families according to amino acid sequence similarity (Henrissat and Bairoch, 1993). Unlike the IUBMB enzyme classification system, which is based on the chemical reaction, the CAZy is informationally-rich, because it accounts for enzyme structure and, providing that the biochemical

characteristics of at least one family member are known, it reveals fundamental mechanistic and structural information for whole families. Importantly, the CAZy system pinpoints family relationships among enzymes that display different substrate specificities and thus accounts for divergent evolution. Thus far, up to 125 GH families have been defined, though regular updates will without doubt increase this number. Beyond the family level, CAZy also defines clans, which are composed of CAZy families that display similar 3D architecture. Clans also result from divergent evolution and thus probably reflect the differential adaption of a common ancestor (Henrissat and Bairoch, 1996). The 14 existing clans and their associated folds are summarized in Table B-1.

Table B-1. Clans of glycoside hydrolases.

Clan	Structural fold	GH family
GH-A	(β/α)8	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113
GH-B	β-jelly roll	7, 16
GH-C	β-jelly roll	11, 12
GH-D	(β/α)8	27, 31 36
GH-E	6-fold β-propeller	33, 34 83 93
GH-F	5-fold β-propeller	43, 62
GH-G	(α/α)6	37, 63
GH-H	(β/α)8	13, 70 77
GH-I	α+β	24, 46 80
GH-J	5-fold β-propeller	32, 68
GH-K	(β/α)8	18, 20 85
GH-L	(α/α)6	15, 65 125
GH-M	(α/α)6	8, 48
GH-N	β-helix	28, 49

Endo-1,4-β-xylanases have been classified into CAZy GH families 5, 8, 10 and 11, all of them only comprising a single catalytic domain. Additionally, some bifunctional enzymes bearing xylanase activity have also been found in GH families 16, 43, 52 and 62 (Collins et al., 2005; Henrissat and Bairoch, 1993; Lagaert et al., 2009).

Both GH 5 and GH 10 belong to clan GH-A and thus share similar 3D architecture. However, enzymes from GH 8 and GH 11 belong to clans GH-M and GH-C respectively (Table B-1). The majority of xylanases known so far belong to GH 10 and 11 families, which have been

extensively studied due to their industrial relevance (Collins et al., 2005; Kulkarni et al., 1999).

B.2 Active site and catalytic mechanism

The active site (AS) of glycosidases is made of a catalytic dyad, which is surrounded by binding sites, where substrate recognition and binding happen. The concept of a subsite is used to define the region of the active site that accommodates a single glycosyl unit. Glycoside hydrolases display n subsites. Negatively-labeled (-n) subsites accommodate sugar moieties that are located on the non-reducing part (glycon region) of the glycoside to hydrolyzed, whereas positively-numbered (+n) subsites accommodate sugar moieties that are located on the reducing end (aglycon region). The point of hydrolysis or catalytic site is located at the interface between the glycon and aglycon regions (Figure B-1) (Davies et al., 1997). Like the vast majority of glycoside hydrolases, xylanases perform catalysis via the intervention of two acidic amino acids, with one acting as an acid–base and the other as an electrophile/nucleophile (Zvelebil and Sternberg, 1988).
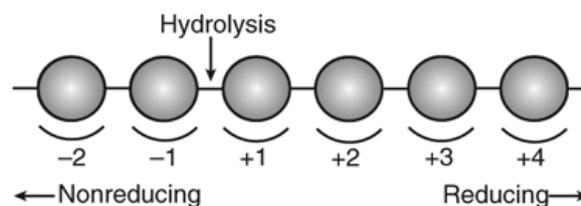
Figure B-1. Schematic representation of substrate binding subsites in glycosidases (Dodd and Cann, 2009).

Among glycoside hydrolases and, *a fortiori*, xylanases, there are two major mechanisms for hydrolysis, differing in whether they lead to overall retention or inversion of the configuration of the anomeric carbon at the cleavage point (Henrissat and Bairoch, 1996). Retaining catalysis employs two carboxylic amino acids, designated as the acid/base and the nucleophile, located about 5.5 Å apart. These residues provide assistance to the reaction, achieved via two separate chemical steps, known as a double-displacement mechanism (Figure B-2.A). In the first step, the concomitant action of two carboxylic amino acids, one acting as a Lewis acid and the other as a nucleophile, leads to the formation of a covalent glycosyl-enzyme intermediate displaying inverted anomeric configuration and a leaving group (aglycon product). In the second step, the now deprotonated carboxylate, acting as a

general base, captures a proton from an incoming water molecule. This simultaneously hydroxylates the anomeric carbon of the covalent glycosyl-enzyme intermediate, once again inverting the anomeric configuration. Upon release of the glycon product, which now displays the same anomeric configuration as the substrate, the enzyme is recycled into its ground state (McCarter and Withers, 1994; Sinnott ML, 1990).

Inverting glycoside hydrolases also employ two carboxylic amino acids (Figure B-2.B), which perform catalysis in a single step. Like in the double displacement mechanism, one carboxylic residue provides acid catalysis, provoking leaving group departure, while another carboxylic residue provides base-assistance to the nucleophile attack of a water molecule, which occurs on the opposite side of the sugar ring. Overall, this leads to the cleavage of the glycosidic bond and the inversion of the anomeric configuration. The key feature that differentiates retaining enzymes from inverting ones is the distance that separates the two catalytic carboxylic amino acids. In inverting enzymes this distance is in the range of 6.5-9.5 Å (McCarter and Withers, 1994).
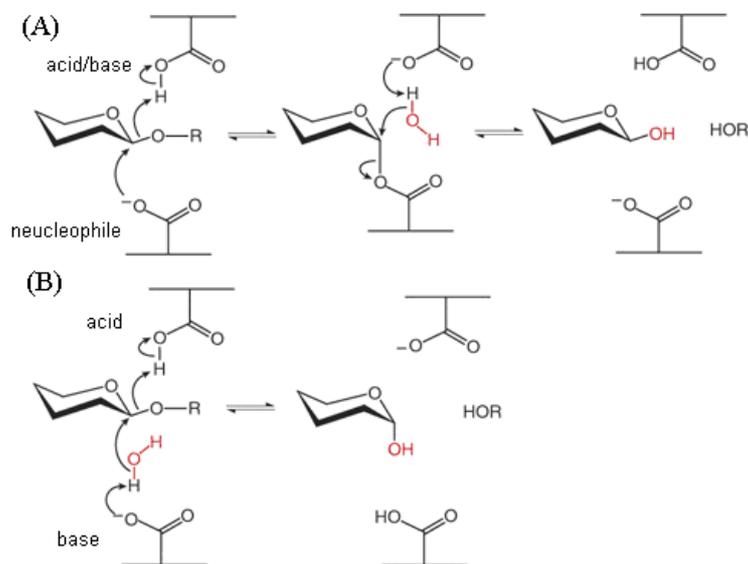
Figure B-2. **Retaining (A) and inverting (B) mechanisms of glycosidases.** Adapted from (Dodd and Cann, 2009).

As indicated earlier, the catalytic mechanism is generally conserved within each GH family (Henrissat and Bairoch, 1996). Regarding xylanases, enzymes in GH families 5, 10, 11 and 16 act via the retaining mechanism using two glutamic acid residues as the catalytic dyad. In

contrast, GH 8 and 43 xylanases perform hydrolysis via an inverting mechanism and employ a glutamate and an aspartate residue respectively (Berrin and Juge, 2008; Collins et al., 2005; Jeffries, 1996).

B.3 Industrial application

In the past decades the industrial applications of xylanases have mainly concerned paper pulping and the improvement of the nutritive value and technological properties of human foodstuffs and animal feeds (Collins et al., 2005; Kulkarni et al., 1999).

In the pulp and paper industry, xylanases are used to replace chlorine-mediated bleaching of paper pulp. Here, xylanases disrupt the intimate contacts that are made between cellulose and lignin, which is mediated by heteroxylans (Ragauskas et al., 1994; Roncero et al., 2005; Viikari et al., 1994). This enzymatic bleaching process is usually carried out at high temperatures (50-70°C) and under alkaline conditions. Therefore thermo-alkaliphilic xylanases are preferred for this process. Compared to chemical bleaching, xylanase usage is not only environmental friendly, but also allows the recovery of lignins after treatment, thus reducing the overall cost (Kulkarni et al., 1999; Olsson et al., 2007).

In the food industry, xylanases are used for the production of beverages, such as fruit and vegetable juices, beer and wine (Bhat, 2000; Collins et al., 2005). The use of xylanases improves the maceration step and can also play a useful role in juice clarification and for the reduction of liquid viscosity. Such effects improve the yield and the efficiency of the extraction and filtration steps. A second important application for xylanases in the food industry is within the baking sector (Autio et al., 1996; Courtin and Delcour, 2001; Courtin and Delcour, 2002). Here, xylanases have a significant impact on dough preparation and proofing for bakery products, where they improve the elasticity and make the dough soft and slack and thus allow easier handling and provide larger loaf volumes. Since dough preparation is often carried out at around 35°C, xylanases having high activity at low or intermediate temperatures are more suitable (Collins et al., 2005; Collins et al., 2006).

Another important application of xylanases is in the animal feed industry (Beauchemin et al., 2003; Choct, 2006; Francesch and Geraert, 2009), where they are used as additives in feed

for non-ruminant animals that lack appropriate digestive enzymes. When applied to feed such as wheat bran or wheat flour, xylanases act upon the non-starch polysaccharide (NSP) components. The depolymerization of NSP leads to a reduction in viscosity (Beauchemin et al., 2003; Vahjen et al., 1998), which favours digestion, and generates short xylo-oligosaccharides, which are thought to have a beneficial prebiotic effect on digestive microbiota (Hsu et al., 2004; Swennen et al., 2006). Preferably, xylanases that are employed as feed additives should display a certain degree of thermostability, because ideally the xylanase component is introduced into the feed before pelleting, which incurs a short duration rise in temperature to 70-95°C. Also, it is beneficial if the xylanases are highly active in the acidic environment of the animal's stomach (approximately 40°C and pH 4.8) (Collins et al., 2005).

Increasingly, the importance of xylanases in biorefinery processes is being taken into account. Therefore, in the longer term, it is certain that biorefining will become a major application sector for xylanases. Specifically, xylanases (and other hemicellulases) will probably be recognized as a key component of cellulase cocktails that are used for the deconstruction of biomass and the production of fermentable sugars. Accordingly, future improvement of these cocktails will also focus on the improvement of the xylanase components and will aim at overall goals such as cost reduction and robustness. Regarding cost reduction this will be achieved by improving enzyme production and increasing enzyme activity on complex substrates. More details on these goals and challenges are given in the Chapter A.3.

B.4 Xylanases from GH11 family

B.4.1 Biochemical properties

A key feature of the CAZy family GH11 is that it is exclusively composed of endo-1,4-β-xylanases that hydrolyze xylans. In other terms, no other activities have been so far associated with this family. The catalytic domain of GH11 enzymes is characterized by a low molecular weight, between 20 – 26 kDa, and relatively high pI and a wide optimal pH window, in the range of 2.0 – 9 .0. In fact, most bacterial GH11 xylanases present an optimal pH above pH 5.5, in contrast to the majority of fungal xylanases which are more acidic: the most acidophilic xylanase derived from *Aspergillus kawachii* has an optimal pH at 2.0 (Ito et

al., 1992). Regarding catalysis, GH11 family members hydrolyse β-1,4 linkages between D-xylosyl units via a retaining mechanism.
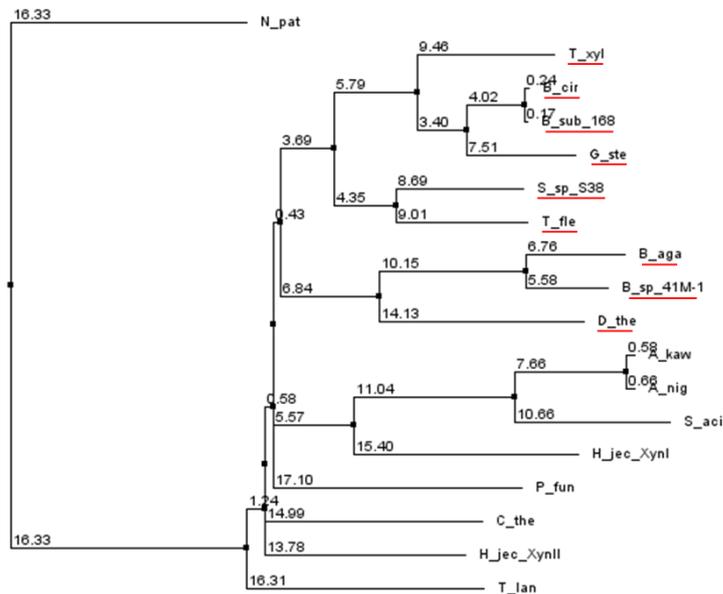
Figure B-3. Phylogenetic tree of eighteen GH11 xylanases whose 3D structures are known. Bacterial xylanases are underlined in red. The tree is calculated using the neighbour joining algorithm, and the distance is based on the percentage identity between the two sequences. The numbers at each node are distance values and the tree is built by Jalview (Waterhouse et al., 2009).

B.4.2 Overall structure

To date, 24 GH11 xylanases have been assigned good quality, published 3D structures by X-ray crystallographic studies. Table B-2 mentions 18 of these. Nine are of bacterial origin, and the other nine are of fungal origin. A phylogenetic analysis of this set of xylanases (Figure B-3) reveals several distant branches, which indicate that these sequences are sufficiently diverse to be representative of GH11 family diversity. In addition, branch lengths suggest that the bacterial group is more inter-related from an evolutionary point of view than the fungal group of GH11 xylanases, which is in accordance with previous alignment studies on eighty-two GH11 sequences (Sapag et al., 2002). In other words, the evolution of fungal GH11 xylanases is more divergent. As shown on the multiple sequence alignment (Figure B-5), the highly conserved residues (>80% identity) of GH11 xylanase account for around 28% of total length.

The secondary structure of the catalytic domain of GH11 xylanases consists of two large β-pleated sheets and one α-helix (Figure B-4.A). β-sheet A includes a maximum of six antiparallel β-strands -from A1 to A6- and β-sheet B contains nine β-strands, from B1 to B9. In some enzymes, the N-terminal region is shorter and β-strand A1 is then replaced by a loop.

Table B-2. Selection of eighteen crystallized GH11 xylanases.

Abbreviation	Organisms	PDB code	Genbank	Reference
Bacterial origin				
T_xyl	*Thermobacillus xylanilyticus*	Cryst	CAJ87325.1	(Paës, 2005)
B_cir	*Bacillus circulans*	1XNB	ACR92575.1	(Ludwiczek et al., 2007)
B_aga	*Bacillus agaradhaerens AC13*	1QH7	CAB42305.1	(Sabini et al., 1999)
B_sp_41M-1	*Bacillus sp. 41M-1*	2DCJ	AAS31755.1	(Ihsanawati *et al*, unpublish)
B_sub_168	*Bacillus subtilis subsp. subtilis str. 168*	1AXK	AAA22897.1	(Ay et al., 1998)
D_the	*Dictyoglomus thermophilum RT46B.1*	1F5J	AAC46361.1	(McCarthy et al., 2000)
G_ste	*Geobacillus stearothermophilus 236*	Cryst	AAB72117.1	(Cho and Choi, 1995)
S_sp_S38	*Streptomyces sp. S38*	1HIX	CAA67143.1	(Wouters et al., 2001)
T_fle	*Thermopolyspora flexuosa DSM43186 (ATCC35864)*	1M4W	AAO97628.1	(Hakulinen et al., 2003)
Fungal origin				
A_kaw	*Aspergillus kawachii*	1BK1	AAC60542.1	(Fushinobu et al., 1998)
A_nig	*Aspergillus niger*	1UKR	CAA01470.1	Krengel_1996
C_the	*Chaetomium thermophilum CBS730.95*	1H1A	AAR67679.1	(Hakulinen et al., 2003)
H_jec_XynI	*Hypocrea jecorina*	1XYN	AAE16058.1	(Törrönen and Rouvinen, 1995)
H_jec_XynII	*Hypocrea jecorina*	1ENX	AAB29346.1	(Törrönen and Rouvinen, 1995)
N_pat	*Neocallimastix patriciarum*	2C1F	CAA46498.1	(Vardakou et al., 2008)
P_fun	*Penicillium funiculosum*	1TE1	CAC15487.1	(Payan et al., 2004)
S_aci	*Scytalidium acidophilum*	3M4F	AAQ22691.1	(Michaux et al., 2010)
T_lan	*Thermomyces lanuginosus SSBP / ATCC 46882*	1YNA	AAB94633.1	(Gruber et al., 1998)

The structural elements of GH11 xylanases pack closely to fold into a β-jelly roll tertiary structure, belonging to clan GH-C. This structure has been compared to a right hand and different structurally regions have been named accordingly (i.e. fingers, palm and thumb). As shown in Figure B-4.B, fingers are made up of β-sheet A and part of β-sheet B, the palm is formed by the α-helix and a twisted β-sheet B, the thumb is made of the loop between

strands B7 and B8 and finally the cord is a long loop between B6 and B9 (Purmonen et al., 2007). The active site is a cleft located at the inner surface of the palm, and is partially surrounded by the fingers, cord and thumb (Ludwiczek et al., 2007; Purmonen et al., 2007). Except for a few GH11 xylanases from *Cellulomonas fini, Clostridium thermocellum and Bacillus sp.*, most GH11, members don't have additional carbohydrate binding domains and only present a single catalytic domain.

(A)

(B)
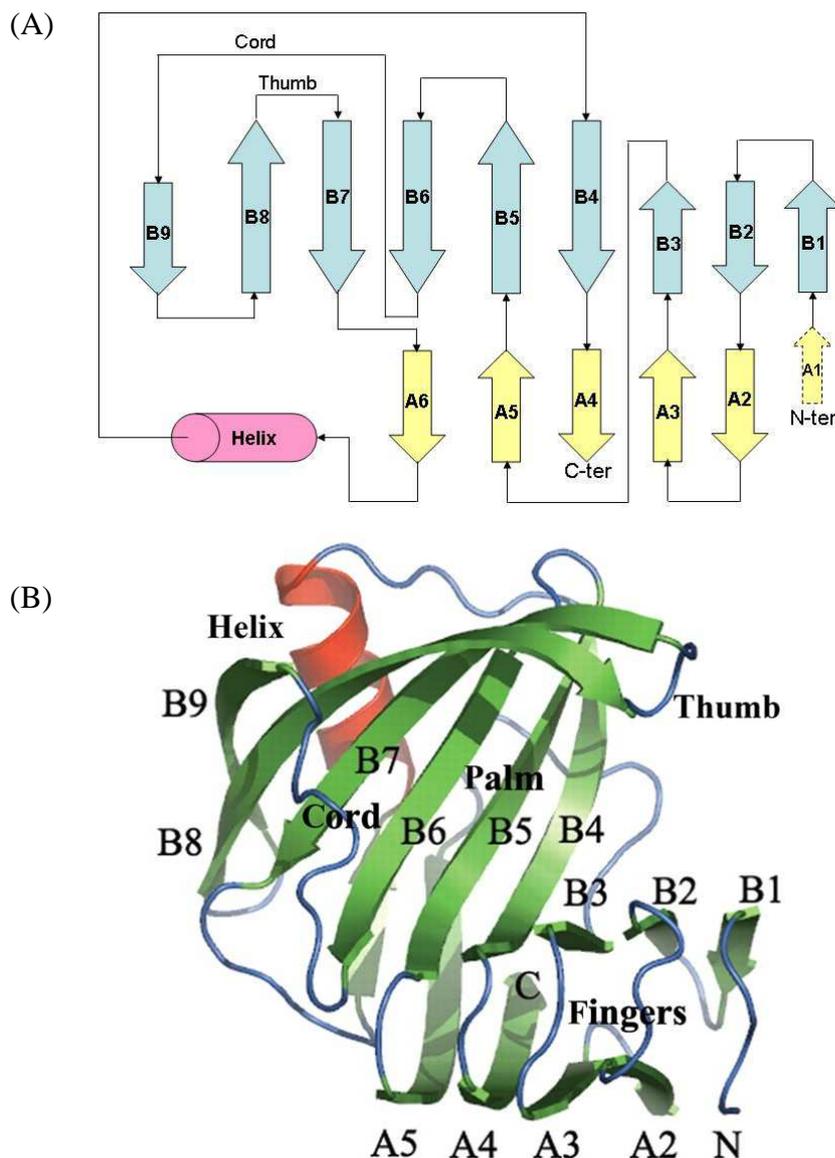
Figure B-4. The typical topology diagram (A) and cartoon representation (B) of GH11 xylanases. The letters A1-A6 and B1-B9 represent numbered β-strands for β-sheet A and β-sheet B respectively. N and C represent the N-terminus and C-terminus in the figure B. In addition, the relevant thumb, palm, fingers and cord regions in the tertiary structure are also indicated. Figure A is adapted from (Paës, 2005) and figure B originates from (Purmonen et al., 2007).
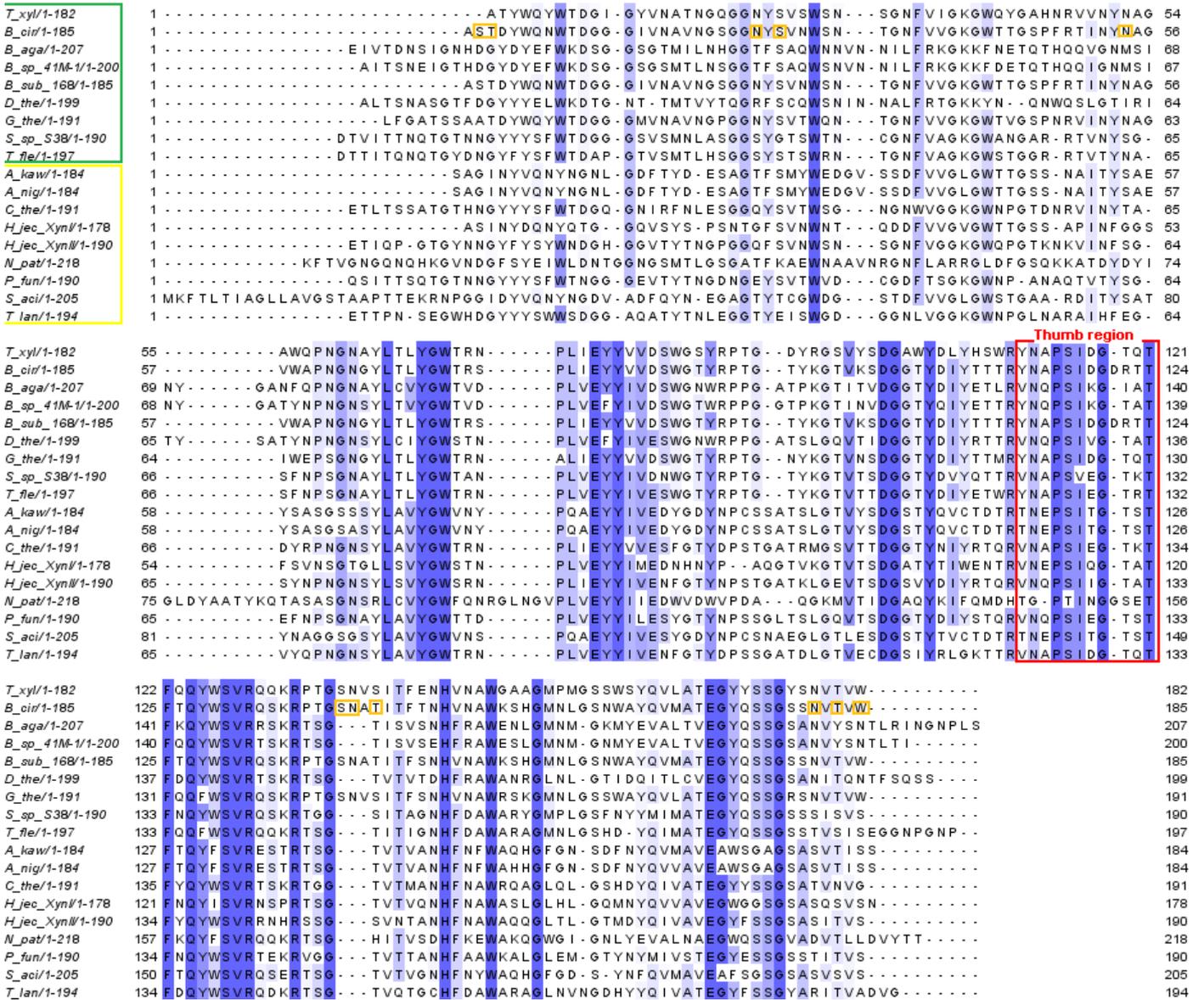
Figure B-5. **Multiple sequence alignment of catalytic domains (without signal peptide) of eighteen GH11 xylanases possessing a resolved 3D structure.** Database and biographic references of the xylanases is presented in Table B-2. The nine bacterial xylanases and nine fungal xylanases are indicated by green and yellow boxes respectively. Conserved residues (>80% identity) are highlighted using a blue background, in which the deeper colour corresponds to higher conservation. The thumb region in the sequence is indicated by a red box. The residues that constitute the secondary binding site of the *Bacillus circulans* xylanase are highlighted using orange boxes along the sequence of B_cir. The alignment was performed and edited using Jalview (Waterhouse et al., 2009).

B.4.3 Structure-function relationships

B.4.3.1 Highly conserved region

The sequence alignment of eighteen representative GH11 xylanases reveals fifty-two highly conserved residues with more than 80% identity (Figure B-5). Conserved residues are found around the catalytic site, on the loops linking adjacent β-sheets, and on the inner side of the α-helix (Figure B-6.A). Logically, invariant residues in the catalytic site maintain the tight enzymatic specificity of GH11 xylanase. Other conserved residues can be regarded as ''linker'' elements, that are localized at the junctions between the structural elements. These contribute to the definition of the overall protein fold and probably ensure the adoption of an active conformation. Figure B-6.B and C reveals that most of the conserved residues cluster together and that more than half of these clusters are located in buried, solvent inaccessible regions. An interesting feature is the high conservation of amino acids observed in the thumb region. This suggests that this area might play an important role in enzyme function.
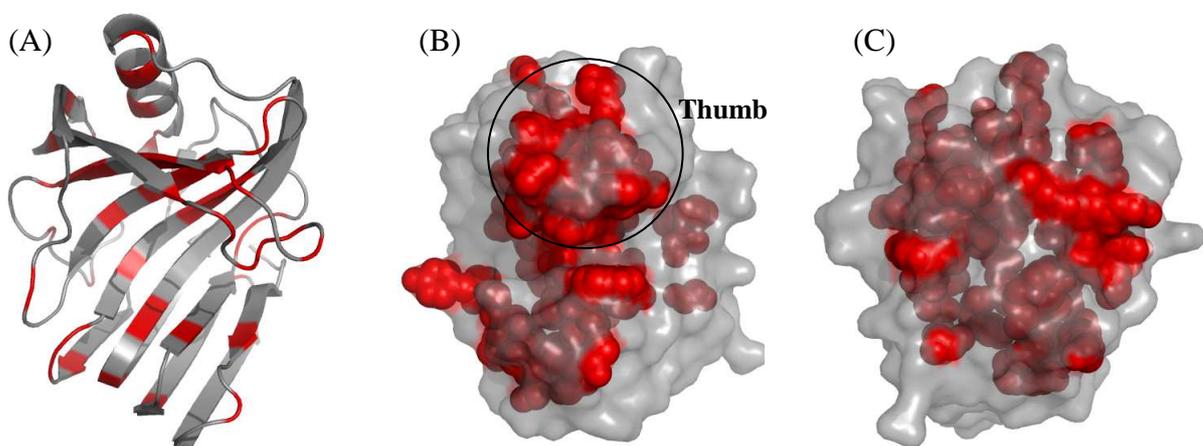
Figure B-6. **(A) Cartoon representation of the GH11 xylanase from *T. xylanilyticus* indicating the positions of highly conserved residues (red colour). Front (B) and back (C) views of conserved residues (bright red for exposed residues and dark red for buried residues) in the surface presentation.** The position of thumb region is highlighted by a circle.

B.4.3.2 Active site and substrate specificity

The active site of GH11 xylanases has been described as a deep cleft that generally spans three glycon subsites and two or three aglycon subsites (Vardakou et al., 2008). Like other hydrolases, the active site of GH11 xylanases is mainly composed of aromatic residues (Figure B-7), tightly packed together and providing a hydrophobic environment for substrate accommodation. In addition, the presence of aliphatic hydrophobic side chains not only enhances the overall hydrophobicity of the active site, but also provides possibilities for Van der Waals interactions with xylan substrates. The two conserved glutamate residues responsible for catalysis are located at the centre of the active site.
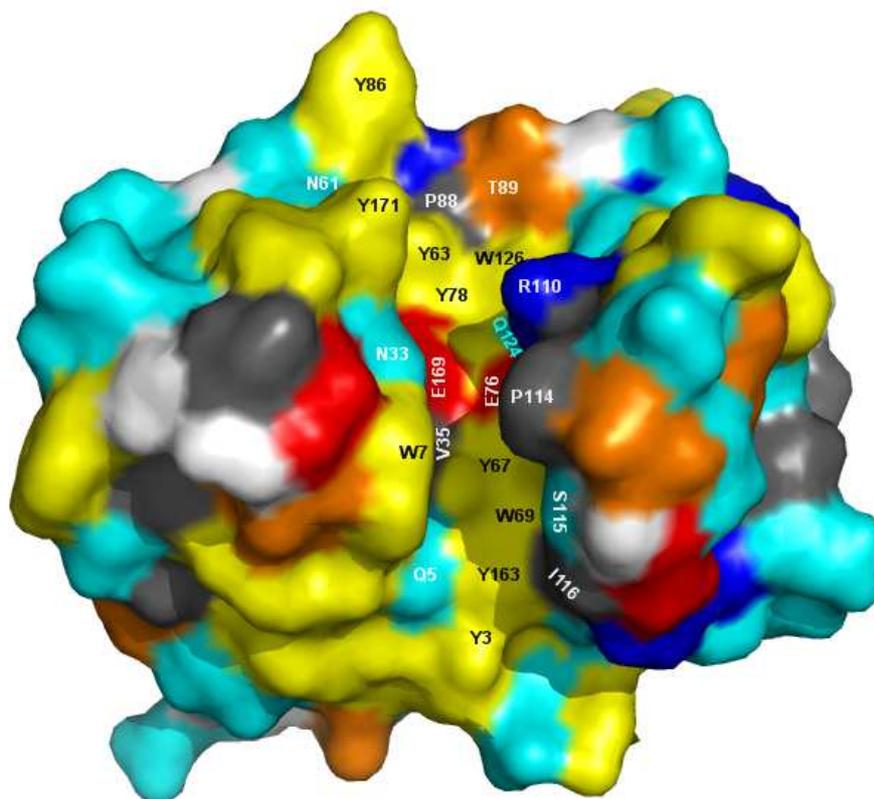
Figure B-7. Active site residues in the surface representation of *T. xylanilyticus* GH11 xylanase. The residues are labelled with name and position number. Different colours are used to represent the various residue groups: white – glycine (G); yellow - aromatic hydrophobic residues (P, Y and W); grey – aliphatic hydrophobic residues (A, I, L, V and P); orange – amphipathic residues (T, L and M); red – hydrophilic residues with negative charge (D, E and C); cyan – neutral hydrophilic residues (S, N and Q); blue – hydrophilic residues with positive charge (R and H);. The amino acid classification is based on (Petsko and Ringe, 2004).

To understand the substrate binding in the active site, it is important to describe substrate specificity and related biochemical features. In general, the most important forces for enzyme-substrate are hydrogen bonds, and Van der Waals interactions (Petsko and Ringe, 2004). To investigate enzyme-substrate interactions in xylanases, X-ray crystallization has often been the method of choice (Havukainen et al., 1996; Janis et al., 2005; Sabini et al., 1999; Sidhu et al., 1999; Vardakou et al., 2008), sometimes coupled to mutational studies (Vandermarliere et al., 2008; Wakarchuk et al., 1994).

Among the different subsites that comprise the active site of GH11 xylanases, subsites -2 and -1 subsites have been the major focus of characterization. This is because they make tight contacts with the substrate and are essentially invariant in GH11 xylanases (Havukainen et al., 1996; Janis et al., 2005; Sabini et al., 1999; Wakarchuk et al., 1994). In contrast, little is known about aglycon interactions and the substrate binding across the -1/+1 subsites. The binding in the aglycon subsites is usually suggested to be energetically unfavourable, because it is believed that weak binding is necessary for the efficient release of the products (Janis et al., 2005; Vandermarliere et al., 2008). Moreover, obtaining a crystal structure of an active enzyme with a substrate molecule bound in the -1/+1 subsites is technically impossible, because of its fast hydrolysis (microsecond timescale). Nevertheless, so far, two publications have revealed some information about aglycon subsites binding using X-ray crystallography (Vandermarliere et al., 2008; Vardakou et al., 2008), although only Vandermarliere et al actually solved the complex structure with a substrate across the -1/+1 subsites, using an inactive mutant.

In the following, the details of xylanase-ligand interactions are presented for each subsite, and summarized in Table B-3.

(1) -3 subsite: only weak electron densities have been described for this subsite, which indicates that the binding does not involves strong forces (Vandermarliere et al., 2008; Vardakou et al., 2008). Indeed, the xylosyl moiety in -3 subsite is located just outside the active site, so if hydrogen bonds are formed, they are either weak or are mediated by water molecule intermediates with a flanking isoleucine residue (Ile116 in Figure B-7).

(2) -2 subsite: the α-face of the xylosyl unit always stacks against the indole ring of ipsilateral tryptophan in alkaline xylanases or tyrosine in acidic xylanases (Trp7 in Figure B-7) (Sabini et al., 1999; Wakarchuk et al., 1994). In addition, the hydrogen bonds formed between O2 and/or O3 of xylosyl unit and vicinal tyrosine residues (Tyr63 and Tyr167 in Figure B-7) are also highly conserved (Table B-3).

(3) -1 subsite: here there are hydrogen bond interactions between the sugar and several amino acids (Glu76, Pro114 and Arg110 in Figure B-7), which are,

- catalytic (nucleophile) glutamate residue

- carbonyl group of the adjacent proline residue

- positively charged arginine residue in the thumb region.

All of these contribute equally to the stabilization of the xylosyl moiety's twisted boat ($B^{2,5}$), which is necessary for catalysis (Sabini et al., 1999). Additionally, a pH-related residue Asn/Asp (Asn33 in Figure B-7) can also be hydrogen-bonded to the endocyclic oxygen and/or O1 position of the xylosyl unit and a 90° conformational rotation of this residue before/after ligand binding has been found in *B. subtilis* and *A. niger* xylanases (Vandermarliere et al., 2008). It has also been proposed that steric hindrance caused by the conserved hydrophobic residue Val/Ile (Val35 in Figure B-7), adjacent to the acid/base catalytic residue, plays an important role in preventing the accommodation of a glucose at -1 subsite for GH11 xylanases (Sabini et al., 1999).

(4) +1 subsite: the study of the crystal structure of a *B. subtilis* xylanase-xylotriose complex has clearly indicated that the xylosyl moiety at +1 subsite is presented in a twisted boat conformation (Vandermarliere et al., 2008). Even though very few crystal structures reveal +1 subsite binding, conserved hydrogen bonds between the catalytic (acid/base) glutamate residue and two neighbouring residues (Asn124 and Tyr78 in Figure B-7) are likely to occur, as these three residues are strictly conserved in GH11 family (Vardakou et al., 2008).

(5) +2 and +3 subsites: the binding affinity in these two subsites is much weaker than in any other parts of the active site. The xylosyl moieties are speculated to make hydrogen bonds with the threonine and/or asparagine residues lying at the end of the aglycon cleft (Thr89 and Asn61 in Figure B-7) (Vardakou et al., 2008). In addition, a stacking interaction is thought to occur, which involves a Trp/Tyr side chain (Tyr86 in Figure B-7).

In the hydrolysis of heteroxylan, GH11 xylanases prefer to attack unsubstituted regions of the polymer backbone due to the stereo conformation of -2, -1 and +1 subsites, which are too narrow to accommodate main-chain substitutions, especially at the -2 subsite (Gruber et al., 1998; Pollet et al., 2010; Vandermarliere et al., 2008). Since the predominant substrate binding occurs at +1 and glycon subsites, xylobiose and xylotriose are the major end products of hydrolysis (Pollet et al., 2010).

However, the analysis of reaction products released from decorated xylans suggests that at least some GH11 xylanases might accommodate/tolerate 4-methyl-glucuronic acid substitution of xylosyl moieties bound in the -3 and +2 subsites (Katapodis et al., 2003; Kolenov et al., 2006). In addition, a structure of a co-crystal clearly indicated that a xylanase, isolated from an environmental source, was able to accommodate an α-1,3 linked arabinosyl ramification, itself bearing a *o*-5-linked ferulate moiety, in the +2 subsite (Vardakou et al., 2008). Nevertheless, whether GH11 xylanases can commonly accept side chains at +2 and -3 subsites is still unclear, as the identified enzyme-substrate side chain interactions are rarely reported.

Table B-3. **Summary of binding interactions in GH11 xylanase-ligand crystal complexes.** The hydrogen bonds are shown in dots that link atoms from sugar to amino acid residue, the stacking interaction and covalent bonds are noted in brackets.

Organism	Ligand	PDB code	-3	-2	-1	+1	+2	+3	Reference
Bacillus circulans	Xylobiose	1BCX		W9 (stacking) O2··OH (Y69) O3··OH (Y166)	C1··OE1 (E78) C1··OH (Y80) O1··OE1 (E172) O2··OE2 (E78) O2··NE (R112) O3··O (P116) O3··NH2 (R112)				(Wakarchuk et al., 1994)
Trichoderma reesei (Xyn II)	2,3-epoxypropyl-β-D-xyloside (X-O-C3)	1REF		W18 (stacking) O1··OH (Y77) O2··OH (Y77) O2··OH (Y171) O3··OH (Y171)					(Havukainen et al., 1996)
Bacillus agaradhaerens	2-deoxy-2-fluoro-4-O-β-D-xylopyranosyl-α-D-xylopyranose	1QH6		W19 (stacking) O2··OH (Y85) O2··NH2 (R49) O3··NH1 (R49) O3··OE2 (E17)	C1 - OE2 (E94) (ester bond) F2··NH (R129) O3··NH2 (R129) O3··O (P133) O5··OH (Y85)				(Sabini et al., 1999)
Chaetominum thermophilum	Methyl 4,4^{II},4^{III},4^{IV}-tetrathio-α-xylopentoside (S-Xyl5-Me)	1XNK		O2··OH (Y78) O2··OH (Y172) O3··OH (Y172)	CH3O··OE2 (E178)				(Janis et al., 2005)
Environmental source, extracted from soil (EnXyn11A)	Xylobiose (glycon subsites); Ferulic acid-1,5-arabinofuranose-α1,3-xylotriose (FAX$_3$) (aglycon subsites)	2VGD	O4··HOH··O (I132)	W22 (stacking) O1··OE2 (E89) O1··OH (Y80) O2··OH (Y80) O2··OH (Y175) O3··OH (Y175)		O2··OH (Y76) O3··OH (Y91) O3··NH2 (Q140) O4··OE2 (E181) O4··ND2 (N48)	O5··ND2 (N74)	W99 (stacking) O3··ND2 (N74)	(Vardakou et al., 2008)
Bacillus subtilis	Xylotriose (-3, -2 and -1 subsites in asymmetric unit A, and -2, -1 and +1 subsites in unit B)	2QZ3 and 2Z79	O4··O (I118)	W9 (stacking) O1··O (P116) O2··OH (Y69) O2··OH (Y166) O3··OH (Y166) O3··NE2 (Q7) O5··O (P116) O5··N (P116)	O1··OD1 (N35) O2··OE1 (E78) O2··OE2 (E78) O2··NE (R112) O3··O (P116) O3··NH2 (R112) O3··NE (R112)	O1··OH (Y80) O2··N (G173) O3··OH (Y80)			(Vandermarliere et al., 2008)
Aspergillus niger	Xylotriose	2QZ2		Y10 (stacking) O1··O (P119) O2··OH (Y70) O2··OE1 (Q8) O3··NE2 (Q8) O5··O (S120) O5··N (S120)	O1··OD1 (D37) O1··OD2 (D37) O1··OH (Y81) O2··OE1 (E79) O2··OE2 (E79) O2··NE2 (N129) O3··O (P119) O5··OD2 (D37)				(Vandermarliere et al., 2008)

B.4.3.3 Secondary binding site

As mentioned above, the majority of GH11 xylanases are composed of a single catalytic domain. However, a xylan-specific secondary binding site (SBS) on the surface of the catalytic domain has been discovered and identified in the xylanase from *Bacillus circulans* (B_cir in Table B-2) (Ludwiczek et al., 2007). Using the "right-hand" analogy to describe the location of this binding site, the SBS crosses the 'knuckles' of the hand. In addition, the SBS is formed by a line of flanking asparagine, serine and threonine residues and is terminated by a C-terminal tryptophan, which provides hydrogen bond and aromatic stacking interactions. Finally, although the SBS that have been described are located on the outer surface of the enzyme, this feature appears to be linked to the active site (located on the 'palm side' of the 'hand') via a small shallow groove (Figure B-8).
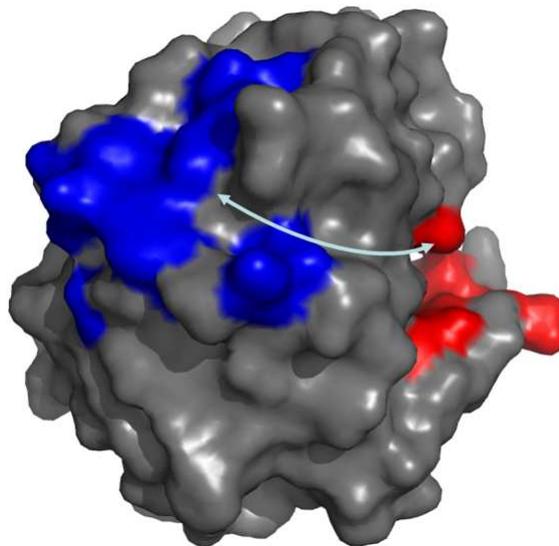
Figure B-8. Surface representation of the secondary binding site (blue) and active site (red) of GH11 xylanases from *B. circulans* (PBD 1XBN). The arrow indicates the shallow groove that links the SBS and the active site. Adapted from (Ludwiczek et al., 2007).

In *B. circulans* xylanase, the scope of the SBS allows it to span three to four xylosyl units. Unlike the active site, the SBS does not participate in any additional conformational change in the binding or release of the substrate, because its association/disassociation rate is measured ten-fold higher than that for the active site. It has been hypothesized that the SBS and the active site might associate independently to short XOS, but corporately bind longer xylan polymers (Ludwiczek et al., 2007). Ludwiczek *et al* (2007) have inferred the

cooperation between the active site and SBS enhances the binding affinity and specific activity of *B. circulans* xylanase towards both soluble and insoluble xylans.

However, except for Ser27, Thr143 and Thr183, most of SBS residues in *B. circulans* xylanase are not conserved in GH11 family, as revealed from an alignment of 473 structure-known GH11 xylanase sequences (Figure B-5 shows the results of a partial alignment). Nevertheless, through crystallographic study, Vandermarliere *et al* (2008) have identified SBS in two xylanases from evolutionarily distant species – *Bacillus subtilis* and *Aspergillus niger*. This suggests that the function of the SBS might be conserved, but involves different amino acids according to the exact biological origin of the enzyme. Overall, it is possible to speculate that the SBS constitutes an alternative strategy to the addition of a carbohydrate binding module.

C.4.3.4 Thumb region

The thumb region in the structure of GH11 xylanase is composed of a loop that links the B7 and B8 β-strands and is a unique feature in GH11 xylanases when compared to other enzymes that display jelly-roll folding. This thumb loop is formed by 11 residues (from Tyr111 to Thr121 numbered in Figure B-9), where the invariant Asn112 and Thr121 constitute the base of the thumb and the highly conserved motif Pro114-Ser115-Ile116-X117-Gly118 (X can be any residue) form the tip. Structurally, the thumb displays a typical hairpin topology with a type I β-turn (Ser115 – Gly118) and six internal hydrogen bonds. The presence of Asn112 terminates the series of internal hydrogen bonds and sharply distorts the thumb downwards, forming a half-folded shape.

The thumb loop is structurally and functionally important for catalysis. Previous structural studies have revealed that the thumb region partially closes the glycon side of active site in the presence of the ligand (Havukainen et al., 1996), and causes a steric hindrance to avoid the accommodation of xylan side chains (Paës et al., 2007; Vandermarliere et al., 2008). In addition, the thumb position determines the width of the cleft, the narrowest part being formed on one side by the conserved Pro residue (thumb tip) and the opposite side by a Trp residue (Trp7 in *T. xylanilyticus* xylanase). Moreover, thumb residues Pro114 and Ser115 (numbering in Figure B-9) appear to be in direct contact with the substrate at -1 and -2

subsites via hydrogen bonds (Table B-3) (Sabini et al., 1999; Vandermarliere et al., 2008; Wakarchuk et al., 1994).
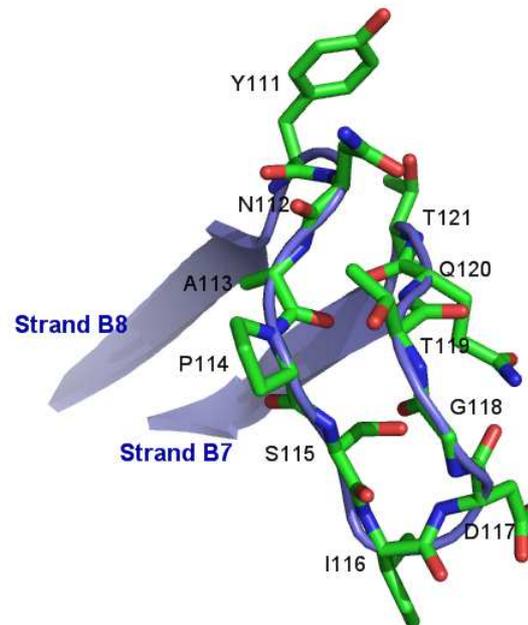
Figure B-9. Stick and cartoon representation of thumb region in GH11 xylanase from *T. xylanilyticus*.

The mobility of the thumb region during catalysis is thought to be an intrinsic feature of this element and an essential characteristic of GH11 xylanases. It has been hypothesized that upon substrate binding, the thumb leaves a ground state position, closing downwards onto the substrate. Supposedly, reopening of the closed conformation occurs during the release of reaction products, although precise details are not yet available (Muilu et al., 1998; Paës et al., 2007; Pollet et al., 2009). Pollet *et al* (2009) have inferred that the possible movement of the thumb proceeds through three steps: first, the thumb adopts a loose and open conformation as the substrate enters into the active site; then the thumb moves downwards into the cleft to enhance the stabilization of the ligand; finally, the thumb returns to its initial position, the tip residue (Ile116 in Figure B-9) aiding to push the product out of the active site to diffuse away. The mobility of the thumb highly depends on three residues: Arg110, Asn112 and Thr124 at the base of the thumb (numbering in Figure B-9) (Paës et al., 2007; Pollet et al., 2009).

As mentioned earlier, GH family 11 together with family 12 constitutes clan GH-C (Table I-5), and therefore enzymes in these two families share the same β-jelly roll architecture. However, GH12 endoglucanases are characterized by wider substrate specificity, acting on both cellulose and xylan in contrast to GH11 xylanases (Shimokawa et al., 2008; Vincken et al., 1997). Importantly, in GH12 enzymes the thumb structure is little more than a protuberance and thus cannot play the same functional role as that of the thumb of GH11 xylanases (Paës et al., 2007; Sabini et al., 1999). Among the functions attributed to the thumb of GH11 xylanases is that of gatekeeper. It is postulated that the thumb would be one of the major obstacles for the binding of glucose-based substrates in the active site of GH11 xylanases. In support of this postulate, Paës *et al* (2007) have shown, via the complete deletion of the thumb of the xylanase from *T. xylanilyticus,* that a thumbless GH11 xylanase can bind cellulo-oligosaccharides, although hydrolysis does not occur.

D.4.3.5 N-terminal end

The N-terminal region of GH11 xylanases is usually considered to span from the N-terminal amino acid through to β-strand B3, hence including five or six β-strands (**Figure B-4**.A). Sequence alignments indicate that the N-terminal region varies drastically in sequence and length among family 11. Based on thermodynamic simulation, this N-terminal part has been suggested to be an unstable region, wherein unfolding of the xylanase could be initiated (Purmonen et al., 2007). Therefore, abnormal folding or loss of N-terminal β-strands are likely to cause protein instability or denaturation (Purmonen et al., 2007).

Compared to mesophilic GH11 xylanases, thermophilic ones generally possess an additional β-strand at the N-terminus. Therefore, one might suppose that the elevated number of strands rigidify the protein structure and thus stabilize the molecular conformation (Hakulinen et al., 2003; Purmonen et al., 2007; Ruller et al., 2008). However, a longer N-terminal strand does not necessarily correlate with higher thermostability. This is demonstrated by the xylanase from *Neocallimastix patriciarum,* which has an unusually long N-terminal sequence, but is a mesophilic enzyme (Gilbert et al., 1992; Vandermarliere et al., 2008). This is because thermostability is the result of an intricate combination of several features such as a high Thr : Ser ratio, high numbers of charged residues and/or aromatic

residues or the presence of ion pairs at the surface. (Gruber et al., 1998; Hakulinen et al., 2003; Pack and Yoo, 2004).

Since the N-terminal region appears to have a casual relationship with regard to the thermo-tolerance of xylanases, a lot of engineering work has been focused on it. Disulphide bonds were introduced into this N-terminal region and were shown to successfully stabilize the structure of *T. reesei* and *T. xylanilyticus* xylanases (Fenel et al., 2004; Paës and O'Donohue, 2006). In another experiment, extension of the N-terminal side of the protein with an arginine-rich sequence has improved the thermal performance of the fusion protein compared to the wild-type (Sung, 2007). Similarly, the replacement of the N-terminal region of a mesophilic xylanase by that of a thermophilic one has also procured higher thermal tolerance (Shibuya et al., 2000; Sun et al., 2005). Finally, Zhang and colleagues have managed to enhance the thermostability of 3 mesophilic xylanases – from *Streptomyces olivaceovirdis, Streptomyces lividans, Aspergillus niger* – using a combination of five thermophilic mutations in the N-terminal sequence (Zhang et al., 2010).

From a functional point of view, the N-terminal region might also partially contribute to the more remote glycon subsites of the active site cleft. However, the electron density around the N-terminal region is difficult to analyze and model when xylanases are bound to substrates, suggesting that this region is disordered (Vandermarliere et al., 2008; Vardakou et al., 2008). Nevertheless, the N-terminal region might display kinetic assistance in catalysis, as revealed from a study focused on the *Neocallimastix patriciarum* xylanase. Here, it was proposed that the full β-strands A1 and B1 would make extensive binding at the -3 subsite, and thus provide an explanation for the unusually high catalytic activity of this enzyme (Vardakou et al., 2008).

Part C The endo-β-1,4-xylanase from *Thermobacillus xylanilyticus* (Tx-Xyl)

C.1 Biological and biochemical characteristics

C.1.1 General characteristics of Tx-Xyl

C.1.1.1 Origin of Tx-Xyl

The microorganism *Thermobacillus xylanilyticus*, previously designated as *Bacillus* sp. D3, is an aerobic, thermophilic, xylanolytic and spore-forming bacterium. It was first isolated from soil beneath a manure heap in northern France, using an enrichment and a functional screening approach with oat spelt xylan being used as the main carbon source (Samain, 1991; Samain, 1992). *T. xylanilyticus* is a moderate thermophile, growing at temperatures up to 63°C, and in a wide range of pH from 6.5 to 8.5 (Samain, 1992; Touzel et al., 2000). Based on 16S rDNA sequence comparisons, this microorganism has been identified as a new genus and species that belongs to the Gram-negative group of bacteria (Touzel et al., 2000).

When cultivating the type strain *T. xylanilyticus* XE at 50°C over 16 hours, an expression level of 110 U ml^{-1} xylanase production was obtained (Samain, 1992). This activity was attributed to a single protein species, designated as Tx-Xyl. This enzyme is an extracellular, cellulose-free, low-molecular-weight (20.6 KDa) endo-β-1,4-endoxylanase, which belongs to the GH11 family (Samain et al., 1997; Samain, 1992). To maximize expression, strain XE was submitted to *in vivo* mutagenesis using ethyl methanesulfonate. This led to the isolation of a hyper-producing strain designated *T. xylanilyticus* D3 (Samain, 1991), which produces >1000 U ml^{-1} of xylanase activity in less than 15 hours of cultivation (Samain et al., 1997). The sequence encoding the mature Tx-Xyl (Figure C-1) was determined and then expressed in *Escherichia coli* (Harris et al., 1997).

```
  1 aacacgtactggcagtattggacggatggcatcgggtatgtgaac
    N  T  Y  W  Q  Y  W  T  D  G  I  G  Y  V  N
 46 gcgacgaacggacaaggcggcaactacagcgtaagctggagcaac
    A  T  N  G  Q  G  G  N  Y  S  V  S  W  S  N
 91 agcggcaacttcgtcatcggcaagggctggcaatacggtgcgcac
    S  G  N  F  V  I  G  K  G  W  Q  Y  G  A  H
136 aaccgggttgtcaactacaacgccggcgcatggcagccgaacggc
    N  R  V  V  N  Y  N  A  G  A  W  Q  P  N  G
181 aacgcgtatctgacgctgtacggctggacgcgcaacccgctcatc
    N  A  Y  L  T  L  Y  G  W  T  R  N  P  L  I
226 gaatactacgtcgtcgacagctggggcagctaccgcccgaccggc
    E  Y  Y  V  V  D  S  W  G  S  Y  R  P  T  G
271 gactaccggggcagcgtgtacagcgacggcgcatggtatgacctc
    D  Y  R  G  S  V  Y  S  D  G  A  W  Y  D  L
316 tatcacagctggcgctacaacgcaccgtccatcgacggcacgcag
    Y  H  S  W  R  Y  N  A  P  S  I  D  G  T  Q
361 acgttccaacaatactggagcgttcgtcagcagaaacgcccgacg
    T  F  Q  Q  Y  W  S  V  R  Q  Q  K  R  P  T
406 ggcagcaacgtctccatcacgttcgagaaccacgtgaacgcatgg
    G  S  N  V  S  I  T  F  E  N  H  V  N  A  W
451 ggcgctgccggcatgccgatgggcagcagctggtcttaccaggtg
    G  A  A  G  M  P  M  G  S  S  W  S  Y  Q  V
496 ctcgcaaccgaaggctattacagcagcggatactccaacgtcacg
    L  A  T  E  G  Y  Y  S  S  G  Y  S  N  V  T
541 gtttggtaa 549
    V  W  *
```

Figure C-1. **Nucelotide sequence and deduced amino acid sequence of the wild-type of Tx-Xyl xylanase from *T. xylanilyticus*.**

C.1.1.2 Activity of Tx-Xyl

Tx-Xyl xylanase displays high activity on soluble xylan (e.g. birchwood xylan) and yields mainly xylo-oligosaccharides of DP (degree of polymerization) 2-4 as end products (Benamrouche et al., 2002; Zilliox and Debeire, 1998; Beaugrand 2004). On lignocellulosic biomass, such as wheat straw and bran, Tx-Xyl is found to be only active on the AX fraction of the substrate, and in the case of bran hydrolysis, Tx-Xyl focused on AXs displaying a high Xyl : Ara ratio (Benamrouche et al., 2002; Zilliox and Debeire, 1998). The pH optimum of Tx-Xyl is around 6, but the enzyme is active over a broad pH range.

C.1.1.3 Thermostability of Tx-Xyl

Thermostability measurements have shown that Tx-Xyl is relatively thermostable and that this stability is dependent on enzyme concentration, indicating that the formation of

intermolecular interactions is important for this physical property (Harris et al., 1997). At 20 mg ml^{-1}, Tx-Xyl loses 60-70% of its initial activity when incubated at 60°C for 24 hours, whereas at a concentration of 200 mg ml^{-1}, the half-life at 75°C and 80°C are 40 and 25 minutes respectively (Harris et al., 1997). Other characteristics of Tx-Xyl are listed in Table C-1.

Table C-1. **Principle parameters of Tx-Xyl xylanase.**

Enzyme identity	
Name	Tx-Xyl
Fonction	endo-β-1,4-xylanase
Microorganism	*Thermobacillus xylanilyticus*
GH family	GH11
EC Classification	EC 3.2.1.8
Genebank number	CAJ87325.1
Structural characteristics	
Nucleotide number	549
Amino acid number	182
Top 3 abundant residues	Gly (12.6%), Tyr (11.5%) and Ser (10.4%)
Folding	β-jelly roll
Catalytic residues	E76 (nuclophile); E169 (acid/base)
Composition	57% β-sheet and 6% α-helix
Biochemical characteristics	
Molecular weight	20.693 kDa
pI	7.7
Optimal pH	5.8 – 6.0
pH activity	> 50% residual activity in the range of pH 4.2 - 8.2
Optimal temperature	~ 68°C
Apparent T_m	75.5°C
Extinction coefficient	218592 (experimental) / 102790 (theoretical) M^{-1}cm^{-1}
Thermostability	Several hours at 60°C and several days at 50°C

C.1.2 Action on lignocellulosic biomass

Previous studies indicate that the hydrolytic behaviour of Tx-Xyl is different depending on the nature of the substrate (Beaugrand et al., 2004; Benamrouche et al., 2002; Lequart et al., 1999; Zilliox and Debeire, 1998). On wheat bran, Tx-Xyl can solubilise 49% of available AXs. This reaction follows Michaelian kinetics and is characterized by a fast initial phase (over 1h) and completion after 24 hours, achieved using low enzyme loadings (i.e. as little as 50 IU kg^{-1} bran) (Beaugrand et al., 2004; Benamrouche et al., 2002). Previously, microscopic analyses

have revealed that Tx-Xyl initially disorganizes the homogeneous aleurone layer and then penetrates towards the inner bran layer, where 80% and 50% of container polysaccharides would be degraded respectively (Figure C-2). However, the further penetration of Tx-Xyl is impeded and the outer bran layer remains intact, possibly because of the high content of phenolic compounds in the outer layer or because of the highly substituted nature of the AXs therein (Beaugrand et al., 2005; Benamrouche et al., 2002). Significantly, the hydrolysis of wheat bran using Tx-Xyl in combination with a GH 10 xylanase did not reveal any evidence of synergy, and only led to a modest increase of liberation of ferulic acid and *p*-coumaric acid (Beaugrand et al., 2004).
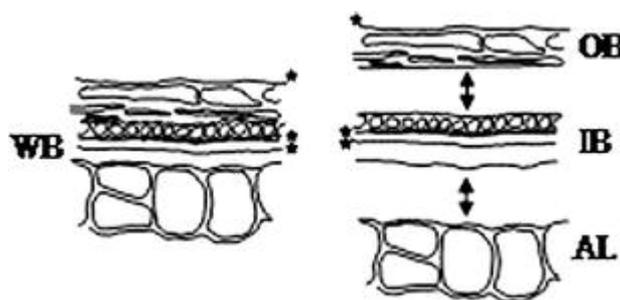
Figure C-2. **Schematic structure of a cross-section of wheat bran (WB)**. From the interior to the exterior of the grain, wheat bran is composed of outer bran layer (OB), inner bran layer (IB) and aleurone layer (AL) (Beaugrand et al., 2005).

When wheat straw is used as the substrate, Tx-Xyl action is considerably more limited, since it can only release 18 to 20% of total AXs over a 24-H period and at high enzyme loadings (Lequart et al., 1999; Zilliox and Debeire, 1998). This reaction does not obey Michaelian kinetics and increasing enzyme loads do not lead to proportional increases in reaction rate. Likewise, the reaction does not appear to reach a clear endpoint, even after extended incubation periods (Lequart et al., 1999; Rémond et al., 2010; Zilliox and Debeire, 1998). Absorption studies indicated that Tx-Xyl binds tightly and irreversibly to wheat straw, up to a maximum of 512 µg enzyme g^{-1} straw. However, importantly Tx-Xyl was shown to bind to isolated wheat straw lignin, but not to microcrystalline cellulose, which might be due to the hydrophobic character of Tx-Xyl (Zilliox and Debeire, 1998).

Undoubtedly, the different hydrolytic profiles displayed by Tx-Xyl on wheat bran and wheat straw are correlated to the obvious differences between these two substrates. A significant

proportion (approximately 50%) of AXs in wheat bran is localized in the aleurone layer, which is composed of living cells. AXs in this layer are moderately substituted by arabinosyl units, with an approximate average X/A ratio of 2. In the Tx-Xyl-recalcitrant wheat bran residue, the X/A ratio is 0.98. Therefore, as expected Tx-Xyl is less active on highly substituted AX. Nevertheless, in wheat straw, AXs are quite unsubstituted, displaying a xylose : arabinose ratio of approximately 10, thus this factor is not a limitation for Tx-Xyl (Lequart et al., 1999). However, the ultra-structure of wheat straw is quite different from wheat bran, and the former contains approximately 20% of lignin, whereas wheat bran only contains about 3.4% (Lequart et al., 1999) (Figure C-3). Therefore, the presence of lignin is probably one of the key barriers for Tx-Xyl, though the presence of diferulic acid bridges, linking lignin to AX could also contribute to the resistance of wheat straw to Tx-Xyl action (Grabber et al., 1998; Lequart et al., 1999).
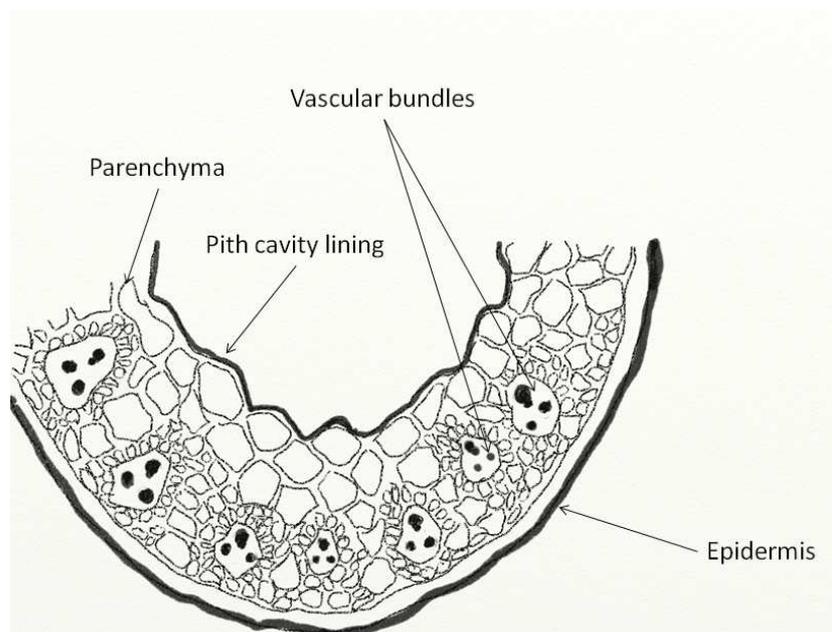
Figure C-3. Schematic structure of a cross-section of an inter-node portion of wheat straw. The outermost ring is termed epidermis which is rich in cellulose. Inside the epidermis is a lignin containing (25-27%) layer loosely composed of vascular bundles and parenchyma (Hornsby et al., 1997a). The pith cavity lining is a thin layer of modified parenchyma cells that lines the centre void (pith cavity) (Hansen et al., 2010). The diagram is drawn referring to the microscopy images from (Hansen et al., 2010; Liu et al., 2005).

C.2 Structure of Tx-Xyl

C.2.1 Overall structure

The crystal structure of Tx-Xyl has been determined by Harris *et al* (1997). Like all GH11 xylanases, Tx-Xyl displays a typical β-jelly roll framework, consisting of two antiparallel β-pleated sheets and one α-helix (Figure C-4). The catalytic dyad is composed of two glutamate residues – Glu76 (nucleophile) and Glu169 (acid/base catalyst) – that are situated in the centre of the active site cleft. When comparing the 3D structure of Tx-Xyl to other GH11 xylanases, two individual features can be distinguished i.e. (1) a significantly large number of surface exposed aromatic residues and (2) a distinctly shorter N-terminal region.
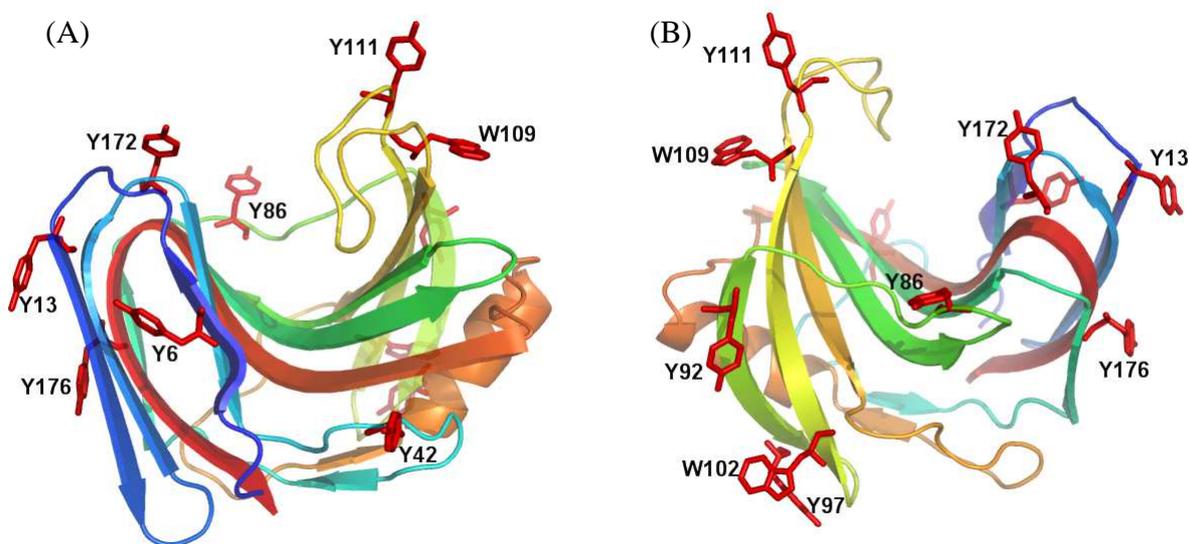
Figure C-4. **Front view (A) and back view (B) of the cartoon tertiary structure of Tx-Xyl xylanase.** The protein schematic is "colour-ramped" from the N-terminus (blue) to the C-terminus (red). The eleven surface aromatic residues (Y6, Y13, Y42, Y86, Y92, Y97, W102, W109, Y111, Y172 and Y176) are indicated in stick presentation with red colour.

Harris et al have proposed that the high number of surface-exposed aromatic residues may form the basis of thermostability in Tx-Xyl. Precisely, it has been postulated that the eleven surface aromatics, Tyr6/13/42/86/92/97/102/111/172/176 and Trp102/109 (Figure C-4), form clusters or ''sticky patches'' that mediate intermolecular protein interactions between monomers via aromatic-aromatic interactions (Harris et al., 1997). The formation of dimers or higher order polymers is assumed to help the stabilization of enzyme at high temperature.

On the other hand, the shorter N-terminal end of Tx-Xyl is a curious feature that is apparently in contradiction with the notion of thermostability, because usually thermophilic GH11 xylanase possess an extended N-terminal region (Hakulinen et al., 2003).

C.2.2 Topology of the active site

Using crystallographic data showing a *Bacillus circulans* GH11 xylanase complexed to xylobiose as template (Wakarchuk et al., 1994), a xylohexaose molecule has been built into the active site of 3D model of Tx-Xyl, thus providing a theoretical view of subsite (-3 to +3) occupancy (Paës et al., 2007). This docking model has provided useful means to investigate binding energies in each subsite and, in particular, has furnished data on the putative -3 subsite. According to the model, the existence of this subsite is uncertain, since only a weak binding energy has been calculated and the xylosyl moiety is visualized as being partially exposed to the solvent (Paës et al., 2007) (Figure C-5 and Table C-2). In contrast, the other subsites appear to be associated with much more favourable binding energies. Two stacking interactions are identified at the -2 and +3 subsites, while Van der Waals interactions appear to be the main driving force for interactions in the -1, +1 and +2 subsites. A total of six hydrogen bonds were predicted through -2 to +3 subsites. Apart from the predicted H bond formed between Tyr89 and the xylosyl moiety in +2 subsite, which is only seen in equivalent modelled structures (Andre-Leroux et al., 2008; Cervera Tison et al., 2009), the other equivalent H bonds have been identified through X-ray crystallographic studies and have been proposed to be highly conserved among GH11 xylanases (Janis et al., 2005; Sabini et al., 1999; Vandermarliere et al., 2008; Vardakou et al., 2008; Wakarchuk et al., 1994).

Table C-2. Binding interactions between the Tx-Xyl xylanase and docked xylohexaose (Paës, 2005).

Subsite	Stacking	Hydrogen bond	Van der Waals interactions
-3		O2···NE1 (W7)	S115
-2	W7	O3···OH (Y163)	Q5, S115, I116
		O2···OH (Y67)	
-1		O2···NH1 (R110)	W7, N33, V35, Y67, E76, Y78, P114, S115
+1		O···OE2 (E169)	N33, Y78, R110, Q124, W126, Y171
+2		O2···N (T89)	Y63, P88, T89, W126, Y171
+3	Y86	O2···ND2 (N61)	Y63, Y86, Y171
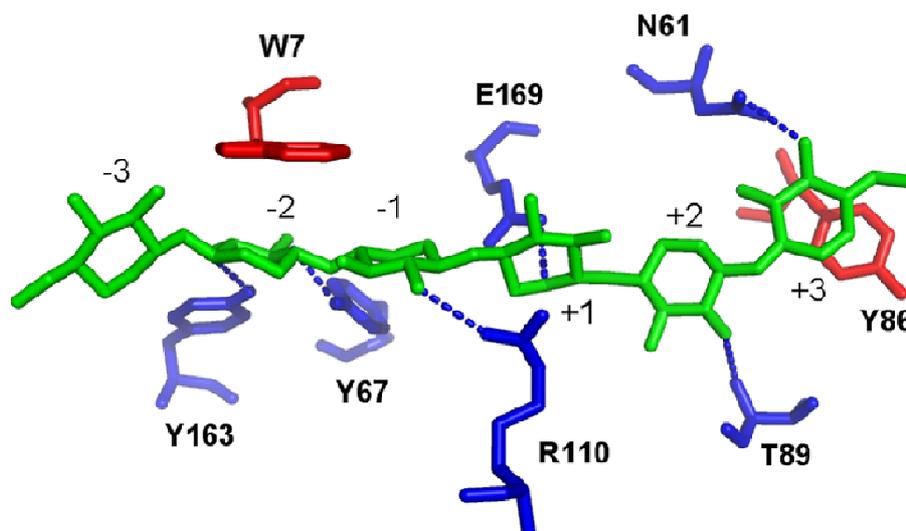
Figure C-5. **Binding of xylohexaose in the active site of Tx-Xyl.** The xylohexaose molecule is shown in green, indicated with subsite number corresponding to each xylosyl moiety. Hydrogen-bonded amino acid residues are in blue, and hydrogen bonds are drawn in blue dash lines. Stacking residues are in red.

C.2.3 Thumb region of Tx-Xyl

To better understand the function of the thumb loop in Tx-Xyl, Paës *et al* (2007) have probed the conserved triplet Pro114-Ser115-Ile116 using site-directed saturation. An unnatural combination, Pro114-Gly115-Cys116, has been generated and surprisingly, the k_{cat} value of this novel triplet variant was increased by 20%. In the same study, site-directed mutagenesis was also performed on amino acids Tyr111 and Thr121 that together form the N- and C-terminal extremities of the thumb loop. The analysis of mutant enzymes in which either Tyr111 or Thr121 were deleted revealed that these amino acids are tightly linked to catalysis, probably determining the 3D trajectory of the loop's motion. Finally, the complete deletion of the loop, creating a "thumbless" enzyme, abolished catalysis but allowed the mutant to bind both xylo-oligosaccharides and cellotetraose. Taking into account the fact that wild type Tx-Xyl is unable to bind cello-oligosaccharides, this result indicates that the thumb plays a substrate filtering role, exclusively admitting xylose-based oligosaccharides and polymers into the active site (Paës et al., 2007).

C.2.4 Protein engineering of Tx-Xyl

The cloning and expression in *E. coli* of the sequence encoding mature Tx-Xyl (i.e. without a signal peptide) led to the production of a recombinant enzyme that displayed abnormally

low thermostability. Closer analysis of the recombinant protein revealed that the initial Met residue, which is absent in 90% of *E. coli*-expressed proteins was still present and that this extra residue was the source of protein instability (Paës, 2005). Normally, the initial Met should be removed by the combined post-translational action of a methionyl-deformylase and a methionyl-aminopeptidase (Ben-Bassat et al., 1987; Schmitt et al., 1996). To circumvent this problem, a new recombinant Tx-Xyl was produced, which bears an Ala in the place of Asn at position 1. This single amino acid modification is sufficient to alleviate post-translational protein processing problems and allows the production of a protein that presents characteristics identical to that of wild type Tx-Xyl.

Similarly, in an attempt to increase the thermostability of Tx-Xyl, Paës *et al* (2006) have created several variants via the introduction of cysteine residues that can mediate the formation of disulphide bridges. The first disulphide bridge was introduced by mutating Ser98 and Asn145, residues that are located on β-strand B9 and on the α-helix respectively. This disulphide bridge led to increased thermostability, with a 4-fold increase in half-life at 70°C, but slightly decreased thermoactivity. The second disulphide bridge, linking the N-terminal and C-terminal extremities, was produced by adding Cys residues at either extremity. This led to increases in both thermostability and thermoactivity. Likewise, the combination of the two disulphide bridges procured an additive effect, increasing by 10-fold the half-life of the resulting mutant at 70°C and nearly doubling the specific activity at 75°C. The results suggested that the disulphide bonds slowed down the denaturation of Tx-Xyl.

Part D Irrational engineering of xylanase

Over the past two decades, random enzyme engineering strategies have been increasingly applied with success to a wide variety of protein engineering projects (Arnold and Moore, 1997; Arnold and Volkov, 1999; Johannes and Zhao, 2006). Likewise, several examples of random engineering applied to xylanases are available. Compared to rational approaches, random engineering does not require detailed structural or mechanistic information, but simply relies on good knowledge of the reaction that is catalyzed and a robust, pertinent screening method that allows the identification of improved enzyme variants. Moreover, the success of random approaches (including DNA shuffling) is highly dependant on two factors, which are the diversity of the mutant library and the reliability of screening method (Johannes and Zhao, 2006).

D.1 Mutagenesis techniques

D.1.1 Error-prone PCR

Directed evolution is a popular strategy to improve enzyme characteristics, which can be used to improve the fitness of enzymes for industrial processes. Likewise, when coupled to detailed analyses of improved enzymes, directed enzyme evolution is a valuable approach for probing protein structure-function relationships (Arnold and Volkov, 1999).

Generally, to achieve directed evolution, both random *in vitro* mutagenesis and gene shuffling techniques are employed. Several methods can introduce random mutations into a target gene. These include UV irradiation (Bagg et al., 1981), chemical mutagenesis (e.g. deamination, alkylation, base-analog mutagens, etc.) (Encell et al., 1998; Lai et al., 2004), cloning genes in mutator strains (e.g. XL1-red) (Greener et al., 1997; Henke and Bornscheuer, 1999) and PCR-based mutagenesis (e.g. error-prone PCR, PCR using randomly synthesized oligonucleotides, gene shuffling etc.) (Botstein and Shortle, 1985; Cadwell and Joyce, 1992; Fromant et al., 1995). However, among the various methods, error-prone PCR (epPCR) is the most attractive one, because of its versatility and simplicity.

In epPCR, the introduction of mutations relies on the low fidelity of polymerase-catalyzed DNA replication, which is induced by several factors, including dNTP imbalances, high concentrations of divalent metal ions, and the intrinsic low fidelity of certain DNA polymerases (Cadwell and Joyce, 1994; Cirino et al., 2003). The very common *Taq* DNA polymerase (hereafter called *Taq*) is a suitable choice for epPCR, because of its lack of fidelity when used with inappropriate buffers. The natural error rate generated by *Taq* usually varies between 0.001 and 0.002% per nucleotide per round of replication, (Cadwell and Joyce, 1992),which is insufficient for directed evolution projects. To enhance the error rate of *Taq*, epPCR typically contains a high concentration of Mg^{2+} (normally around 7 mM), which helps to stabilize non-complementary nucleotide pairing. In addition, the error rate can also be adjusted by adding Mn^{2+} into the PCR reaction. This increases mis-incorporation of nucleotides into the amplified product and the effect of Mn^{2+} ions is proportional to the concentration used (LinGoerke et al., 1997). However, one important drawback of *Taq* utilization is linked to a higher frequency of transition mutations (A $\leftrightarrow$ G and T $\leftrightarrow$ C) compared to transversions (A/G $\leftrightarrow$ C/T) (Cadwell and Joyce, 1992; Cadwell and Joyce, 1994; Cirino et al., 2003). To reduce this inherent bias, the use of an unbalanced mixture of dNTPs is thus recommended. For instance, Cirino *et al* (2003) published a protocol that uses 1.0 mM dCTP/dTTP and 0.2 mM dGTP/dATP, which apparently solved the problem and provided relatively even occurrence of mutations involving the four nucleotides.

In addition to the influence of external factors, one must also consider internal factors. The first is inherent codon bias, which must be taken into account when performing random mutagenesis. For example, the chance of a Val being substituted by a Phe is greater than if Val is mutated to Cys or Glu, simply because the Val→Phe substitution only requires a single nucleotide mutation, whereas the latter mutations require either double or triple mutations. This bias linked to codon degeneracy can be considered as a self-protection for living organisms, since it limits the loss of function caused by point mutations (Neylon, 2004). However, this is clearly undesirable in *in vitro* random mutagenesis, because it reduces the diversity of mutant libraries, compared to that which is theoretically achievable. Other factors that can affect error rate, and thus library quality, are template concentration and the number of amplification cycles. Importantly, mutation frequency will vary between different templates, even if the same amplification conditions are employed, because

sequence length, nucleotide composition, etc… play extremely important roles (Cirino et al., 2003). Therefore, it is important to measure error rate for each individual epPCR experiment.

D.1.2 *In vitro* DNA recombination

Normally, advantageous mutations generated by epPCR are randomly distributed throughout the sequence and, most of the time, mixed with deleterious ones. Therefore, the overall result is often modest improvements of the targeted function, or even complete failure to detect beneficial mutants, that are masked by deleterious ones. As a consequence, in order to concentrate beneficial point mutations and facilitate their identification, homologous recombination can be performed. This technique significantly increases enzyme fitness, thanks to synergistic interactions between the beneficial mutations. To achieve this, several robust DNA recombination technologies have been developed.

DNA shuffling, also known as sexual PCR, is the original homologous DNA recombination approach (Stemmer, 1994b; Stemmer, 1994a). In this process, two or more parental genes are first randomly cleaved by DNase I, producing small fragments (≥50 bp) that are then reassembled in a PCR reaction that is performed without primers. Homologous and partially homologous fragments anneal together and are extended progressively cycle after cycle. The outcome of the process is chimeric self-primer extension products, which can then be amplified using standard PCR reaction, thus creating a library of daughter genes (Figure D-1.A) (Joern, 2003; Neylon, 2004; Zhao and Arnold, 1997).

Another classical DNA recombination method is the **Staggered Extension Process** (StEP) (Zhao et al., 1998), which does not require DNA fragmentation, but is based on a very simple PCR process using short elongation times (Figure D-1.B). Compared to standard PCR, StEP PCR also employs primers that anneal to the extremities of the sequence to be amplified, but the amplification program merges annealing and extension into one single step characterized by a very short incubation time. Consequently, in each cycle, the DNA polymerase synthesizes stretches of DNA of limited length along the template. Then, in the next cycle, single stranded fragments can anneal to a different template and therefore generate a crossover. After the cycle is repeated until the full length of the sequence is built up. The key point of StEP PCR is to limit the DNA replication rate, in order to increase the

incidence of template switching or crossover (Zhao and Zha, 2006). A lower annealing/extension temperature reduces the turnover rate of the DNA polymerase and promotes the complementation between fragments that are only partially homologous. Nevertheless, caution is required, because a too low temperature would cause non-specific annealing. When using *Taq*, Zhao *et al*(2006) recommended that an optimal annealing temperature would be in the range of $(T_m - 25)°C$ to $(T_m - 5)°C$. The *Taq* can also be replaced by other polymerases with proofreading activity, especially enzymes with slow extension rate such as the *Vent* polymerase, which only incorporates 1000 bases per minute comparing to > 4000 bases per minute for *Taq* (Aguinaldo and Arnold, 2003; Zhao and Zha, 2006).

In addition to DNA shuffling and StEP, other well-known DNA recombination methods include **Random Chimeragenesis on Transient Templates** (RACHITT) and some fragmentation-free PCR methods, derived from the StEP method. RACHITT involves the hybridation of single-stranded fragments onto the full-length complementary ssDNA of a template parental gene. Therefore, this technique does not require any thermocycling, but only needs the removal of non-annealed flaps and filling in of remaining gaps (Figure D-1.C) (Coco et al., 2001; Pelletier, 2001).

When all of these DNA recombination methods are compared, StEP offers the simplest procedure, though RACHITT is the most advantageous, because it involves elimination of parental genes in the chimeric library. Even though the protocol for DNA shuffling is a little labour intensive, the average number of crossovers per chimeric gene is similar or higher than that of StEP (Chaparro-Riggers et al., 2007; Zhao and Zha, 2006). Among the three methods, RACHITT produces more crossovers than the other two methods (Coco et al., 2001).
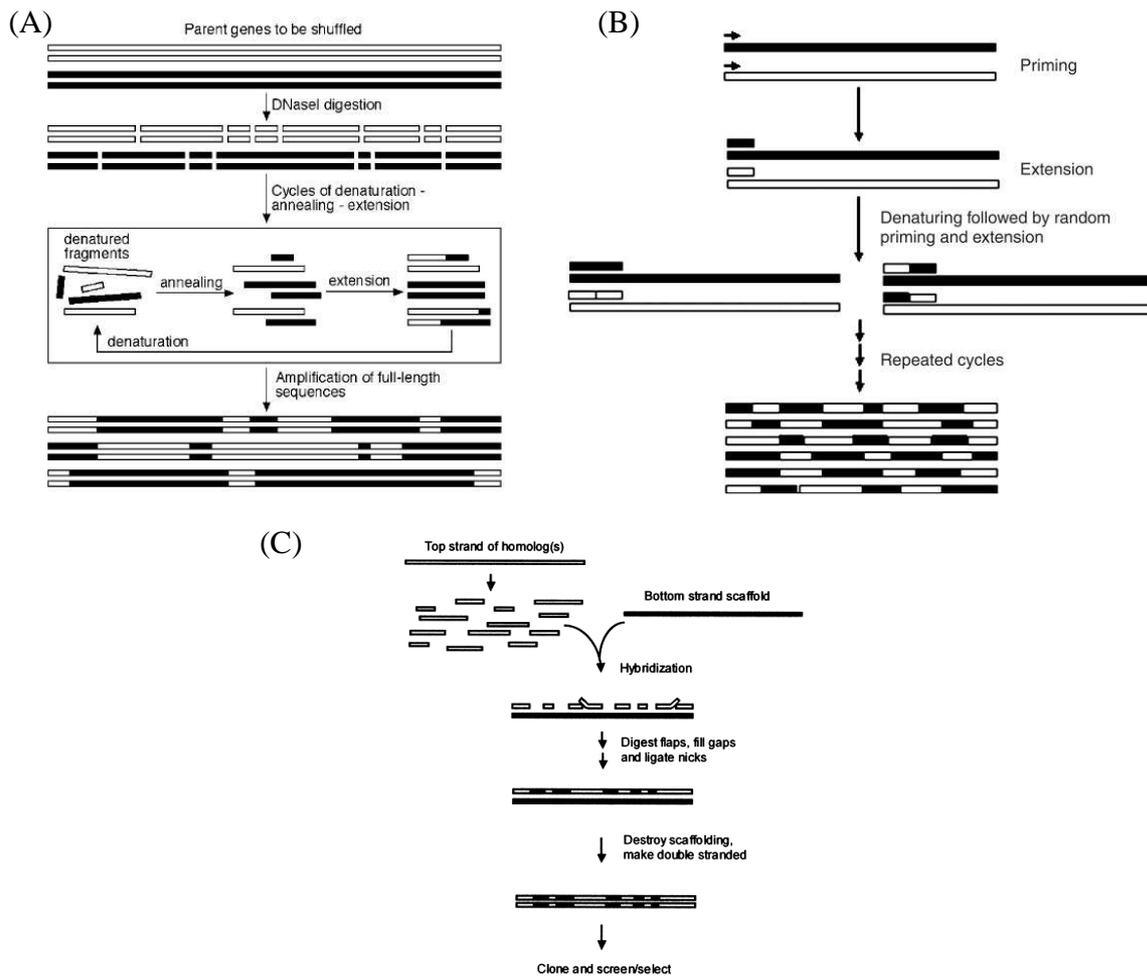
Figure D-1. Schematic representation of DNA recombination techniques: (A) DNA shuffling (Joern, 2003), (B) Staggered Extension Process PCR (StEP) (Zhao and Zha, 2006), and (C) RAndom CHImeragenesis on Transient Templates (RACHITT) (Coco et al., 2001).

D.1.3 Site-saturation mutagenesis

Site-saturation mutagenesis is well adapted to the further investigation of mutational hot-spots that have already been identified. Basically, this technique allows the substitution of a specific amino acid for the 19 other amino acids. Saturated mutagenesis is either achieved in a 'one pot' approach using degenerate mutagenic primers, in which the target codon is randomized, or in a more laborious parallel approach, in which each amino acid variant is created by site-directed mutagenesis. In the former strategy the design of the mutagenic primers can be achieved in a straightforward way by introducing the sequence NNN (N = A, T, G or C). This will lead to the generation of all 64 codons in a mutational library, theoretically, in equal proportions. Nevertheless, the number of possible codon triplets can be reduced to

32 by restricting the mutation of the third base to G or C only, as this does not influence the overall amino acid diversity, but only decreases the size of library and the probability of introduction of stop codons (Georgescu et al., 2003). Similarly, it is possible to further restrict the sequence of the targeted codon in order to avoid the generation of the original codon. However, in this case the overall number of possible amino acid substitutions is also likely to be decreased. In addition, when designing the primers, it is critical to leave enough correct bases on both sides to ensure the annealing occurs normally. The PCR reaction is the same as when doing site-directed mutagenesis, in which the full length of plasmid is amplified (Georgescu et al., 2003; Ho et al., 1989).

D.1.4 Family shuffling

In Nature, evolution endows each species with inherent traits, such as thermostability, psychrophilicity, acidophilicity or alkaliphilicity, to adapt to various natural surroundings. Family shuffling is a powerful process that permits to combine these natural advantages in one chimeric protein through *in vitro* DNA recombination. Since Stemmer's group first formulated family shuffling in 1998, this method has been shown to be efficient for the acceleration of protein evolution (Crameri et al., 1998). As described in Chapter D.1.1, epPCR accumulates point-mutations in a single DNA resource. In contrast, family shuffling simply recombines homologous genes from a wide variety of related organisms, and therefore avoids degenerate mutations (i.e. in principle all of the parental genes encode active proteins) and greatly expands the sequence space which is explored (Figure D-2).

In theory, all of the methodologies suitable for simple DNA recombination (described in Chapter D.1.2) can be used in family shuffling. However, without modification, their direct application in family shuffling leads to poor mixing of the parental genes, because of their low similarity. For example, shuffling of two parental genes encoding catechol 2,3-dioxygenases (84% nucleotide identity) by traditional DNA shuffling method, only resulted in a low frequency of hybrid mutants (<1%) (Kikuchi et al., 1999). StEP-mediated recombination gives a better result when parental genes share more than 80% of identity (Aguinaldo and Arnold, 2003; Chaparro-Riggers et al., 2007). The main problem is that, in the course of reassembly, most re-annealing occurs between fully homologous segments, derived from

the same parental gene (homo-duplex formation) instead of heterogeneous genes (hetero-duplex formation).
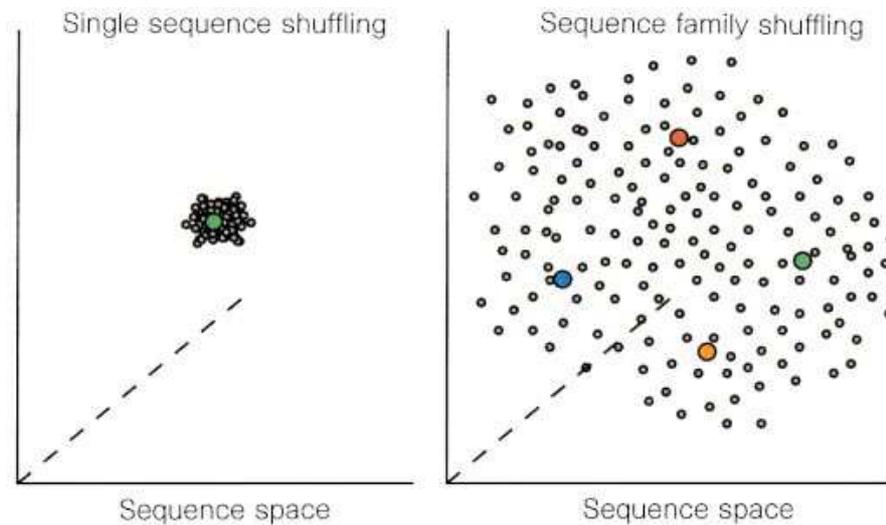
Figure D-2. Sequence space of chimeric library generated by family shuffling versus single sequence shuffling (Crameri et al., 1998). The coloured dots represent the parent(s) which are used as template in the different strategies.

To overcome the problem of homo-annealing, Kikuchi *et al* (1999) have utilized restriction enzymes to cleave parental genes independently; thus DNA elongation only appears on hetero-duplex molecules in the reassembly process. The success of this method relies heavily on the location of the cleavage sites: indeed the hetero-duplex will be hard to amplify if the 3'-ends of the two fragments are closer than 20 bp. That's why, a set of multiple restriction enzymes is recommended for the gene fragmentation. Alternatively, single stranded (ss) DNA shuffling can also be used, in which a single-stranded DNA template is first prepared, either through the use of a plasmid bearing a M13 bacteriophage origin of replication and helper phage (Kikuchi et al., 2000), or by λ exonuclease digestion of phosphorylated dsDNAs (Zha et al., 2003). Once ssDNA is obtained, it is randomly digested with DNaseI and reassembled. Nevertheless, in the ssDNA-based method, extra care needs to be provided to produce and purify single-stranded DNA, avoiding contamination with dsDNA or simply the complementary strand, which would otherwise result in failure of the experiment. However, when properly performed, both approaches powerfully improve the frequency of hybrid genes to reach 100% of chimera in the library.

D.2 Application of random approaches to xylanases

D.2.1 High-throughput screening approaches

The success of combinatorial engineering does not require detailed 3D protein structures but critically depends on the availability of an appropriate high-throughput screening methodology. A good screen should be i) be specific to the experimental demand, ii) sensitive enough to discriminate between negative and positive mutants and iii) perfectly reproducible (Arnold and Moore, 1997; Moore et al., 1997).

Concerning the screening of xylanase mutants, the choice of substrate is a key issue. Selection based on the degradation of a substrate and the formation of a lytic halo on solid medium is widely used to screen active clones in libraries. Generally, synthetic chromogenic substrates, such as the widely used RBB-xylan (Shibuya et al., 2000; Stephens et al., 2007) or Blue-xylan (Andrews et al., 2004; Xie et al., 2006) are preferred because substrate hydrolysis automatically leads to the formation of a clear zone. However, unmodified xylan can also be used, but in this case it is necessary to reveal the lytic halo using a polysaccharide-specific colouring agent such as Congo red (Chen et al., 2001; Ruller et al., 2008). Other screening criteria can for example be pH-based, where lytic haloes are allowed to develop in medium containing acid or alkaline compounds (Chen et al., 2001). Overall lytic halo detection is rapid and intuitive, but it does not give any quantitative indication of the targeted improvements.

In order to perform rigorous screening, it is desirable to use individually cultured clones and cell-free extracts to measure activity. Likewise, the library variants can be exposed to stringent conditions before residual enzyme activity is quantified. For example, to screen for thermostability a heat shock selection at the targeted temperature is often employed (Miyazaki et al., 2006; Palackal et al., 2004). Similarly, for pH adaptation, a high pH reaction condition can be used for alkali-stability selection (Stephens et al., 2009). Residual activity is often measured in 96-well microtitre plates and quantitative measurements are facilitated by the use of chromogenic reagents. For example, for the hydrolysis of birchwood xylan or oat spelt xylan, solubilised reducing sugars can be quantified through the use of 3,5-dinitrosaylicylic (DNS) or bicinchoninic acid (BCA), which are the two most popular methods.

The DNS assay functions in a wide range of sugar concentration (6.7 - 600 mg l^{-1}) and does not cross-react with proteins (Miller, 1959; Zhang et al., 2006). On the other hand, the BCA assay is more sensitive than the DNS assay, being able to detect sugar concentrations as low as 0.2 mg l^{-1}. However, this method is subject to interference by the presence of proteins (Kenealy and Jeffries, 2003; Zhang et al., 2006). Azo-xylan, a soluble chromogenic substrate, is also widely used to quantitatively measure xylanase activity and does not require additional colorimetric reagents (Dumon et al., 2008; Palackal et al., 2004). However, this substrate is chemically modified.

One of the central dilemmas for the identification of high performance enzymes for biorefining processes is the gulf that exists between simple substrates that are often used for screening and the highly complex nature of actual lignocellulosic biomass. The latter is insoluble, amorphous and is composed of numerous macromolecules, whose accessibility to the enzyme is widely variable. To address this problem, several attempts have been made to devise screening methods that employ 'real' substrates (Chundawat et al., 2008; Navarro et al., 2010). Notably, Chundawat *et al* (2008) devised a screening method that uses finely milled, pretreated lignocellulosic biomass (diameter < 100 µm). The use of such particles allowed the constitution of a nearly homogeneous aqueous slurry, which could be pipetted using manually cut pipette tips and an automated workstation. However, Chundawat's method requires specific grinding equipment that is not widely available, and relies on a rather artisanal pipetting of slurry (which can decant) method that hinges on the use of manually modified tips. Overall, these potentially represent major sources of error, linked to biomass delivery.

D.2.2 Examples of xylanase engineering projects

Table **D-1** summarizes a number of xylanase engineering studies that have been undertaken over the last ten years. Mostly, epPCR has been employed and in some cases, thermostability has been significantly increased via only one round of random mutagenesis (Chen et al., 2001; Stephens et al., 2009; You et al., 2010). Nevertheless, directed evolution on xylanases has also been shown to be very powerful and relevant, especially for increasing the thermostability of a *Bacillus subtilis* xylanase (Miyazaki et al., 2006; Ruller et al., 2008).

Interestingly, the extent of the improvements is not proportional to the number of generated mutations. Often the best performing mutants are the result of subtle changes, while others result from more than ten mutations. Therefore, at this stage it is hard to make generic predictions concerning the residues that are important for any given aspect of xylanase activity, even when detailed tertiary structures are available.

Nevertheless, one might expect that the ever increasing use of random approaches for xylanase engineering will lead to greater insight and will eventually allow reliable prediction and rational design. In this respect, some advances have been made, via the structural analyses of thermostabilized mutants, arising from directed evolution strategies. In some cases, increased thermostability has been associated with surface hydrophobic properties and/or local hydrophobic interactions (Dumon et al., 2008; Miyazaki et al., 2006; Xie et al., 2006), and for others with a rigidified N-terminal region (Palackal et al., 2004; Shibuya et al., 2000), an increase in the potential for salt bridge formation (Stephens et al., 2009), or reinforced interactions between newly engineered Cys residues (You et al., 2010). Likewise, it has been proposed that an increase of surface polarity is highly beneficial for the development of alkaphilicity in xylanases (Chen et al., 2001; Stephens et al., 2009).

So far, the main aims of xylanase combinatorial engineering studies have concentrated on improving protein stability at high temperatures and under alkaline environments, in order to fit industrial applications, such as bio-bleaching processes used in the paper industry. However, for biorefining, these characteristics may or may not be useful, depending on the exact process design and the precise use of the enzymes. What is certain, is that biorefining will require robust enzymes that can efficiently hydrolyze xylans contained within complex lignocellulosic material or derivatives thereof (e.g. pretreated pulps). In this case, novel screening approaches using complex biomass medium coupled with powerful protein engineering strategies should be applied for xylanases.

Table D-1. **Summary of irrational engineering of xylanases over the last ten years.**

Objectives	Organism of parental gene(s)	Combinatorial engineering method	Screening strategy	Number of mutations in best mutant	Hypothesis/Explanation	Reference
More thermostable	*S. lividans* and *T. fusca*	Family shuffling between 2 parental genes	Halo selection on plate (RBB-xylan)	-	Thermostable N-termninus of *S. lividans* xylanase	(Shibuya et al., 2000)
	XYL7746 xylanase	Gene site saturation mutagenesis + Gene reassembly technology	High temperature incubation + activity assay on Azo-xylan	9	Rigidified N-terminal β-strand; Increased hydrophobic interactions	(Palackal et al., 2004)
	Bacillus subtilis	1 round of epPCR + 1 round of DNA shuffling + Site-saturation mutagenesis	High temperature incubation+ activity assay on birchwood xylan	3	Increased hydrophobicity	(Miyazaki et al., 2006)
	Thermomyces lanuginosus	2 rounds of epPCR	Halo selection on plate (RBB-xylan)	1	Not mentioned	(Stephens et al., 2007)
	Cellvibrio mixtus	1 round of epPCR + mutation recombined by site-directed mutagenesis	Halo selection on plate (overlaying with blue-xylan)	2	Increased hydrophobic interactions; Potentially enhanced substrate binding affinity	(Xie et al., 2006)
	Enviromental resource xylanase	Gene site saturation mutagenesis + Gene reassembly technology	High temperature incubation + activity assay on Azo-xylan	7	Increased hydrophobic interactions; Locked conformation of a surface loop	(Dumon et al., 2008)
	Thermotoga maritima	Family shuffling between 2 parental genes	Activity assay on oat spelt xylan and CD measurement	-	Increased interactions between N- and C-terminus	(Kamondi et al., 2008)
	Bacullus subtilis	2 round of epPCR + 1 round of DNA shuffling	Halo selection on plate (oat spelt xylan + Congo red staining)	4	Change of protein-solvent interface; Decrease in the heat capacity change	(Ruller et al., 2008)
	Thermomyces lanuginosus	Mixture of 7 independent epPCR	High temperature incubation + activity assay on birchwood xylan	4	Increased number of arginine residues on the surface	(Stephens et al., 2009)
	Geobacillus stearothermophilus	2 round of epPCR + mutation recombined by shuffling and site-directed mutagenesis	High temperature incubation + activity assay on birchwood xylan	13	Potentially enhanced substrate binding affinity	(Zhang et al., 2010)
	Neocallimastix patriciarum	1 round of epPCR + mutation recombined by site-directed mutagenesis	High temperature incubation + activity assay on birchwood xylan	2	Increased hydrophobic interaction between two cystein residues	(You et al., 2010)
More alkaliphilic/alkaline-stable	*Neocallimastix patriciarum*	1 round of epPCR + mutation recombined by site-directed mutagenesis	Halo selection on alkaline plate (oat spelt xylan + Congo red staining)	7	Increase of the negative charge on the surface	(Chen et al., 2001)
	Bacillus sp. Strain 41M-1	1 round of epPCR	Halo selection on alkaline plate	4	Not mentioned	(Inami et al., 2003)
	Thermomyces lanuginosus	Mixture of 7 independent epPCR	Halo selection on alkaline plate (birchwood xylan)	1	Increased polarity	(Stephens et al., 2009)
More stable in the absence of Ca^{2+}	*Cellvibrio japonicus*	3 rounds of epPCR	Halo selection on plate (overlaying with blue-xylan)	3	Influence in calcium binding site; New generated disulfide bridge, hydrogen bond and hydrophobic interactions	(Andrews et al., 2004)

References

Ademark, P., Varga, A., Medve, J., Harjunpaa, V., Torbjon, D., Tjerneld, F., Stabrand, H., 1998. Softwood hemicellulose-degrading enzymes from Aspergillus niger: Purification and properties of a [beta]-mannanase. *Journal of Biotechnology*, 63, 199-210.

Aguinaldo, A.M., Arnold, F.H., 2003. Staggered Extension Process (StEP) In Vitro Recombination Methods in Molecular Biology pp. 105-110.

Andre-Leroux, G., Berrin, J.G., Georis, J., Arnaut, F., Juge, N., 2008. Structure-based mutagenesis of Penicillium griseofulvum xylanase using computational design. *Proteins*, 72, 1298-307.

Andrews, S.R., Taylor, E.J., Pell, G., Vincent, F., Ducros, V.M., Davies, G.J., Lakey, J.H., Gilbert, H.J., 2004. The use of forced protein evolution to investigate and improve stability of family 10 xylanases. The production of Ca2+-independent stable xylanases. *J Biol Chem*, 279, 54369-79.

Arnold, F.H., Moore, J.C., 1997. Optimizing industrial enzymes by directed evolution. *Adv Biochem Eng Biotechnol*, 58, 1-14.

Arnold, F.H., Volkov, A.A., 1999. Directed evolution of biocatalysts. *Curr Opin Chem Biol*, 3, 54-9.

Arshadi, M., Sellstedt, A., 2009. Production of energy from biomass Introduction to Chemicals from Biomass. Edited by James H. Clark and Fabien E. I. Deswarte. WILEY, pp. 1-18.

Ash, L., 2004. Fermentation of lignocellulosic biomass. Wisconsin biorefining development initiative http://www.wisbiorefine.org/proc/fermlig.pdf.

Autio, K., Härkönen, H., Parkkonen, T., Frigård, T., Poutanen, K., Siika-aho, M., Åman, P., 1996. Effects of Purified Endo-[beta]-xylanase and Endo-[beta]-glucanase on the Structural and Baking Characteristics of Rye Doughs. *Lebensmittel-Wissenschaft und-Technologie*, 29, 18-27.

Ay, J., Gotz, F., Borriss, R., Heinemann, U., 1998. Structure and function of the Bacillus hybrid enzyme GluXyn-1: native-like jellyroll fold preserved after insertion of autonomous globular domain. *Proc Natl Acad Sci U S A*, 95, 6613-8.

Bagg, A., Kenyon, C.J., Walker, G.C., 1981. Inducibility of a gene product required for UV and chemical mutagenesis in Escherichia coli. *Proceedings of the National Academy of Sciences*, 78, 5749-5753.

Banowetz, G.M., Boateng, A., Steiner, J.J., Griffith, S.M., Sethi, V., El-Nashaar, H., 2008. Assessment of straw biomass feedstock resources in the Pacific Northwest. *Biomass and Bioenergy*, 32, 629-634.

Bayer, E.A., Chanzy, H., Lamed, R., Shoham, Y., 1998. Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology*, 8, 548-557.

Beauchemin, K.A., Colombatto, D., Morgavi, D.P., Yang, W.Z., 2003. Use of Exogenous Fibrolytic Enzymes to Improve Feed Utilization by Ruminants. *Journal of Animal Science*, 81, E37-E47.

Beaugrand, J., Chambat, G., Wong, V.W., Goubet, F., Remond, C., Paës, G., Benamrouche, S., Debeire, P., O'Donohue, M., Chabbert, B., 2004. Impact and efficiency of GH10 and GH11 thermostable endoxylanases on wheat bran and alkali-extractable arabinoxylans. *Carbohydr Res*, 339, 2529-40.

Beaugrand, J., Paës, G., Reis, D., Takahashi, M., Debeire, P., O'Donohue, M., Chabbert, B., 2005. Probing the cell wall heterogeneity of micro-dissected wheat caryopsis using both active and inactive forms of a GH11 xylanase. *Planta*, 222, 246-57.

Beg, Q.K., Kapoor, M., Mahajan, L., Hoondal, G.S., 2001. Microbial xylanases and their industrial applications: a review. *Applied Microbiology And Biotechnology*, 56, 326-338.

Ben-Bassat, A., Bauer, K., Chang, S.Y., Myambo, K., Boosman, A., Chang, S., 1987. Processing of the initiation methionine from proteins: properties of the Escherichia coli methionine aminopeptidase and its gene structure. *J Bacteriol*, 169, 751-7.

Benamrouche, S., Crônier, D., Debeire, P., Chabbert, B., 2002. A Chemical and Histological Study on the Effect of (1-->4)-[beta]-endo-xylanase Treatment on Wheat Bran. *Journal of Cereal Science*, 36, 253.

Berrin, J.G., Juge, N., 2008. Factors affecting xylanase functionality in the degradation of arabinoxylans. *Biotechnol Lett*, 30, 1139-50.

Bhat, M.K., 2000. Cellulases and related enzymes in biotechnology. *Biotechnology Advances*, 18, 355-383.

Boerjan, W., Ralph, J., Baucher, M., 2003. Lignin Biosynthesis. *Annual Review of Plant Biology*, 54, 519-546.

Borland, A.M., Griffiths, H., Hartwell, J., Smith, J.A.C., 2009. Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *Journal of Experimental Botany*, 60, 2879-2896.

Botstein, D., Shortle, D., 1985. Strategies and applications of in vitro mutagenesis. *Science*, 229, 1193-201.

Brett, C.T., Waldron, K.W., 1996. Physiology and biochemistry of plant cell walls. 2nd edition ed. Chapman & Hall, London.

Buranov, A.U., Mazza, G., 2008. Lignin in straw of herbaceous crops. *Industrial Crops And Products*, 28, 237-259.

Cadwell, R.C., Joyce, G.F., 1992. Randomization of genes by PCR mutagenesis. *PCR Methods Appl*, 2, 28-33.

Cadwell, R.C., Joyce, G.F., 1994. Mutagenic PCR. *PCR Methods Appl*, 3, S136-40.

Campbell, J.E., Lobell, D.B., Genova, R.C., Field, C.B., 2008. The global potential of bioenergy on abandoned agriculture lands. *Environmental Science & Technology*, 42, 5791-5794.

Carvalheiro, F., Duarte, L.C., Girio, F.M., 2008. Hemicellulose biorefineries: a review on biomass pretreatments. *Journal of Scientific & Industrial Research*, 67, 849-864.

Cervera Tison, M., Andre-Leroux, G., Lafond, M., Georis, J., Juge, N., Berrin, J.-G., 2009. Molecular determinants of substrate and inhibitor specificities of the Penicillium griseofulvum family 11 xylanases. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics*, 1794, 438-445.

Chandrakant, P., Bisaria, V.S., 1998. Simultaneous Bioconversion of Cellulose and Hemicellulose to Ethanol. *Critical Reviews in Biotechnology*, 18, 295-331.

Chang, M.C.Y., 2007. Harnessing energy from plant biomass. *Current Opinion in Chemical Biology*, 11, 677-684.

Chaparro-Riggers, J.F., Loo, B.L., Polizzi, K.M., Gibbs, P.R., Tang, X.S., Nelson, M.J., Bommarius, A.S., 2007. Revealing biases inherent in recombination protocols. *BMC Biotechnol*, 7, 77.

Chen, Y.L., Tang, T.Y., Cheng, K.J., 2001. Directed evolution to produce an alkalophilic variant from a Neocallimastix patriciarum xylanase. *Can J Microbiol*, 47, 1088-94.

Cho, S., Choi, Y., 1995. Nucleotide sequence analysis of an endo-xylanase gene (xynA) from Bacillus stearothermophilus. *J. Microbiol. Biotechnol.*, 5, 117-124.

Choct, M., 2006. Enzymes for the feed industry: past, present and future. *World's Poultry Science Journal*, 62, 5-16.

Choi, J., Choi, D.-H., Faix, O., 2007. Characterization of lignin-carbohydrate linkages in the residual lignins isolated from chemical pulps of spruce (*Picea abies*) and beech wood (*Fagus sylvatica*) *Journal of Wood Science*, 53, 309-313.

Chundawat, S.P., Balan, V., Dale, B.E., 2008. High-throughput microplate technique for enzymatic hydrolysis of lignocellulosic biomass. *Biotechnol Bioeng*, 99, 1281-94.

Chundawat, S.P., Venkatesh, B., Dale, B.E., 2007. Effect of particle size based separation of milled corn stover on AFEX pretreatment and enzymatic digestibility. *Biotechnol Bioeng*, 96, 219-31.

Cirino, P.C., Mayer, K.M., Umeno, D., 2003. Generating Mutant Libraries Using Error-Prone PCR. in: F.H. Arnold, G. Georgiou (Eds.), Directed Evolution Library Creation. Humana Press, pp. 3-9.

Clark, J.H., Dewarte, F.E.I., 2009. The Biorefinery Concept - An Integrated Approach Introduction to Chemicals from Biomass. Edited by James H. Clark and Fabien E. I. Deswarte. WILEY, pp. 1-18.

Coco, W.M., Levinson, W.E., Crist, M.J., Hektor, H.J., Darzins, A., Pienkos, P.T., Squires, C.H., Monticello, D.J., 2001. DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotech*, 19, 354.

Collins, T., Gerday, C., Feller, G., 2005. Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiology Reviews*, 29, 3.

Collins, T., Hoyoux, A., Dutron, A., Georis, J., Genot, B., Dauvrin, T., Arnaut, F., Gerday, C., Feller, G., 2006. Use of glycoside hydrolase family 8 xylanases in baking. *Journal of Cereal Science*, 43, 79-84.

Courtin, C.M., Delcour, J.A., 2001. Relative Activity of Endoxylanases Towards Water-extractable and Water-unextractable Arabinoxylan. *Journal of Cereal Science*, 33, 301-312.

Courtin, C.M., Delcour, J.A., 2002. Arabinoxylans and Endoxylanases in Wheat Flour Bread-making. *Journal of Cereal Science*, 35, 225-243.

Crameri, A., Raillard, S.-A., Bermudez, E., Stemmer, W.P.C., 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*, 391, 288.

Davies, G.J., Wilson, K.S., Henrissat, B., 1997. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J*, 321 (Pt 2), 557-9.

Demain, A.L., Newcomb, M., Wu, J.H.D., 2005. Cellulase, Clostridia, and Ethanol. *Microbiol. Mol. Biol. Rev.*, 69, 124-154.

Deniaud, E., Quemener, B., Fleurence, J., Lahaye, M., 2003. Structural studies of the mix-linked [beta]-(1-->3)/[beta]-(1-->4)--xylans from the cell wall of Palmaria palmata (Rhodophyta). *International Journal of Biological Macromolecules*, 33, 9-18.

Dodd, D., Cann, I.K.O., 2009. Enzymatic deconstruction of xylan for biofuel production. *GCB Bioenergy*, 1, 2-17.

Dodds, D.R., Gross, R.A., 2007. Chemicals from Biomass. *Science*, 318, 1250-1251.

Dogaris, I., Vakontios, G., Kalogeris, E., Mamma, D., Kekos, D., 2009. Induction of cellulases and hemicellulases from Neurospora crassa under solid-state cultivation for bioconversion of sorghum bagasse into ethanol. *Industrial Crops And Products*, 29, 404-411.

Dumon, C., Varvak, A., Wall, M.A., Flint, J.E., Lewis, R.J., Lakey, J.H., Morland, C., Luginbuhl, P., Healey, S., Todaro, T., DeSantis, G., Sun, M., Parra-Gessert, L., Tan, X., Weiner, D.P., Gilbert, H.J., 2008. Engineering hyperthermostability into a GH11 xylanase is mediated by subtle changes to protein structure. *J Biol Chem*, 283, 22557-64.

Ebringerová, A., Hromádková, Z., Heinze, T., 2005. Hemicellulose Polysaccharides I. Springer Berlin / Heidelberg, pp. 1-67.

Encell, L.P., Coates, M.M., Loeb, L.A., 1998. Engineering Human DNA Alkyltransferases for Gene Therapy Using Random Sequence Mutagenesis. *Cancer Research*, 58, 1013-1020.

Engel, R., Long, D., Carlson, G., Wallander, R., . 2005. Estimating straw production of spring and winter wheat. in: C. Jones, J. Jacobsen, E. D'imperio (Eds.), Fertilizer Facts. Montana State University Agricultural Experiment Station and Extension Service, Montana, U.S.

FAO, June 2009. Food outlook of wheat. FAO Corporate Document Repository.

Fenel, F., Leisola, M., Janis, J., Turunen, O., 2004. A de novo designed N-terminal disulphide bridge stabilizes the Trichoderma reesei endo-1,4-[beta]-xylanase II. *Journal of Biotechnology*, 108, 137-143.

Fonseca, C., Spencer-Martins, I., Hahn-Hagerdal, B., 2007. L-Arabinose metabolism in Candida arabinofermentans PYCC 5603T and Pichia guilliermondii PYCC 3012: influence of sugar and oxygen on product formation. *Appl Microbiol Biotechnol*, 75, 303-10.

Francesch, M., Geraert, P.A., 2009. Enzyme complex containing carbohydrases and phytase improves growth performance and bone mineralization of broilers fed reduced nutrient corn-soybean-based diets. *Poult Sci*, 88, 1915-1924.

Fraser-Reid, B.O., Tatsuta, K., Thiem, J., Glasser, W.G., 2008. Cellulose and Associated Heteropolysaccharides Glycoscience. Springer Berlin Heidelberg, pp. 1473-1512.

Frei, E., Preston, R.D., 1964. Non-Cellulosic Structural Polysaccharides in Algal Cell Walls. I. Xylan in Siphoneous Green Algae. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 160, 293-313.

French, C.E., 2009. Synthetic biology and biomass conversion: a match made in heaven? *J R Soc Interface*, 6 Suppl 4, S547-58.

Fromant, M., Blanquet, S., Plateau, P., 1995. Direct Random Mutagenesis of Gene-Sized DNA Fragments Using Polymerase Chain Reaction. *Analytical Biochemistry*, 224, 347-353.

Fushinobu, S., Ito, K., Konno, M., Wakagi, T., Matsuzawa, H., 1998. Crystallographic and mutational analyses of an extremely acidophilic and acid-stable xylanase: biased distribution of acidic residues and importance of Asp37 for catalysis at low pH. *Protein Engineering*, 11, 1121-1128.

Georgescu, R., Bandara, G., Sun, L., 2003. Saturation Mutagenesis Methods in Molecular Biology, pp. 75-83.

Gilbert, H.J., 2007. Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. *Molecular Microbiology*, 63, 1568-1576.

Gilbert, H.J., Hazlewood, G.P., Laurie, J.I., Orpin, C.G., Xue, G.P., 1992. Homologous catalytic domains in a rumen fungal xylanase: evidence for gene duplication and prokaryotic origin. *Molecular Microbiology*, 6, 2065-2072.

Girio, F.M., Fonseca, C., Carvalheiro, F., Duarte, L.C., Marques, S., Bogel-Lukasik, R., 2010. Hemicelluloses for fuel ethanol: A review. *Bioresource Technology*, 101, 4775-4800.

Gorgens, J.F., Planas, J., van Zyl, W.H., Knoetze, J.H., Hahn-Hagerdal, B., 2004. Comparison of three expression systems for heterologous xylanase production by S. cerevisiae in defined medium. *Yeast*, 21, 1205-17.

Goyal, H.B., Seal, D., Saxena, R.C., 2008. Bio-fuels from thermochemical conversion of renewable resources: A review. *Renewable and Sustainable Energy Reviews*, 12, 504-517.

Grabber, J.H., Hatfield, R.D., Ralph, J., 1998. Diferulate cross-links impede the enzymatic degradation of non-lignified maize walls. *Journal of the Science of Food and Agriculture*, 77, 193-200.

Grabber, J.H., Ralph, J., Lapierre, C., Barriere, Y., 2004. Genetic and molecular basis of grass cell-wall degradability. I. Lignin-cell wall matrix interactions. *C R Biol*, 327, 455-65.

Greener, A., Callahan, M., Jerpseth, B., 1997. An efficient random mutagenesis technique using an *E.coli* mutator strain. *Molecular Biotechnology*, 7, 189-195.

Gruber, K., Klintschar, G., Hayn, M., Schlacher, A., Steiner, W., Kratky, C., 1998. Thermophilic xylanase from Thermomyces lanuginosus: High-resolution X-ray structure and modeling studies. *Biochemistry*, 37, 13475-13485.

Hahn-Hagerdal, B., Karhumaa, K., Fonseca, C., Spencer-Martins, I., Gorwa-Grauslund, M.F., 2007a. Towards industrial pentose-fermenting yeast strains. *Appl Microbiol Biotechnol*, 74, 937-53.

Hahn-Hagerdal, B., Karhumaa, K., Jeppsson, M., Gorwa-Grauslund, M.F., 2007b. Metabolic engineering for pentose utilization in Saccharomyces cerevisiae. *Adv Biochem Eng Biotechnol*, 108, 147-77.

Hakulinen, N., Ossi, T., Janne, J., Matti, L., Juha, R., 2003. Three-dimensional structures of thermophilic β-1,4-xylanases from Chaetomium thermophilum and Nonomuraea flexuosa. *European Journal of Biochemistry*, 270, 1399-1412.

Haltrich, D., Nidetzky, B., Kulbe, K.D., Steiner, W., Zupancic, S., 1996. Production of fungal xylanases. *Bioresource Technology*, 58, 137-161.

Hansen, M.A.T., Kristensen, J.B., Felby, C., Jogensen, H., 2010. Pretreatment and enzymatic hydrolysis of wheat straw (Triticum aestivum L.) - The impact of lignin relocation and plant tissues on enzymatic accessibility. *Bioresource Technology*, 102, 2804-2811.

Harris, G.W., Pickersgill, R.W., Connerton, I., Debeire, P., Touzel, J.P., Breton, C., Perez, S., 1997. Structural basis of the properties of an industrially relevant thermophilic xylanase. *Proteins*, 29, 77-86.

Havukainen, R., Törrönen, A., Laitinen, T., Rouvinen, J., 1996. Covalent Binding of Three Epoxyalkyl Xylosides to the Active Site of endo-1,4-Xylanase II from Trichoderma reesei. *Biochemistry*, 35, 9617-9624.

Hendriks, A.T.W.M., Zeeman, G., 2009. Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource Technology*, 100, 10.

Henke, E., Bornscheuer, U.T., 1999. Directed Evolution of an Esterase from Pseudomonas fluorescens. Random Mutagenesis by Error-Prone PCR or a Mutator Strain and Identification of Mutants Showing Enhanced Enantioselectivity by a Resorufin-Based Fluorescence Assay. *Biological Chemistry*, 380, 1029-1033.

Henrissat, B., Bairoch, A., 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J*, 293 (Pt 3), 781-8.

Henrissat, B., Bairoch, A., 1996. Updating the sequence-based classification of glycosyl hydrolases. *Biochem J*, 316 (Pt 2), 695-6.

Himmel, M.E., Ding, S.-Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W., Foust, T.D., 2007. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science*, 315, 804-807.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., Pease, L.R., 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, 77, 51-9.

Hornsby, P.R., Hinrichsen, E., Tarverdi, K., 1997a. Preparation and properties of polypropylene composites reinforced with wheat and flax straw fibres: Part I Fibre characterization. *Journal of Materials Science*, 32, 443-449.

Hsu, C.K., Liao, J.W., Chung, Y.C., Hsieh, C.P., Chan, Y.C., 2004. Xylooligosaccharides and fructooligosaccharides affect the intestinal microbiota and precancerous colonic lesion development in rats. *J Nutr*, 134, 1523-8.

Huang, H.J., Ramaswamy, S., Tschirner, U.W., Ramarao, B.V., 2008. A review of separation technologies in current and future biorefineries. *Separation And Purification Technology*, 62, 1-21.

Hubbard, W., Biles, L., Mayfield, C., Ashton, S., (Eds.). 2007. Sustainable Forestry for Bioenergy and Bio-based Products: Trainers Curriculum Notebook. Athens, GA: Southern Forest Research Partnership, Inc.

Inami, M., Morokuma, C., Sugio, A., Tamanoi, H., Yatsunami, R., Nakamura, S., 2003. Directed evolution of xylanase J from alkaliphilic Bacillus sp. strain 41M-1: restore of alkaliphily of a mutant with an acidic pH optimum. *Nucleic Acids Res Suppl*, 315-6.

Ito, K., Ikemasu, T., Ishikawa, T., 1992. Cloning and sequencing of the xynA gene encoding xylanase A of Aspergillus kawachii. *Biosci Biotechnol Biochem*, 56, 906-12.

Janis, J., Hakanpaa, J., Hakulinen, N., Ibatullin, F.M., Hoxha, A., Derrick, P.J., Rouvinen, J., Vainiotalo, P., 2005. Determination of thioxylo-oligosaccharide binding to family 11 xylanases using electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and X-ray crystallography. *FEBS J*, 272, 2317-33.

Jeffries, T.W., 1996. Biochemistry and genetics of microbial xylanases. *Current Opinion in Biotechnology*, 7, 337-342.

Jiang, Z.Q., Deng, W., Yan, Q.J., Zhai, Q., Li, L.T., Kusakabe, I., 2006. Subunit composition of a large xylanolytic complex (xylanosome) from Streptomyces olivaceoviridis E-86. *Journal of Biotechnology*, 126, 304-312.

Joern, J.M., 2003. DNA Shuffling Methods in Molecular Biology, pp. 85-89.

Johannes, T.W., Zhao, H., 2006. Directed evolution of enzymes and biosynthetic pathways. *Curr Opin Microbiol*, 9, 261-7.

Kamondi, S., Szilagyi, A., Barna, L., Zavodszky, P., 2008. Engineering the thermostability of a TIM-barrel enzyme by rational family shuffling. *Biochem Biophys Res Commun*, 374, 725-30.

Katapodis, P., Vrsanska, M., Kekos, D., Nerinckx, W., Biely, P., Claeyssens, M., Macris, B.J., Christakopoulos, P., 2003. Biochemical and catalytic properties of an endoxylanase purified from the culture filtrate of Sporotrichum thermophile. *Carbohydr Res*, 338, 1881-90.

Kenealy, W.R., Jeffries, T.W., 2003. Rapid 2,2'-bicinchoninic-based xylanase assay compatible with high throughput screening. *Biotechnol Lett*, 25, 1619-23.

Kerstetter, J.D., Lyons, J.K., 2001. Wheat straw for ethanol production in Washington: a resource, technical, and economic assessment. Washington State University. Cooperative Extension Energy Program. http://cru.cahe.wsu.edu/CEPublications/wsuceep01-084/wsuceep01-084.pdf.

Kikuchi, M., Ohnishi, K., Harayama, S., 1999. Novel family shuffling methods for the in vitro evolution of enzymes. *Gene*, 236, 159.

Kikuchi, M., Ohnishi, K., Harayama, S., 2000. An effective family shuffling method using single-stranded DNA. *Gene*, 243, 133-7.

Kolenov, K., Vrsansk, M., Biely, P., 2006. Mode of action of endo-[beta]-1,4-xylanases of families 10 and 11 on acidic xylooligosaccharides. *Journal of Biotechnology*, 121, 338-345.

Koshijima, T., Watanabe, T., 2003. Association between Lignin and Carbohydrates in Wood and Other Plant Tissues. Springer Series in Wood Science. Springer.

Koutinas, A.A., Du, C., Wang, R.H., Webb, C., 2009. Production of chemicals from biomass Introduction to Chemicals from Biomass. Edited by James H. Clark and Fabien E. I. Deswarte. WILEY, pp. 1-18.

Kristensen, J.B., Thygesen, L.G., Felby, C., Jorgensen, H., Elder, T., 2008. Cell-wall structural changes in wheat straw pretreated for bioethanol production. *Biotechnology For Biofuels*, 1.

Kulkarni, N., Shendye, A., Rao, M., 1999. Molecular and biotechnological aspects of xylanases. *FEMS Microbiol Rev*, 23, 411-56.

Kumar, R., Singh, S., Singh, O.V., 2008. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *Journal Of Industrial Microbiology & Biotechnology*, 35, 377-391.

Kumar, R., Wyman, C.E., 2009b. Effect of xylanase supplementation of cellulase on digestion of corn stover solids prepared by leading pretreatment technologies. *Bioresour Technol*, 100, 4203-13.

La Grange, D.C., Pretorius, I.S., Claeyssens, M., van Zyl, W.H., 2001. Degradation of Xylan to D-Xylose by Recombinant Saccharomyces cerevisiae Coexpressing the Aspergillus nigerbeta -Xylosidase (xlnD) and the Trichoderma reesei Xylanase II (xyn2) Genes. *Appl. Environ. Microbiol.*, 67, 5512-5519.

Lagaert, S., Belien, T., Volckaert, G., 2009. Plant cell walls: Protecting the barrier from degradation by microbial enzymes. *Seminars In Cell & Developmental Biology*, 20, 1064-1073.

Lai, Y.-P., Huang, J., Wang, L.-F., Li, J., Wu, Z.-R., 2004. A new approach to random mutagenesis in vitro. *Biotechnology and Bioengineering*, 86, 622.

Lapierre, C., Monties, B., Rolando, C., 1986. Thioacidolysis of poplar lignins - identification of monomeric syringyl products and characterization of guaiacyl-syringyl lignin fractions. *Holzforschung*, 40, 113-118.

Lee, J., Heo, S.-Y., Lee, J.-W., Yoon, K.-H., Kim, Y.-H., Nam, S.-W., 2009. Thermostability and xylan-hydrolyzing property of endoxylanase expressed in yeast *Saccharomyces cerevisiae*. *Biotechnology and Bioprocess Engineering*, 14, 639-644.

Lequart, C., Nuzillard, J.-M., Kurek, B., Debeire, P., 1999. Hydrolysis of wheat bran and straw by an endoxylanase: production and structural characterization of cinnamoyl-oligosaccharides. *Carbohydrate Research*, 319, 102.

Lilly, M., Fierobe, H.P., van Zyl, W.H., Volschenk, H., 2009. Heterologous expression of a Clostridium minicellulosome in Saccharomyces cerevisiae. *FEMS Yeast Res*, 9, 1236-49.

Lin, L.L., Thomson, J.A., 1991. An analysis of the extracellular xylanases and cellulases of butyrivibrio-fibrisolvens H17C. *FEMS Microbiology Letters*, 84, 197-204.

Lin, Y., Tanaka, S., 2006. Ethanol fermentation from biomass resources: current state and prospects. *Appl Microbiol Biotechnol*, 69, 627-42.

LinGoerke, J.L., Robbins, D.J., Burczak, J.D., 1997. PCR-based random mutagenesis using manganese and reduced dNTP concentration. *Biotechniques*, 23, 409-412.

Liu, R., Yu, H., Huang, Y., 2005. Structure and morphology of cellulose in wheat straw. *Cellulose*, 12, 25-34.

Ludwiczek, M.L., Heller, M., Kantner, T., McIntosh, L.P., 2007. A secondary xylan-binding site enhances the catalytic activity of a single-domain family 11 glycoside hydrolase. *J Mol Biol*, 373, 337-54.

Lynd, L.R., 1996. Overview and Evaluation of Fuel Ethanol from Cellulosic Biomass: Technology, Economics, the Environment, and Policy. *Annual Review of Energy and the Environment*, 21, 403-465.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., Pretorius, I.S., 2002. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev*, 66, 506-77, table of contents.

Lynd, L.R., Zyl, W.H.v., McBride, J.E., Laser, M., 2005. Consolidated bioprocessing of cellulosic biomass: an update. *Current Opinion in Biotechnology*, 16, 577.

Martin, C., Thomsen, M.H., Hauggaard-Nielsen, H., Belindathomsen, A., 2008. Wet oxidation pretreatment, enzymatic hydrolysis and simultaneous saccharification and fermentation of clover-ryegrass mixtures. *Bioresour Technol*, 99, 8777-82.

McCarter, J.D., Withers, S.G., 1994. Mechanisms of enzymatic glycoside hydrolysis. *Curr Opin Struct Biol*, 4, 885-92.

McCarthy, A.A., Morris, D.D., Bergquist, P.L., Baker, E.N., 2000. Structure of XynB, a highly thermostable beta-1,4-xylanase from Dictyoglomus thermophilum Rt46B.1, at 1.8 A resolution. *Acta Crystallogr D Biol Crystallogr*, 56, 1367-75.

Mckean, W.T., Jacobs, R.S., 2007. Wheat straw as a paper fiber source. The Clean Washington Center http://www.cwc.org/paper/pa971rpt.pdf.

McKendry, P., 2002. Energy production from biomass (part 2): conversion technologies. *Bioresource Technology*, 83, 47.

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

Michaux, C., Pouyez, J., Mayard, A., Vandurm, P., Housen, I., Wouters, J., 2010. Structural insights into the acidophilic pH adaptation of a novel endo-1,4-beta-xylanase from Scytalidium acidophilum. *Biochimie*, 92, 1407-1415.

Miller, G.L., 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*, 31, 426.

Miyazaki, K., Takenouchi, M., Kondo, H., Noro, N., Suzuki, M., Tsuda, S., 2006. Thermal stabilization of Bacillus subtilis family-11 xylanase by directed evolution. *J Biol Chem*, 281, 10236-42.

Möller, R., Toonen, M.A.J., Beilen, J.B., Salentijn, E.M.J., Clayton, D., 2007. Crop Platforms for cell wall biorefining : lignocellulose feedstocks. CPL Press.

Moore, J.C., Jin, H.M., Kuchner, O., Arnold, F.H., 1997. Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. *J Mol Biol*, 272, 336-47.

Muilu, J., Törrönen, A., Peräkylä, M., Rouvinen, J., 1998. Functional conformational changes of endo-1,4-xylanase II from Trichoderma reesei: a molecular dynamics study. *Proteins*, 31, 434-44.

Murashima, K., Kosugi, A., Doi, R.H., 2003. Synergistic effects of cellulosomal xylanase and cellulases from Clostridium cellulovorans on plant cell wall degradation. *J Bacteriol*, 185, 1518-24.

Navarro, D., Couturier, M., da Silva, G., Berrin, J.-G., Rouau, X., Asther, M., Bignon, C., 2010. Automated assay for screening the enzymatic release of reducing sugars from micronized biomass. *Microbial Cell Factories*, 9, 58.

Neylon, C., 2004. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res*, 32, 1448-59.

Nieves, R.A., Ehrman, C.I., Adney, W.S., Elander, R.T., Himmel, M.E., 1998. Survey and analysis of commercial cellulase preparations suitable for biomass conversion to ethanol. *World Journal of Microbiology and Biotechnology*, 14, 301-304.

Nishiyama, Y., Langan, P., Chanzy, H., 2002. Crystal structure and hydrogen-bonding system in cellulose 1 beta from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society*, 124, 9074-9082.

Numan, M., Bhosle, N., 2006. α-L-Arabinofuranosidases: the potential applications in biotechnology. *Journal of Industrial Microbiology and Biotechnology*, 33, 247.

Ohara, H., 2003. Biorefinery. *Applied Microbiology And Biotechnology*, 62, 474-477.

Okazaki, F., Tamaru, Y., Hashikawa, S., Li, Y.-T., Araki, T., 2002. Novel Carbohydrate-Binding Module of beta-1,3-Xylanase from a Marine Bacterium, Alcaligenes sp. Strain XY-234. *J. Bacteriol.*, 184, 2399-2403.

Olofsson, K., Bertilsson, M., Liden, G., 2008. A short review on SSF - an interesting process option for ethanol production from lignocellulosic feedstocks. *Biotechnol Biofuels*, 1, 7.

Olsson, L., Viikari, L., Alapuranen, M., Puranen, T., Vehmaanperä, J., Siika-aho, M., 2007. Thermostable Enzymes in Lignocellulose Hydrolysis Biofuels. Springer Berlin / Heidelberg, pp. 121-145.

Pack, S.P., Yoo, Y.J., 2004. Protein thermostability: structure-based difference of amino acid between thermophilic and mesophilic proteins. *Journal of Biotechnology*, 111, 269.

Paës, G., 2005. Etude sturcture/fonctiond'hémicellulases thermostables: la xylanase GH-11 et l'arabinofuranosidase GH-51 de Thermobacillus xylanilyticus. Université de Reims Champagne-Ardenne, INRA UMR Fare 614.

Paës, G., O'Donohue, M.J., 2006. Engineering increased thermostability in the thermostable GH-11 xylanase from Thermobacillus xylanilyticus. *J Biotechnol*, 125, 338-50.

Paës, G., Skov, L.K., O'Donohue, M.J., Rémond, C., Kastrup, J.S., Gajhede, M., Mirza, O., 2008. The Structure of the Complex between a Branched Pentasaccharide and Thermobacillus xylanilyticus GH-51 Arabinofuranosidase Reveals Xylan-Binding Determinants and Induced Fitâ€ â€¡. *Biochemistry*, 47, 7441.

Paës, G., Tran, V., Takahashi, M., Boukari, I., O'Donohue, M.J., 2007. New insights into the role of the thumb-like loop in GH-11 xylanases. *Protein Eng Des Sel*, 20, 15-23.

Palackal, N., Brennan, Y., Callen, W.N., Dupree, P., Frey, G., Goubet, F., Hazlewood, G.P., Healey, S., Kang, Y.E., Kretz, K.A., Lee, E., Tan, X., Tomlinson, G.L., Verruto, J., Wong, V.W., Mathur, E.J., Short, J.M., Robertson, D.E., Steer, B.A., 2004. An evolutionary route to xylanase process fitness. *Protein Sci*, 13, 494-503.

Payan, F., Leone, P., Porciero, S., Furniss, C., Tahir, T., Williamson, G., Durand, A., Manzanares, P., Gilbert, H.J., Juge, N., Roussel, A., 2004. The Dual Nature of the

Wheat Xylanase Protein Inhibitor XIP-I. *Journal of Biological Chemistry*, 279, 36029-36037.

Pedersen, M., Meyer, A.S., 2009. Influence of substrate particle size and wet oxidation on physical surface structures and enzymatic hydrolysis of wheat straw. *Biotechnology Progress*, 25, 399-408.

Pelletier, J.N., 2001. A RACHITT for our toolbox. *Nature Biotechnology*, 19, 314-315.

Petsko, G.A., Ringe, D., 2004. Protein Structure and Function. Chapter I. From sequence to structure. New Science Press Ltd.

Polizeli, M., Rizzatti, A.C.S., Monti, R., Terenzi, H.F., Jorge, J.A., Amorim, D.S., 2005. Xylanases from fungi: properties and industrial applications. *Applied Microbiology And Biotechnology*, 67, 577-591.

Pollet, A., Delcour, J.A., Courtin, C.M., 2010. Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. *Crit Rev Biotechnol*, 30, 176-91.

Pollet, A., Vandermarliere, E., Lammertyn, J., Strelkov, S.V., Delcour, J.A., Courtin, C.M., 2009. Crystallographic and activity-based evidence for thumb flexibility and its relevance in glycoside hydrolase family 11 xylanases. *Proteins-Structure Function and Bioinformatics*, 77, 395-403.

Prade, R.A., 1996. Xylanases: From biology to biotechnology Biotechnology and Genetic Engineering Reviews, Vol 13. Intercept Ltd, Andover, pp. 101-131.

Purmonen, M., Valjakka, J., Takkinen, K., Laitinen, T., Rouvinen, J., 2007. Molecular dynamics studies on the thermostability of family 11 xylanases. *Protein Engineering, Design and Selection*, gzm056.

Ragauskas, A.J., Poll, K.M., Cesternino, A.J., 1994. Effects of xylanase pretreatment procedures on nonchlorine bleaching. *Enzyme and Microbial Technology*, 16, 492-495.

Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J., Jr., Hallett, J.P., Leak, D.J., Liotta, C.L., Mielenz, J.R., Murphy, R., Templer, R., Tschaplinski, T., 2006. The Path Forward for Biofuels and Biomaterials. *Science*, 311, 484-489.

Ralph, J., Grabber, J.H., Hatfield, R.D., 1995. Lignin-ferulate cross-links in grasses: active incorporation of ferulate polysaccharide esters into ryegrass lignins. *Carbohydrate Research*, 275, 167-178.

Ralph, J., Hatfield Ronald, D., Grabber John, H., Jung Hans-Joachim, G., Quideau, S., Helm Richard, F., 1998. Cell Wall Cross-Linking in Grasses by Ferulates and Diferulates Lignin and Lignan Biosynthesis. American Chemical Society, pp. 209-236.

Rémond, C., Aubry, N., Cronier, D., Noel, S., Martel, F., Roge, B., Rakotoarivonina, H., Debeire, P., Chabbert, B., 2010. Combination of ammonia and xylanase pretreatments: Impact on enzymatic xylan and cellulose recovery from wheat straw. *Bioresource Technology*, 101, 6712-6717.

Roncero, M.B., Torres, A.L., Colom, J.F., Vidal, T., 2005. The effect of xylanase on lignocellulosic components during the bleaching of wood pulps. *Bioresource Technology*, 96, 21-30.

Ruiz, E., Cara, C., Manzanares, P., Ballesteros, M., Castro, E., 2008. Evaluation of steam explosion pre-treatment for enzymatic hydrolysis of sunflower stalks. *Enzyme and Microbial Technology*, 42, 160-166.

Ruller, R., Deliberto, L., Ferreira, T.L., Ward, R.J., 2008. Thermostable variants of the recombinant xylanase A from Bacillus subtilis produced by directed evolution show reduced heat capacity changes. *Proteins*, 70, 1280-93.

Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jorgensen, P.L., Schulein, M., Dupont, C., Davies, G.J., Wilson, K.S., 1999. Catalysis and specificity in enzymatic glycoside hydrolysis: a 2,5B conformation for the glycosyl-enzyme intermediate revealed by the structure of the Bacillus agaradhaerens family 11 xylanase. *Chem Biol*, 6, 483-92.

Saha, B., 2003. Hemicellulose bioconversion. *Journal of Industrial Microbiology and Biotechnology*, 30, 279-291.

Saha, B.C., Cotta, M.A., 2006. Ethanol production from alkaline peroxide pretreated enzymatically saccharified wheat straw. *Biotechnol Prog*, 22, 449-53.

Salmones, D., Mata, G., Waliszewski, K.N., 2005. Comparative culturing of Pleurotus spp. on coffee pulp and wheat straw: biomass production and substrate biodegradation. *Bioresource Technology*, 96, 537-544.

Samain, E., Debeire, P., Touzel, J.P., 1997. High level production of a cellulase-free xylanase in glucose-limited fed batch cultures of a thermophilic Bacillus strain. *Journal Of Biotechnology*, 58, 71-78.

Samain, E., Debeire, P., Debeire-Gosselin, M., Touzel, J.P., , 1991. Xylanase, souches de Bacillus productrices de xylanases et leurs utilisations. Patent FR-9101191.

Samain, E., Touzel, J.P., Brodel, B., Debeire, P.,, 1992. Isolation of a thermophilic bacterium producing high levels of xylanases. in: J. Visser, Beldman, G., Kusters-van Someren, M.A., Voragen, A.G.J., (Ed.) Xlyan and Xylanases. Elsevier Science Publishers B.V., pp. 467-470.

Sánchez, Ó.J., Cardona, C.A., 2008. Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresource Technology*, 99, 5270-5295.

Sapag, A., Wouters, J., Lambert, C., de Ioannes, P., Eyzaguirre, J., Depiereux, E., 2002. The endoxylanases from family 11: computer analysis of protein sequences reveals important structural and phylogenetic relationships. *J Biotechnol*, 95, 109-31.

Saxena, R.C., Seal, D., Kumar, S., Goyal, H.B., 2008. Thermo-chemical routes for hydrogen rich gas from biomass: A review. *Renewable and Sustainable Energy Reviews*, 12, 1909-1927.

Scheller, H.V., Ulvskov, P., 2010. Hemicelluloses. *Annual Review of Plant Biology*, 61, 263-289.

Schmitt, E., Guillon, J.M., Meinnel, T., Mechulam, Y., Dardel, F., Blanquet, S., 1996. Molecular recognition governing the initiation of translation in Escherichia coli. A review. *Biochimie*, 78, 543-54.

Shallom, D., Shoham, Y., 2003. Microbial hemicellulases. *Curr Opin Microbiol*, 6, 219-28.

Shibuya, H., Kaneko, S., Hayashi, K., 2000. Enhancement of the thermostability and hydrolytic activity of xylanase by random gene shuffling. *Biochem J*, 349, 651-6.

Shigeru Deguchi, Kaoru Tsujiib, Horikoshi, K., 2006. Cooking cellulose in hot and compressed water. *Chemical Communications*.

Shimokawa, T., Shibuya, H., Nojiri, M., Yoshida, S., Ishihara, M., 2008. Purification, Molecular Cloning, and Enzymatic Properties of a Family 12 Endoglucanase (EG-II) from Fomitopsis palustris: Role of EG-II in Larch Holocellulose Hydrolysis. *Appl. Environ. Microbiol.*, 74, 5857-5861.

Sidhu, G., Withers, S.G., Nguyen, N.T., McIntosh, L.P., Ziser, L., Brayer, G.D., 1999. Sugar Ring Distortion in the Glycosyl-Enzyme Intermediate of a Family G/11 Xylanase. *Biochemistry*, 38, 5346-5354.

Sinnott ML, 1990. Catalytic mechanisms of enzymic glycoside transfer. *Chem. Rev.*, 90, 1170-1202.

Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredez, A., Persson, S., Raab, T., Vorwerk, S., Youngs, H., 2004. Toward a Systems Approach to Understanding Plant Cell Walls. *Science*, 306, 2206-2211.

Somerville, C., Youngs, H., Taylor, C., Davis, S.C., Long, S.P., 2010. Feedstocks for Lignocellulosic Biofuels. *Science*, 329, 790-792.

Sørensen, H.R., Pedersen, S., Jørgensen, C.T., 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. *Biotechnology Progress*, 23, 100-107.

Stemmer, W.P., 1994b. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences*, 91, 10747-10751.

Stemmer, W.P.C., 1994a. Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, 370, 389-391.

Stenberg, K., Bollok, M., Reczey, K., Galbe, M., Zacchi, G., 2000. Effect of substrate and cellulase concentration on simultaneous saccharification and fermentation of steam-pretreated softwood for ethanol production. *Biotechnol Bioeng*, 68, 204-10.

Stephens, D.E., Rumbold, K., Permaul, K., Prior, B.A., Singh, S., 2007. Directed evolution of the thermostable xylanase from Thermomyces lanuginosus. *Journal of Biotechnology*, 127, 348-354.

Stephens, D.E., Singh, S., Permaul, K., 2009. Error-prone PCR of a fungal xylanase for improvement of its alkaline and thermal stability. *FEMS Microbiol Lett*.

Sticklen, M., 2006. Plant genetic engineering to improve biomass characteristics for biofuels. *Current Opinion in Biotechnology*, 17, 315-319.

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

Sun, J.-Y., Liu, M.-Q., Xu, Y.-L., Xu, Z.-R., Pan, L., Gao, H., 2005. Improvement of the thermostability and catalytic activity of a mesophilic family 11 xylanase by N-terminus replacement. *Protein Expression And Purification*, 42, 122-130.

Sun, R., Sun, X.F., Wang, S.Q., Zhu, W., Wang, X.Y., 2002. Ester and ether linkages between hydroxycinnamic acids and lignins from wheat, rice, rye, and barley straws, maize stems, and fast-growing poplar wood. *Industrial Crops and Products*, 15, 179-188.

Sun, R., Tomkinson, J., 2002. Comparative study of lignins isolated by alkali and ultrasound-assisted alkali extractions from wheat straw. *Ultrasonics Sonochemistry*, 9, 85-93.

Sung, W., L., 2007. Modification of Xylanases to Increase Therophilicity, Thermostability and Alkalophilicity, Patent, Application No. PCT/CA2006/001192

Sunna, A., Antranikian, G., 1997. Xylanolytic enzymes from fungi and bacteria. *Crit Rev Biotechnol*, 17, 39-67.

Swennen, K., Courtin, C.M., Delcour, J.A., 2006. Non-digestible oligosaccharides with prebiotic properties. *Crit Rev Food Sci Nutr*, 46, 459-71.

Takashi, W., 1989. Structural studies on the covalent bonds between lignin and carbohydrate in lignin-carbohydrate complexes by selective oxidation of the lignin with 2,3-dichloro-5,6-dicyano-1,4-benzoquinone,. Wood Research Institute, Kyoto University. http://hdl.handle.net/2433/53279.

Tenkanen, M., Tamminen, T., Hortling, B., 1999. Investigation of lignin-carbohydrate complexes in kraft pulps by selective enzymatic treatments. *Applied Microbiology And Biotechnology*, 51, 241-248.

Törrönen, A., Rouvinen, J., 1995. Structural comparison of two major endo-1,4-xylanases from Trichoderma reesei. *Biochemistry*, 34, 847-56.

Touzel, J.P., O'Donohue, M., Debeire, P., Samain, E., Breton, C., 2000. Thermobacillus xylanilyticus gen. nov., sp. nov., a new aerobic thermophilic xylan-degrading bacterium isolated from farm soil. *Int J Syst Evol Microbiol*, 50, 315-320.

Turvey, J.R., Williams, E.L., 1970. The structures of some xylans from red algae. *Phytochemistry*, 9, 2383-2388.

U.S.DOE, 2006. Breaking the biological barriers to cellulosic ethanol: A joint research agenda. . U.S. Department of Energy Office of Science and Office of Energy Efficiency and Renewable Energy. , pp. 206.

Vahjen, W., Gl, Auml, Ser, K., Sch, Fer, K., Simon, O., 1998. Influence of xylanase-supplemented feed on the development of selected bacterial groups in the intestinal tract of broiler chicks. *The Journal of Agricultural Science*, 130, 489-500.

van Zyl, W.H., Lynd, L.R., den Haan, R., McBride, J.E., 2007. Consolidated bioprocessing for bioethanol production using Saccharomyces cerevisiae. *Adv Biochem Eng Biotechnol*, 108, 205-35.

Vandermarliere, E., Bourgois, T.M., Rombouts, S., Van Campenhout, S., Volckaert, G., Strelkov, S.V., Delcour, J.A., Rabijns, A., Courtin, C.M., 2008. Crystallographic analysis shows substrate binding at the -3 to +1 active-site subsites and at the surface of glycoside hydrolase family 11 endo-1,4-beta-xylanases. *Biochem J*, 410, 71-9.

Vardakou, M., Dumon, C., Murray, J.W., Christakopoulos, P., Weiner, D.P., Juge, N., Lewis, R.J., Gilbert, H.J., Flint, J.E., 2008. Understanding the structural basis for substrate and inhibitor recognition in eukaryotic GH11 xylanases. *J Mol Biol*, 375, 1293-305.

Viikari, L., Kantelinen, A., Sundquist, J., Linko, M., 1994. Xylanases in bleaching: From an idea to the industry. *FEMS Microbiology Reviews*, 13, 335-350.

Vincken, J.-P., Beldman, G., Voragen, A.G.J., 1997. Substrate specificity of endoglucanases: what determines xyloglucanase activity? *Carbohydrate Research*, 298, 299-310.

Wakarchuk, W.W., Campbell, R.L., Sung, W.L., Davoodi, J., Yaguchi, M., 1994. Mutational and crystallographic analyses of the active site residues of the Bacillus circulans xylanase. *Protein Sci*, 3, 467-75.

Wakarchuk, W.W., Sung, W.L., Campbell, R.L., Cunningham, A., Watson, D.C., Yaguchi, M., 1994. Thermostabilization of the Bacillus circulans xylanase by the introduction of disulfide bonds. *Protein Eng*, 7, 1379-86.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25, 1189-1191.

Wong, K.K., Tan, L.U., Saddler, J.N., 1988. Multiplicity of beta-1,4-xylanase in microorganisms: functions and applications. *Microbiol Rev*, 52, 305-17.

Wouters, J., Georis, J., Engher, D., Vandenhaute, J., Dusart, J., Frere, J.M., Depiereux, E., Charlier, P., 2001. Crystallographic analysis of family 11 endo-beta-1,4-xylanase Xyl1 from Streptomyces sp. S38. *Acta Crystallogr D Biol Crystallogr*, 57, 1813-9.

Wyman, C.E., 2003. Potential synergies and challenges in refining cellulosic biomass to fuels, chemicals, and power. *Biotechnol Prog*, 19, 254-62.

Wyman, C.E., 2007. What is (and is not) vital to advancing cellulosic ethanol. *Trends in Biotechnology*, 25, 153.

Xie, H., Flint, J., Vardakou, M., Lakey, J.H., Lewis, R.J., Gilbert, H.J., Dumon, C., 2006. Probing the structural basis for the difference in thermostability displayed by family 10 xylanases. *J Mol Biol*, 360, 157-67.

Yang, H., Yan, R., Chen, H., Lee, D.H., Zheng, C., 2007. Characteristics of hemicellulose, cellulose and lignin pyrolysis. *Fuel*, 86, 1781-1788.

You, C., Huang, Q., Xue, H., Xu, Y., Lu, H., 2010. Potential hydrophobic interaction between two cysteines in interior hydrophobic region improves thermostability of a family 11 xylanase from Neocallimastix patriciarum. *Biotechnology and Bioengineering*, 105, 861-870.

Yuan, J.S., Tiller, K.H., Al-Ahmad, H., Stewart, N.R., Stewart Jr, C.N., 2008. Plants to power: bioenergy to fuel the future. *Trends in Plant Science*, 13, 421.

Zha, W., Zhu, T., Zhao, H., 2003. Family Shuffling with Single-Stranded DNA, pp. 91-97.

Zhang, P.Y.H., Himmel, M.E., Mielenz, J.R., 2006. Outlook for cellulase improvement: Screening and selection strategies. *Biotechnology Advances*, 24, 452-481.

Zhang, S., Zhang, K., Chen, X., Chu, X., Sun, F., Dong, Z., 2010. Five mutations in N-terminus confer thermostability on mesophilic xylanase. *Biochemical and Biophysical Research Communications*, 395, 200-206.

Zhang, Z.-G., Yi, Z.-L., Pei, X.-Q., Wu, Z.-L., 2010. Improving the thermostability of Geobacillus stearothermophilus xylanase XT6 by directed evolution and site-directed mutagenesis. *Bioresource Technology*, 101, 9272.

Zhao, H., Arnold, F.H., 1997. Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res*, 25, 1307-8.

Zhao, H.M., Giver, L., Shao, Z.X., Affholter, J.A., Arnold, F.H., 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology*, 16, 258-261.

Zhao, H.M., Zha, W.J., 2006. In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP). *Nature Protocols*, 1, 1865-1871.

Zilliox, C., Debeire, P., 1998. Hydrolysis of wheat straw by a thermostable endoxylanase: Adsorption and kinetic studies. *Enzyme and Microbial Technology*, 22, 58.

Zimmermann, W., 1991. Degradation of Environmental Pollutants by Microorganisms and their Metalloenzymes. Marcel Dekker, New York

Zvelebil, M.J.J.M., Sternberg, M.J.E., 1988. Analysis and prediction of the location of catalytic residues in enzymes. *Protein Eng.*, 2, 127-138.

ARTICLE I

A HIGH THROUGHPUT SCREENING SYSTEM FOR THE EVALUATION OF BIOMASS-HYDROLYZING

This article has been published in the journal of *Bioresource Technology*, 2010. Here, we described how we designed a novel screening method that uses wheat straw as substrate, and how we validated its relevance in different conditions. In addition, the comparison of this screening method to the other similar screens was mentioned as well. Most importantly, the established screening system provides technique support for the following random engineering study.

A high-throughput screening system for the evaluation of biomass-hydrolyzing glycoside hydrolases.

Letian Song [a,b,c], Sandrine Laguerre [a,b,c], Claire Dumon [a,b,c], Sophie Bozonnet [a,b,c], Michael J. O'Donohue [a,b,c]*

[a] *Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France;*

[b] *INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France*

[c] *CNRS, UMR5504, F-31400 Toulouse, France*

Abstract

To implement a protein engineering strategy for the improvement of enzyme performance on biomass, we have devised a straightforward, robust high throughput method using wheat straw and a recombinant GH11 xylanase as a test case scenario. Inevitably, the method requires automated liquid handling equipment, but it avoids the need for specialized milling and powder weighing devices and the use of labour intensive steps such as manual cutting of pipette tips. After expression in *E. coli* cells grown in microtiter plates, recombinant xylanase was released into the culture medium and used directly for biomass hydrolysis. Reactions were monitored using a micro-DNS assay. The cumulative error of the method was less than 15%. To validate the method, randomly generated xylanase mutants were analyzed. This allowed the detection of one mutant, which produced a 74 % increase in hydrolysis compared to the parental enzyme. Closer analysis revealed that this increase in activity was correlated with a two-fold increase in xylanase expression.

Keywords

I.1 Introduction

Lignocellulosic biomass is the most abundant terrestrial repository of renewable carbon (McKendry, 2002; Wyman, 2007). In this respect, the development of efficient technologies that will give access to its various constituents and provide the means to produce fuels, chemicals and materials is a vital challenge for modern society (Himmel et al., 2007).

The breakdown of lignocellulosic biomass into its constituent subunits is the first step in biorefining and the choice of technologies that are used is critical, because to a large extent it will determine the economic and environmental sustainability of the overall value chain. In state of the art concepts, it is envisaged that hemicellulose and cellulose will be broken down into fermentable sugar syrups in two separate unit operations. Hemicellulose is targeted in the initial pretreatment step that often employs a combination of chemical catalysis and thermo-physical effects, while cellulose is hydrolysed using appropriate enzyme cocktails. However, it is widely acknowledged that current pretreatment technologies lead to both high capital investment and high operational costs (Mosier et al., 2005; Wyman, 2007). Equally, it is recognized that bioconversion processes involving microorganisms and/or enzymes will progressively contribute to the improvement of the environmental and economic performance of today's industrial processes (Demain, 2007), notably through the development of consolidated bioprocessing of lignocellulosic biomass (Lynd et al., 2005). Nevertheless, for biocatalysts to become the linchpins of future biorefining processes it is necessary to accelerate both the discovery and engineering of enzymes and microorganisms. To achieve this, high throughput technologies such as functional metagenomics and *in vitro* enzyme evolution strategies are being increasingly employed to identify and/or engineer enzymes that are particularly adapted to the constraints of the industrial environment and to the specific demands of targeted processes.

To devise strategies to isolate new enzymes arising from natural or artificially-created biodiversity, it is vital to account for the truism "you get what you screen for" (Schmidt-Dannert and Arnold, 1999). In the case of biomass deconstruction, this implies that the chemical and structural complexity of the natural substrate must be reflected in the screen,

thus the choice of substrate is a key issue. To illustrate this point, it is noteworthy that the use of CMC-cellulose as a substrate for the selection of improved activity in mutant cellulase populations has failed to isolate enzymes that display higher activity on crystalline cellulose (Lin et al., 2009; Nakazawa et al., 2009). Presumably, the use of filter paper in these experiments would have been preferable, even though this still does not account for the presence of residual lignin-derived inhibitors present in chemically-pretreated biomass. Similarly, conventional screening methods for hemicellulases are also unsuitable. These are based on the use of isolated, often modified, polysaccharides such as RBB-xylan, which is added to solid agar medium (Shibuya et al., 2000; Stephens et al., 2007). Practically-speaking, these substrates, which allow the detection of lytic haloes around colonies, provide a rapid way to discriminate between active and inactive mutants, but there is no clear relationship between activity on these substrates and activity on complex biomass (Chundawat et al., 2008; Zhang et al., 2006).

In this research study, we have focused on the development of simple and reliable screening approach that can be used for automated high throughput screening (HTS) of biomass deconstructing enzymes or enzyme cocktails, using complex biomass as the substrate. In a previous study, Chundawat *et al.* (2008) described a similar high throughput screening approach, which was used to assay the cellulase-mediated hydrolysis of AFEX-pretreated corn stover. Their method relied upon the use of finely ground biomass (<100 μm) and a glucose-specific detection assay. In the method described here, we have attempted to both design a more generic HTS method using an alternative biomass distribution approach, which does not require high performance milling equipment, and a reducing sugar micro-assay using the 3,5-dinitrosalicylic acid (DNS) reagent. To illustrate the performance of our method we have prepared a library of randomly mutated variants of the GH11 xylanase from *Thermobacillus xylanilyticus* (designated as Tx-Xyl) that has been assayed for improved activity on untreated wheat straw, the aim being to increase hydrolysis efficiency and improve hemicellulose solubilisation beyond the current 18% (w/w, dry weight) level.

I.2 Materials and methods

I.2.1 General materials and chemicals

Unless otherwise stated, all chemicals were analytical grade and purchased from Sigma-Aldrich (St. Louis, MO, USA). Restriction enzymes, *Taq* DNA polymerase and their corresponding buffers were obtained from New England Biolabs (Ipswich, MA, USA). Oligonucleotides were synthesized by Eurogentec (Angers, France). Microtiter plates and their accessories were purchased either from Evergreen Scientific (Los Angeles, CA, USA) (sterile plates for bacterial growth) or from Corning Corp. (NY, USA).

I.2.2 Biomass preparation

Wheat straw (Apache variety), harvested in 2007 in southern France, was obtained from ARD (Pomacle, France). Subsequently, the straw was ground to powder form (0.5 mm) and washed with distilled water at 4°C (> 10 volumes). Straw powder was recovered by filtration using a Büchner funnel with Whatman® No.4 filter paper (pore size: 20-25 µm), then dried in a dry oven. The dried powder was then sterilized by autoclaving.

I.2.3 Xylanase preparation, purification and activity assay

A recombinant form of the GH11 endo-β-1,4-xylanase from *T. xylanilyticus* (Tx-Xyl) was expressed using *E. coli* JM109(DE3) cells bearing a previously prepared plasmid construction, pECXYL-R2 (Paës and O'Donohue, 2006). Purification of Tx-Xyl was performed as Paës *et al*. described (2006). Xylanase activity was determined by monitoring the hydrolysis of birchwood xylan (5 g l^{-1}) at 60°C using the DNS method (Miller, 1959). One unit of xylanase activity (1 IU) was defined as the quantity of xylanase needed to release 1 µmol of equivalent xylose per minute.

I.2.4 Enzymatic hydrolysis of wheat straw powder

Wheat straw (0.6 g in 30 ml 50 mM sodium acetate, pH 5.8) was treated with different xylanase loadings (90 IU, 120 IU and 300 IU) at 60°C in a double jacketed, thermostated reactor with and without stirring. At regular intervals, samples were removed and centrifuged (13 200 rpm, 2 min) and the supernatant was used in a DNS assay to measure the amount of released sugars. In each case, data were derived from triplicate reactions run

in parallel. These were completed by a suitable control performed without the addition of enzyme.

I.2.5 Error-prone PCR and construction of mutant library

Random mutagenesis of the Tx-Xyl encoding DNA was performed using error-prone PCR according to an established method (Cirino et al., 2003). The final PCR product, obtained after 30 cycles of amplification, was purified using QIAquick PCR Purification Kit (Qiagen, Germany). Following digestion using *Eco*RI and *Nde*I restriction enzymes, the amplified fragment was inserted into a similarly digested pRSETa vector. The plasmid library was transformed into chemically competent *E. coli* Novablue(DE3) cells (Novagen, Germany) that were spread onto LB-agar supplemented with ampicillin (100µg/ml) and tetracyclin (12.5ug/ml) in Q-trays (Corning, USA) and incubated at 37°C, 16 hours.

I.2.6 Screening method procedure

I.2.6.1 Robotic handling

All cultivation medium dispensations into microtiter plates were performed by a Biomek® 2000 Laboratory Automation Workstation (Beckman, USA) in sterile conditions. Colony picking and plate replication were performed with a QPix2 (Genetix, UK). A Genesis RSP-200 station (TECAN, Switzerland) was used for non-sterile liquid handling and a TECAN Sunrise microplate reader was employed for spectrophotometric absorbance measurement.

I.2.6.2 Distribution of wheat straw

A transfer device (40 µl capacity Multiscreen® column loader, Millipore Inc., USA) compatible with 96-well microtiter plate format was employed to deliver a fixed amount of wheat straw to all individual wells in polypropylene microtiter plates (**Figure I-1**). The good reproducibility of this method was verified by comparing the weights of twenty microtiter plates before and after the addition of wheat straw powder.

I.2.6.3 Cell growth and expression of xylanase in microtiter plates

For the heterogeneous expression of Tx-Xyl in microplates, the plasmid pECXYL-R2 was transformed into *E. coli* Novablue(DE3). Individual colonies bearing either of wild-type or potential mutant plasmids (from directed evolution) growing on LB-agar plates were picked and automatically arrayed in sterile microtiter plates containing 200 µl LB medium

supplemented with ampicilin (100 µg/ml) and tetracycline (12.5 µg/ml). The microtiter plates were sealed with air permeable membranes and incubated at 30°C for 20-24 hours with shaking (200 rpm). After, the cultures were used to inoculate sterile medium (containing antibiotics and IPTG at 0.01 mM) in microtiter plates for enzyme expression. These fresh cultures were incubated for 16 hours at 30°C with shaking (200 rpm). After growth, cells were lysed for 1 hour at 37°C using 20 µl of lysozyme solution (5 g/l) per well. Bacterial lysis was completed using a freezing/thaw cycle (-80°C followed by rapid heating at 37°C). The resulting cell extracts were then used to perform micro-DNS assays using straw powder as the substrate.
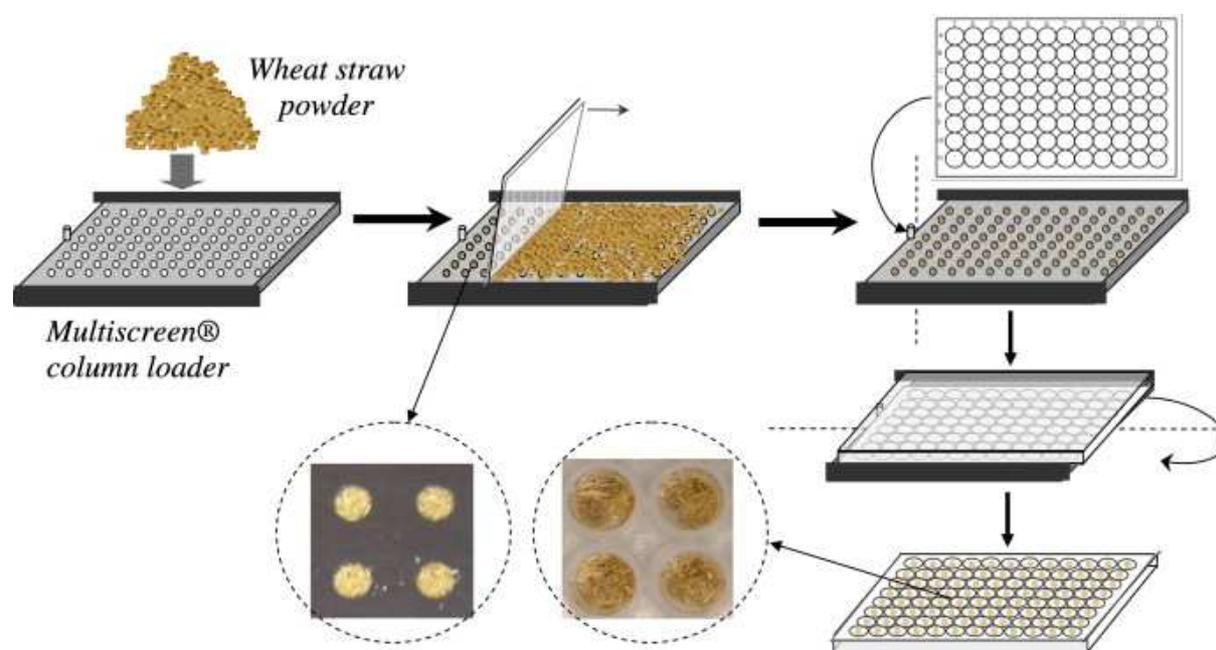
Figure I-1. Distribution of wheat straw powder into microtiter plate via the device of Multiscreen® Column Loader. Firstly, the powder is poured into the wells of device and the excess is removed from the surface, then an empty microtiter plate is covered onto the straw-accommodate device, only through a simple reversion, the straw powder will fall off into each well of microtiter plate.

I.2.6.4 Wheat-straw hydrolysis in microtiter plates

Sodium acetate buffer (50 mM, pH 5.8) and either purified Tx-Xyl xylanase (1 IU and 2 IU per well) or the previously prepared cell extracts were dispensed into individual wells of microtiter plates (96-well format) in a 1:1 ratio with a final volume of 250 µl. The microtiter plates were then sealed with an aluminium film and incubated at 60°C for 4 hours in a dry

oven. After incubation, plates were centrifuged (3700 rpm, 30 min) at 4°C to stop hydrolysis and to pellet residual straw solids. 50 μl of the supernatant were dispensed into a 96-well PCR microtiter plate and mixed with 50 μl of DNS reagent. To allow colour development, the PCR microtiter plate was incubated without sealing on a PCR-bottom heating block (CPAC Ultraflat, Watlow Inc., Germany) at 95°C for 10 min, followed with cooling in an ice bath. Finally, samples were transferred into polystyrene microplates and read absorbance at 540 nm. Generally, experiments were performed in triplicate unless otherwise stated.

I.2.7 Statistical analyses and software

To analyze the errors on micro-DNS measurements in order to evaluate the precision of the screening tests, the mean value (μ), standard deviation (σ) and coefficient of variation (CV = $\sigma/\mu \times 100\%$) were calculated. Analysis of variance was performed to compare the tests using two different concentrations of Tx-Xyl (1 and 2 IU per well). Resulting *p*-value was provided using the R software. When the *p*-value was below 5%, the comparison was considered as statistically different. The other calculations and all statistical figures were achieved using the SigmaPlot software (Systat Software Inc., USA).

I.3 Results and discussion

Previous studies have indicated that Tx-Xyl can solubilise arabinoxylans contained within complex substrates such as wheat bran and straw (Lequart et al., 1999; Zilliox and Debeire, 1998), although the reaction is always incomplete. Taking into account the thermostability of Tx-Xyl, this enzyme can be qualified as a potentially useful enzyme for biorefining applications that specifically target the extraction of arabinoxylans. Therefore, it is pertinent to develop a strategy for the optimization of the catalytic properties of Tx-Xyl, in particular its activity on lignocellulosic substrates. However, because the action of enzymes on plant cell walls is limited by a large number of often ill-defined factors, a random approach to the engineering of Tx-Xyl is preferable. A prerequisite for this strategy is a powerful, pertinent and high-throughput screening procedure that can identify promising variants. To satisfy these criteria, we have developed a microtiter plate-based assay that uses powdered wheat straw as the substrate and recombinant Tx-Xyl-containing *E. coli* extracts as the enzyme inoculums.

I.3.1 Establishment of the basic conditions for a microtiter plate assay

Before developing the microtiter plate-based assay, it was first necessary to ascertain (i) whether agitation is a requisite to achieve enzyme-mediated hydrolysis and (ii) to what extent hydrolysis is proportional to enzyme quantity/activity present in the reaction. To answer these two questions, powdered wheat straw was incubated with three different amounts of enzyme in both stirred and unstirred reactions. The results (**Figure I-2**) indicate that stirring does not affect the course of the reaction, which reached completion after an extended time period. However, in the initial phase (up to 8 h), the reaction rate appeared to be approximately proportional to the number of units of enzyme deployed. Therefore, these preliminary data confirm that microtiter plate agitation is unnecessary and indicate that a 4-h incubation time is suitable for the detection of variants that accelerate or improve the hydrolysis of wheat straw.
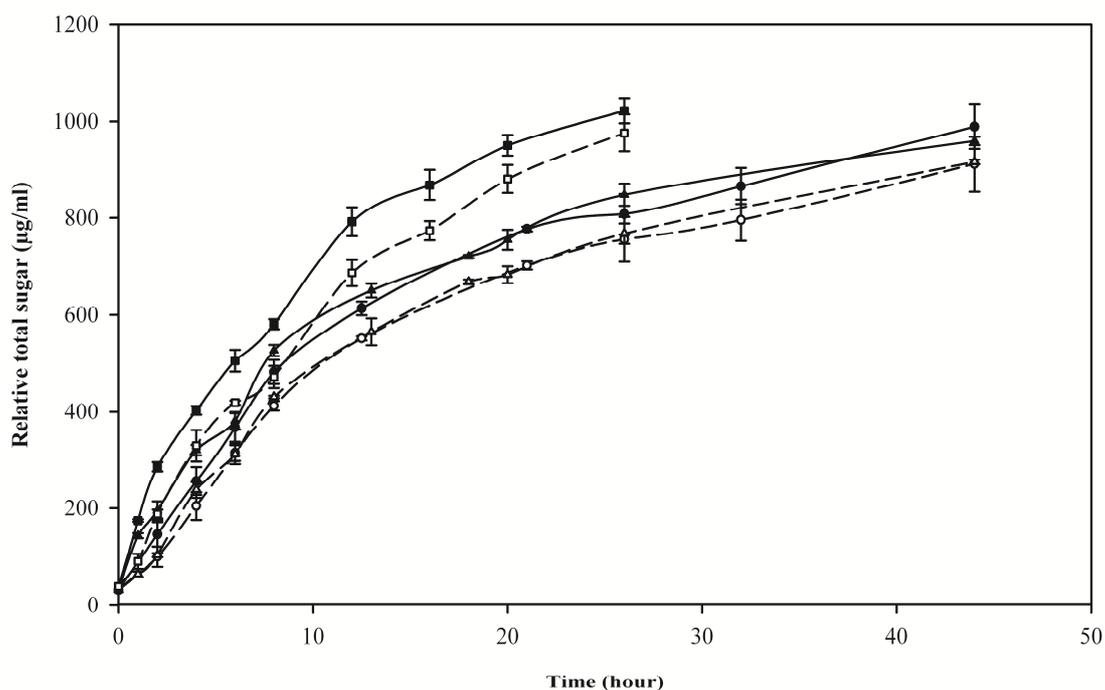
Figure I-2. Tx-Xyl mediated hydrolysis of wheat straw in 30 ml reactor. Solid lines and black symbols represent reactions performed with stirring, whereas dashed lines and open symbols are reactions performed without stirring. The symbols ■ and □, ▲ and ▲, ● and ○, represent enzyme loadings of 500, 200 and 150 IU/ g respectively.

I.3.2 Preparation of the wheat straw and its distribution in microtiter plates

To provide wheat straw in an appropriate form for distribution in microtiter plates, it was ground into a fine powder (0.5 mm). In addition, it is known that size reduction of biomass can be an effective physical pretreatment, which improves enzymatic hydrolysis through increased accessible surface area and lowered cellulose crystallinity (Chundawat et al., 2008; Sánchez and Cardona, 2008). In our experiment, the major consequence of grinding was a significant release of water soluble pigments that had to be removed by washing in pure water, in order to avoid problems with subsequent spectrophotometric measurements. Afterwards, the dried powder was sterilized in order to avoid artifacts linked to the undesirable development of microorganisms.

The homogeneous distribution of powdered biomass into microtiter plates is quite a challenge, because the chosen method needs to be both rapid and reliable. To overcome this challenge, Chundawat *et al* (2008) prepared a slurry of finely milled (<100 µm) corn stover, which was transferred to microplates using a pipettor workstation equipped with hand-cut tips. In our case, using a coarser wheat straw powder, we found that this method was both inconvenient and inconsistent. Therefore, we adopted another method that involves the use of a Multi-screen® Column Loader. This device allows the delivery of dry beads, powders, or resins into the 96-well microtiter plates. Its use is very simple, rapid and, according to our verifications, reliable. The careful weighing of 25 microtiter plates before and after delivery of wheat powder revealed that the average weight of biomass delivered to plates was 424.32±9.61 mg and the coefficient of variation was 2.27%.

I.3.3 Establishment and refinement of screening procedures

The screening approach basically allows xylanase-mediated hydrolysis of wheat straw to be performed in microtiter plates, using the DNS method to monitor the progress of the reaction. The DNS method is among the most rapid and economic assays for measuring reducing sugars and is characterized by a wide detection range (20-2500 µg/ml) and low interference from proteins (Zhang et al., 2006). The use of DNS assay in 96-well microtiter plate format has been previously described (King et al., 2009; Miyazaki et al., 2006). In our micro-DNS assay, a 100 µl final volume was constituted in individual wells of a PCR

microplate, and a 96-well format thermocycler was used for the heating step, this avoids the obvious disadvantages of a water bath. The accuracy of the micro-DNS measurement was verified using a series of standard xylose solutions (data not shown). Moreover, when the hydrolysis of wheat straw by Tx-Xyl was monitored using the micro-DNS assay, it was possible to correlate the colorimetric measurement with enzyme activity. The optimal procedure was considered to be sufficiently reproducible, with a test using a pure xylanase solution producing a low average error (average of CV = 10.91%, quadruplicate repeats).
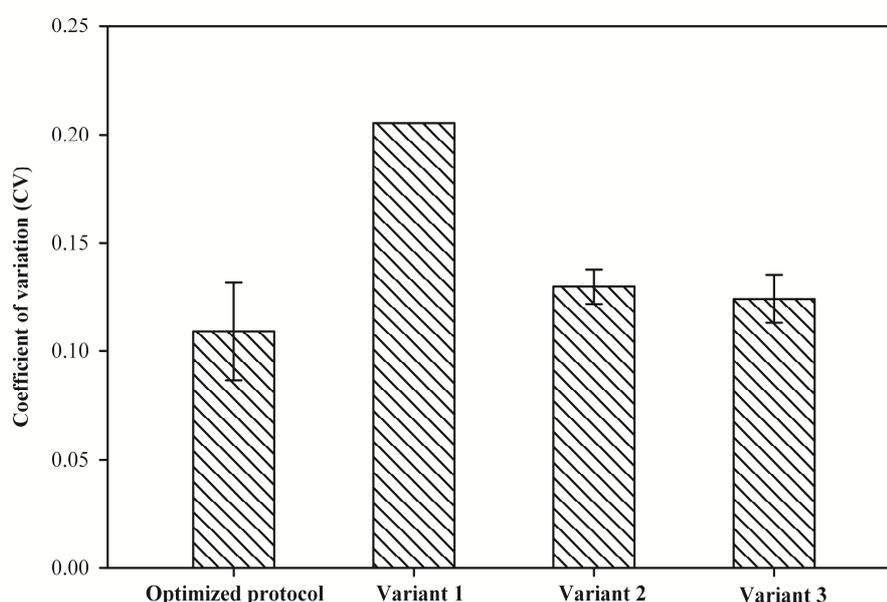
Figure I-3. Statistical analysis of various steps in the HTS protocol. Shown from left to right are: the optimized protocol, the protocol with no microtiter plate sealing (lid only, variant 1), the protocol with the use of a heating block during hydrolysis (variant 2) and the protocol with the use of a water bath for the micro-DNS assay (variant 3).

Various options for microplate sealing (lid or aluminium film) and incubation (heating block or oven) were investigated in order to choose the best overall configuration, which is characterized by low error rate and time-saving advantages. For the evaluation of each protocol, a hydrolysis test was performed using a solution of pure Tx-Xyl and the CV of the final micro-DNS values was calculated (**Figure I-3**). Although sealing with aluminium film constitutes a limiting operation, it avoided the intolerable level of evaporation associated with the use of simple microtiter plate lids. Using aluminium film, the percentage weight loss due to evaporation (in 4-h incubation) was only 1.44 ± 0.16% (average value for 20 replicate

microplates). Advantageously, because stirring does not enhance hydrolysis, it was possible to incubate the microtiter plates in an oven during the hydrolysis reaction. Although, this method does not procure a major gain in accuracy, it does allow the simultaneous processing of a large number of microtiter plates. The key steps in the final optimized protocol are summarized in **Figure I-4**.
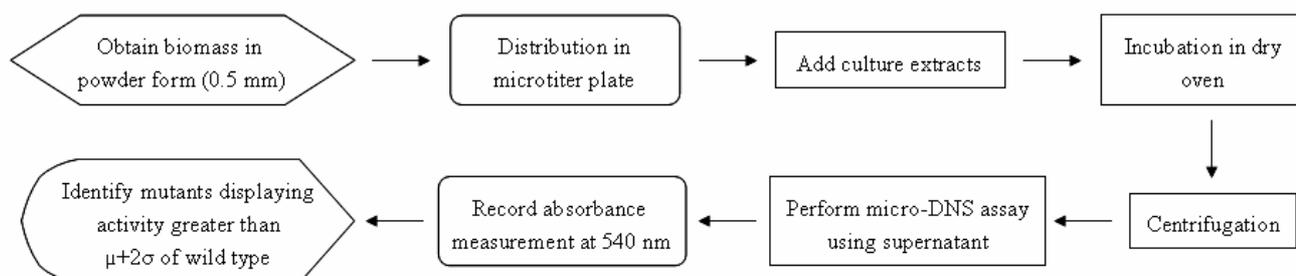
Figure I-4. Flow chart of the HTS protocol.
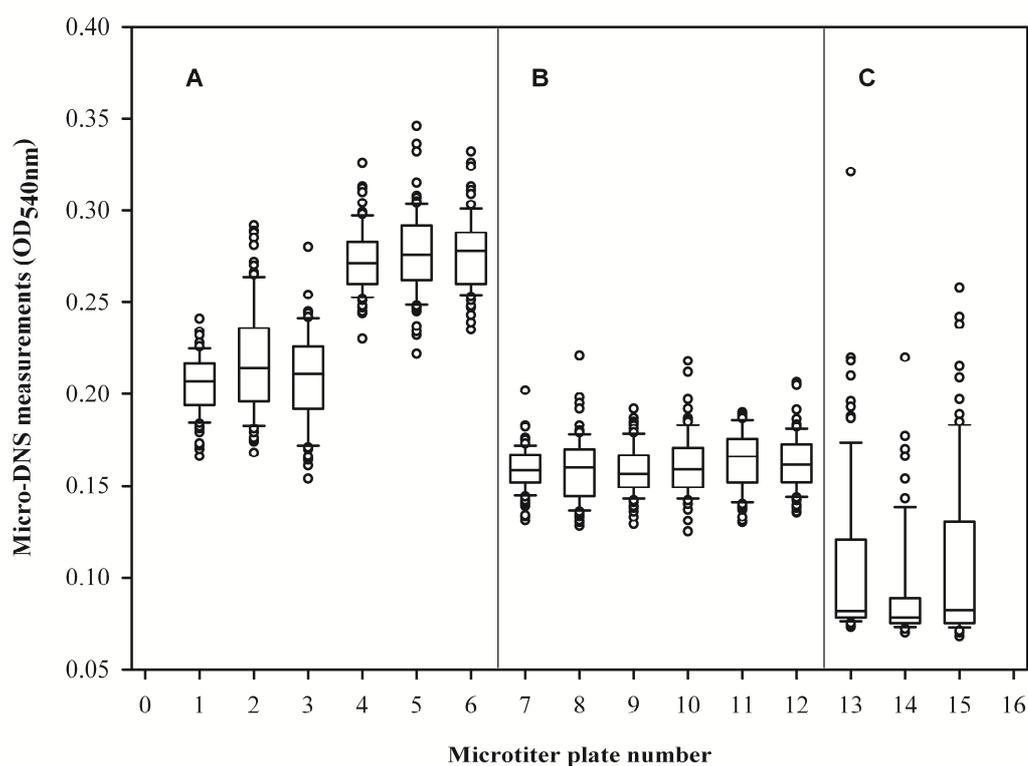
Figure I-5. Box plots of micro-DNS assays performed on reactions involving (A) two different loadings (1 and 2 IU/well) of pure Tx-Xyl (triplicate data); (B) Tx-Xyl-containing cell extracts produced in microtiter plate format (two sets of triplicates) and (C) Cell extracts of xylanase variants from the error-prone PCR library (three microtiter plates). Whiskers extend from 10th to 90th percentiles and outliers contain all data.

I.3.4 Evaluation of the micotiter plate assay using pure Tx-Xyl preparation

Because the amount of straw powder dispensed to each well cannot be directly determined, an indirect measurement was performed by hydrolyzing the straw in individual wells with an identical amount of pure Tx-Xyl enzyme solution. Likewise, using two different concentrations of enzyme (1 and 2 IU per well) it was found that the coefficient of variation in the DNS measurements for individual wells of a microtiter plate was less than 11% ($\mu\pm\sigma$ = 0.211$\pm$0.025), for tests involving 1 IU/well of Tx-Xyl and approximately 7% ($\mu\pm\sigma$ = 0.276$\pm$0.020) in tests using 2 IU/well. The 2-fold increase in enzyme activity which can be clearly seen on **Figure I-5**.A is statistically significant (*p*-value from the analysis of variance < 10^{-15}). Overall, these results show that the straw distribution in the microtiter plate is sufficiently homogeneous and that the assay is sensitive enough to reproducibly reveal 2-fold increases in enzyme activity.

I.3.5 Optimization for microtiter plate-based culture extracts

This high throughput method is intended to screen xylanase activity produced by *E. coli* cells bearing an expression vector. Therefore it was necessary to establish the optimal parameters for the production of Tx-Xyl and the analysis of culture supernatants in microtiter plate format. First, various *E. coli* strains such as JM109(DE3), BL21(DE3), Rosetta(DE3) and Novablue(DE3), expression vectors (pRSETa and pET21a) and culture conditions (liquid media and IPTG concentration) were evaluated in order to pinpoint the best combination for the production of recombinant Tx-Xyl. Only JM109(DE3) and Novablue(DE3), which are *E. coli* K12 derivatives, grown in LB medium produced stable growth in microtiter plates and low background noise in the micro-DNS assay (data not shown). The vector pET21a proved to be unsuitable, so the preferred combination was Novablue(DE3) bearing a pRSETa-based plasmid. The major advantage of Novablue(DE3) over JM109(DE3) was the fact that the former is commercialized in the form of chemically competent cells, which are useful when creating mutant libraries.

The growth of *E. coli* cells at 30°C in microtiter plates was subject to evaporation, with a final percentage weight loss of 5.94±1.54% (cumulative results from 15 microtiter plates). The

average amount of xylanase activity produced by the microtiter plate-based cultures was 4.83±0.34 IU/ml (hydrolysis measurements from 16 microtiter plates using birchwood xylan as substrate) and maximal xylanase production was achieved by cultures displaying an OD_{600nm} of 0.4. Importantly, quite heterogeneous culture densities (CV of OD_{600nm} = 19.05%) did not produce heterogeneous activity measurements (CV of OD_{540nm} = 8.14%), indicating that the method is quite robust with regard to the uneven growth of bacteria in individual wells.

Unlike the tests carried out using pure xylanase solution, the use of culture extracts was a source of further challenges linked to the viscosity and the complexity of the reaction mixture. To ensure good reproducibility in these conditions and to avoid the carry-over of straw powder into the micro-DNS assay, it was found to be important to ensure that (i) the reaction was sufficiently diluted by adding an equivalent volume of buffer solution to the culture extract, and (ii) straw powder was completely eliminated before the micro-DNS assay, by applying an adequate centrifugation step after hydrolysis. When these conditions were satisfied, a negative control (i.e. performed using an extract of *E. coli* Novablue(DE3)/pRSETa) produced a DNS measurement of 0.079±0.004, which was similar to a blank control (0.085±0.003), performed using buffer only. Additionally, using a prolonged incubation time (16 h), it was possible to show that the course of a reaction catalyzed by a xylanase-containing culture extract, prepared and assayed in microtiter plate format, was highly similar to that of a reaction catalyzed by a pure xylanase solution and performed at a 30 ml scale (data not shown).

With regard to the reproducibility of the screening method performed in microtiter plate format, results from six microtiter plates revealed that mean activity measurements (expressed as OD_{540nm}) were in the range 0.158 to 0.164 (**Figure I-5**.B) and that the deviation was below 10%. Only 3% of measurements fell outside the limits of $\mu+2\sigma$, in accordance with a normal distribution (5% expected for a normal distribution). Overall, the small variation between individual wells, corresponding to individual colonies, confirmed that the method is suitable for the screening of clonal libraries.

I.3.6 Implementation of the HTS method on an error-prone PCR clone library

To provide an ultimate validation of the HTS method, a 264-mutant subset of an error-prone PCR clone library was submitted to screening. Three microtiter plates were prepared, each with 88 colonies from the mutant library and eight clones encoding wild-type Tx-Xyl. These latter were placed in column 12 of each microplate and were included as an internal control. The μ±σ values of the wild-type controls in the three microplates were 0.177±0.017, 0.168±0.020 and 0.164±0.013 respectively (**Figure I-5**.C). This reveals variability between microtiter plates, but variability between individual control wells within the same microtiter plate was low. Among the clones from the mutant library, those displaying an activity greater than μ+2σ of wild-type activity were selected for further analysis. In this way, one mutant designated Tx-Xyl-AF7 was found to display 74.5% greater activity than the wild-type controls. DNA sequencing revealed two silent mutations (T27C and C516T) in the Tx-Xyl-AF7 coding sequence, which can be correlated with increased protein expression. Indeed, production of Tx-Xyl using the Tx-Xyl-AF7 clone in 50 ml flasks (performed in quadruplicate) revealed an average two-fold increase in the number IU per ml culture compared to a parental clone grown in the same conditions (**Figure I-6**.A). Further analysis revealed that the mutation T27C is located in the 5'-region of the Tx-Xyl coding sequence, near to the ribosome binding site, which is thought to be a key zone with regard to protein expression levels (Kudla et al., 2009; Seo et al., 2009). Using an mRNA structure prediction webserver (http://rna.tbi.univie.ac.at/) (Gruber et al., 2008; Hofacker et al., 1994), it was possible to reveal the presence of a putative hairpin structure in this region and show that this structure is modified by the mutation T27C (**Figure I-6**.B and C). According to this model, the single base alteration causes a local reorganisation of hydrogen bonding, which correlates to a drop in the minimum free energy (-8.0 to -2.5 kcal/mol) of the hairpin. Therefore, we conclude that the increased protein expression displayed by the clone Tx-Xyl-AF7 is due a modification of the mRNA secondary structure, which produces a favourable effect on the initial rate ribosome-mediated translation rate.
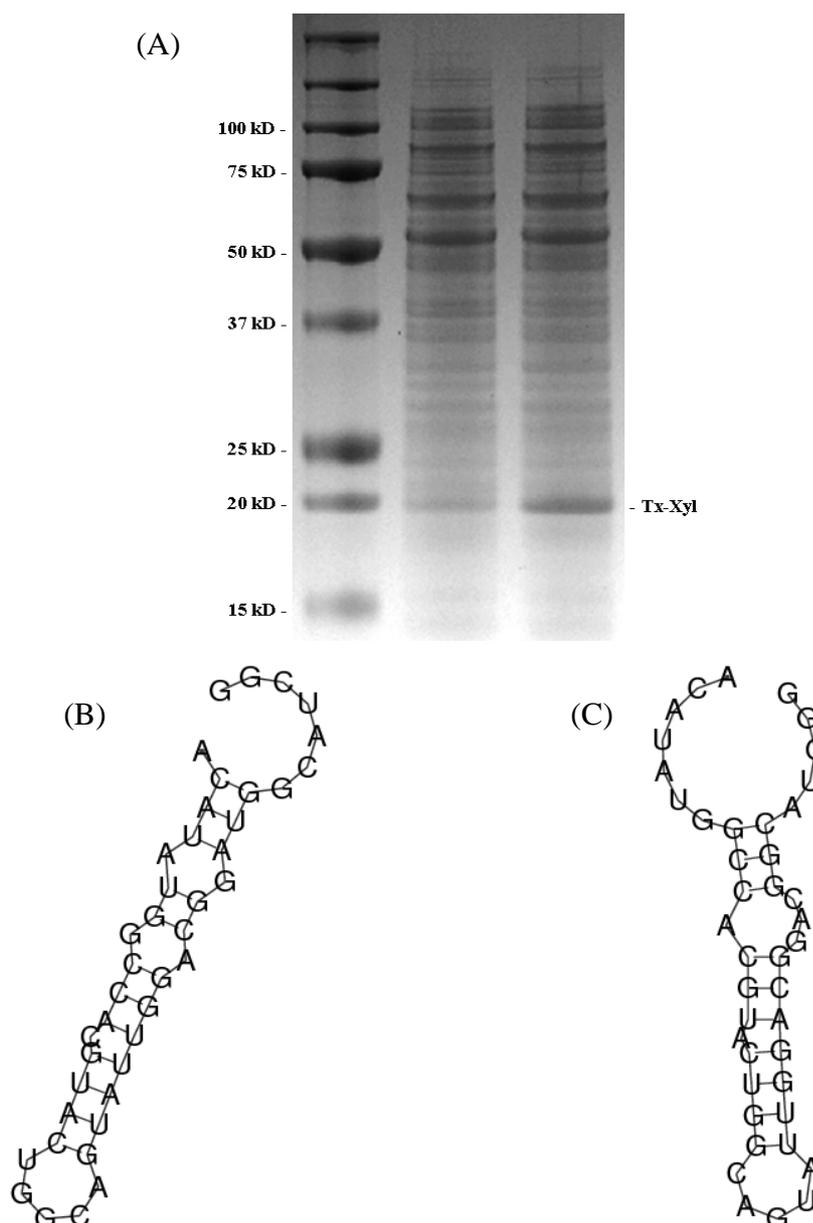
Figure I-6. (A) SDS-PAGE analysis of Tx-Xyl expression from mutant Tx-Xyl-AF7 and wild-type coding sequences: lane 1, *Mw* marker; lane 2, wild-type; lane 3, mutant Tx-Xyl-AF7. **(B)** and **(C)** predicted structures of the 5'-end of Tx-Xyl-encoding mRNA (position -4 - +37) displaying either the original sequence **(B)** or that of the mutant Tx-Xyl-AF7 **(C)**.

I.4 Conclusion

In this study, we have developed a HTS method that can be applied to the screening of glycoside hydrolase activity on lignocellulosic biomass. Because the biomass only needs to be milled to coarse powder form, no specialized micronization equipment is required.

Likewise, because the method employs a very simple and reliable method to distribute the powdered biomass into microtiter plates, the purchase of an automatic powder distributor is avoided. The whole HTS process has been carefully conceived to provide robust technology that can be applied to mutant enzyme libraries, with a screening rate of approximately 2000 clones per week when using 96-well format microtiter plates in a TECAN liquid handling workstation. Working with a xylanase and wheat straw as the substrate, we have demonstrated the reproducibility and robustness of the HTS method. It has been possible to show that any mutants that display more than 15% increase in activity on wheat straw can be detected. In this work, activity measurements have been performed at an early stage in the hydrolytic reaction. Therefore, it has been possible to show that the HTS method can even detect mutants that simply increase protein expression and therefore accelerate the initial rate of the reaction. To detect mutant xylanases that actually solubilise more hemicelluloses than the parental enzyme (i.e. those that go beyond the normal endpoint) a modification of the method is required. This involves either an increase in the incubation time, in order to attain the end-point of the reaction, or the use of depleted wheat straw as the substrate, which has been first treated with the parental xylanase. In recent work, we have tested this latter option and have shown that it constitutes a suitable way to identify high-performance xylanases. The creation and characteristics of these enzymes will be reported in a future communication.

I.5 References

Chundawat, S.P., Balan, V., Dale, B.E., 2008. High-throughput microplate technique for enzymatic hydrolysis of lignocellulosic biomass. *Biotechnol Bioeng*, 99, 1281-94.

Cirino, P.C., Mayer, K.M., Umeno, D., 2003. Generating Mutant Libraries Using Error-Prone PCR. in: F.H. Arnold, G. Georgiou (Eds.), Directed Evolution Library Creation. Humana Press, pp. 3-9.

Demain, A.L., 2007. REVIEWS: The business of biotechnology. *Industrial Biotechnology*, 3, 269-283.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R., Hofacker, I.L., 2008. The Vienna RNA Websuite. *Nucl. Acids Res.*, gkn188.

Himmel, M.E., Ding, S.-Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W., Foust, T.D., 2007. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science*, 315, 804-807.

Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., Schuster, P., 1994. Fast Folding And Comparison Of Rna Secondary Structures. *Monatshefte Fur Chemie*, 125, 167-188.

King, B.C., Donnelly, M.K., Bergstrom, G.C., Walker, L.P., Gibson, D.M., 2009. An optimized microplate assay system for quantitative evaluation of plant cell wall-degrading enzyme activity of fungal culture extracts. *Biotechnol Bioeng*, 102, 1033-44.

Kudla, G., Murray, A.W., Tollervey, D., Plotkin, J.B., 2009. Coding-Sequence Determinants of Gene Expression in Escherichia coli. *Science*, 324, 255-258.

Lequart, C., Nuzillard, J.-M., Kurek, B., Debeire, P., 1999. Hydrolysis of wheat bran and straw by an endoxylanase: production and structural characterization of cinnamoyl-oligosaccharides. *Carbohydrate Research*, 319, 102.

Lin, L., Meng, X., Liu, P., Hong, Y., Wu, G., Huang, X., Li, C., Dong, J., Xiao, L., Liu, Z., 2009. Improved catalytic efficiency of Endo-β-1,4-glucanase from Bacillus subtilis BME-15 by directed evolution. *Applied Microbiology and Biotechnology*, 82, 671.

Lynd, L.R., Zyl, W.H.v., McBride, J.E., Laser, M., 2005. Consolidated bioprocessing of cellulosic biomass: an update. *Current Opinion in Biotechnology*, 16, 577.

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

Miller, G.L., 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*, 31, 426.

Miyazaki, K., Takenouchi, M., Kondo, H., Noro, N., Suzuki, M., Tsuda, S., 2006. Thermal stabilization of Bacillus subtilis family-11 xylanase by directed evolution. *J Biol Chem*, 281, 10236-42.

Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y.Y., Holtzapple, M., Ladisch, M., 2005. Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology*, 96, 673-686.

Nakazawa, H., Okada, K., Onodera, T., Ogasawara, W., Okada, H., Morikawa, Y., 2009. Directed evolution of endoglucanase III (Cel12A) from Trichoderma reesei. *Applied Microbiology and Biotechnology*, 83, 649.

Paës, G., O'Donohue, M.J., 2006. Engineering increased thermostability in the thermostable GH-11 xylanase from Thermobacillus xylanilyticus. *J Biotechnol*, 125, 338-50.

Sánchez, Ó.J., Cardona, C.A., 2008. Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresource Technology*, 99, 5270-5295.

Schmidt-Dannert, C., Arnold, F.H., 1999. Directed evolution of industrial enzymes. *Trends in Biotechnology*, 17, 135.

Seo, S.W., Yang, J., Jung, G.Y., 2009. Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and translational efficiency in Escherichia coli. *Biotechnol Bioeng*, 104, 611-6.

Shibuya, H., Kaneko, S., Hayashi, K., 2000. Enhancement of the thermostability and hydrolytic activity of xylanase by random gene shuffling. *Biochem J*, 349, 651-6.

Stephens, D.E., Rumbold, K., Permaul, K., Prior, B.A., Singh, S., 2007. Directed evolution of the thermostable xylanase from Thermomyces lanuginosus. *Journal of Biotechnology*, 127, 348-354.

Wyman, C.E., 2007. What is (and is not) vital to advancing cellulosic ethanol. *Trends in Biotechnology*, 25, 153.

Zhang, P.Y.H., Himmel, M.E., Mielenz, J.R., 2006. Outlook for cellulase improvement: Screening and selection strategies. *Biotechnology Advances*, 24, 452-481.

Zilliox, C., Debeire, P., 1998. Hydrolysis of wheat straw by a thermostable endoxylanase: Adsorption and kinetic studies. *Enzyme and Microbial Technology*, 22, 58.

ARTICLE II

ENGINEERING BETTER BIOMASS-DEGRADING ABILITY INTO A GH11 XYLANASE USING A DIRECTED EVOLUTION STRATEGY

In this article, the process of random engineering of Tx-Xyl by error-prone PCR, DNA shuffling and site-saturation mutagenesis, in addition to the statistical analyses of screening on wheat straw are described in detail. For the isolated best performing mutants, the enzyme properties of kinetics, thermostability, melting temperature, etc. are characterized respectively. In addition, the hydrolytic performance and synergistic actions with cellulases for the mutant enzymes are studied in the hydrolysis assays of wheat straw. Finally, we develop the insights gained from this study for the understanding of xylanase structure-function effects on complex biomass.

Engineering better biomass-degrading ability into a GH11 xylanase using a directed evolution strategy

Letian Song[a,b,c], Sophie Bozonnet[a,b,c], Béatrice Siguier[a,b,c], Claire Dumon[a,b,c], Michael J. O'Donohue[a,b,c,*]

[a] *Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France*

[b] *INRA, UMR792, F-31400 Toulouse, France*

[c] *CNRS, UMR5504, F-31400 Toulouse, France*

Abstract

Improving the hydrolytic performance of hemicellulases on lignocellulosic biomass is of considerable importance for second generation biorefining. To address this problem, and also to gain greater understanding of structure-function relationships, especially related to xylanase action on complex biomass, we have implemented a combinatorial strategy to engineer the GH11 xylanase (Tx-Xyl) from *Thermobacillus xylanilyticus*. After *in vitro* enzyme evolution and screening on wheat straw, nine best-performing clones were identified, which display mutations at positions 3 (Y3W), 6 (Y6H), 27 (S27T) and 111 (Y111H, Y111S, Y111T). All of these mutants showed increased hydrolytic activity on wheat straw, and solubilised arabinoxylans that were resistant to the parental enzyme. The most active mutants, S27T and Y111T, increased the solubilisation of arabinoxylans 2.3-fold and 2.1-fold respectively, in comparison to the wild type enzyme. In addition, five multiple mutants (all containing mutations at position 27 and/or 111) increased total hemicellulose conversion of intact wheat straw from $16.7\%_{tot.\ xyl}$ (wild-type Tx-Xyl) to $18.6 - 20.4\%_{tot.\ xyl}$. Also, all five mutant enzymes exhibited a better ability to act in synergy with a cellulase cocktail (Accellerase 1500), thus procuring increases in overall wheat straw hydrolysis. Analysis of the results in the light of state of the art knowledge has allowed us to hypothesize that the increased hydrolytic ability of the mutants is linked to i) improved ligand binding in a putative secondary binding site caused by mutation S27T, ii) the diminution of surface hydrophobicity, resulting from mutations at positions 6 and 111, and iii) the modification of thumb flexibility, induced by mutation at position 111. However, the relatively modest improvements that were observed also reveal that enzyme engineering alone cannot overcome the limits imposed by the complex organisation of the plant cell wall and the lignin barrier.

Keywords

Abbreviations

Tx-Xyl: *Thermobacillus xylanilyticus* xylanase; DW: dry weight; In-WS: intact (untreated) wheat straw; Dpl-WS: xylanase-depleted wheat straw; epPCR: error-prone PCR; StEP: staggered extension process; CV: coefficient variation; μ: mean value; σ: standard deviation; BWX: birchwood xylan; LVWAX: low viscosity wheat arabinoxylan; AX: arabinoxylan; SR: ratio of the specificity constants (BWX/LVWAX).

II.1 Introduction

Wheat straw is an abundant co-product of the agri-food industry that is currently considered to be a primary source of lignocellulosic biomass for 2nd generation biorefining. The composition of wheat straw is typical of graminaceous species, containing arabinoxylan (20-25% dry weight or DW), cellulose (35-45% DW) and lignins (15-20% DW) in variable proportions that are determined by both cultivar characteristics and pedo-climatic differences (Fraser-Reid et al. 2008; Scheller and Ulvskov 2010). Regarding the ultrastructure of wheat straw, the internode regions, which in DW terms represent the majority of wheat straw, are characterized by different tissue types, which notably display different levels of lignifications. The central cavity, or lumen, of straw is lined by pith that covers parenchyma cells that possess mainly primary cell walls. Moving further outwards to the external part of wheat straw, one can identify sclerenchyma cells, xylem tissue and finally the outer epidermis, all of which possess lignified secondary cell walls (Hansen et al. 2010; Hornsby et al. 1997a).

Endo-β-1,4-xylanases (EC 3.2.1.8, xylanase) randomly depolymerize the backbone of β-1,4-linked xylans (Berrin and Juge 2008; Zimmermann 1991), including arabinoxylans such as those found in wheat straw. Current commercial uses for xylanases mainly focus on the paper, food and animal feed industries (Collins et al. 2005; Kulkarni et al. 1999), but it is increasingly recognized that these will also be important for biorefining of lignocellulosic biomass (Shallom and Shoham 2003; Wyman 2007). Indeed, recent studies have shown that xylanases are needed in cellulase cocktails in order to alleviate the inhibition of various cellulose-degrading enzymes by xylo-oligosaccharides (Kumar and Wyman 2009b). Also, the development of ambitious approaches such consolidated bioprocesses (Lynd et al. 2002; Lynd et al. 2005), which require the use of microorganisms possessing the dual ability to degrade complex biomass and convert the fermentable sugars into useful products, will also create new demands for highly efficient xylanolytic systems, including xylanases.

So far, most industrial processes that employ xylanases use enzymes that belong to the glycoside hydrolase family, GH11 (CAZy classification). These xylanases are mostly single

domain enzymes that exclusively act on β-1,4-links between xylosyl units in xylans and display a β-jelly roll structure that has been likened to a partially folded human right-hand (Figure II-1) (Törrönen et al. 1994). Likewise, the prominent elements of the 3D structure, which is composed mainly of two β-sheets and one α-helix, have been identified using terms such as 'thumb', which describes a large mobile loop that is located above the active site cleft, 'palm', whose half-folded structure forms the active site cleft, and fingers, which constitute one side of the active site cleft and whose 'knuckles' bear a secondary substrate binding motif (Ludwiczek et al. 2007; Purmonen et al. 2007).
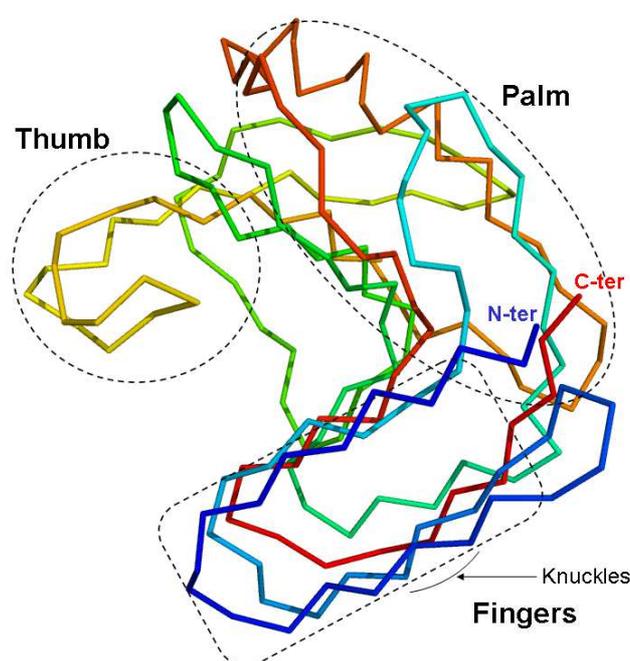
Figure II-1. Ribbon presentation of 3D structure of Tx-Xyl xylanase. The schematic protein is "colour-ramped" from the N-terminus (blue, N-ter) to the C-terminus (red, C-ter). The relevant regions of 'thumb', 'palm' and 'fingers' are highlighted in frames, and the 'knuckles' in the fingers region is indicated by an arrow.

Despite the fact that xylanases will be necessary for biorefining operations, very little R&D has so far being focused on the improvement of xylanases specifically for biorefining purposes, and in particular for increased activity on complex biomass. This is partly because a lot of effort has been focused on cellulase engineering, and also because presently it is unclear on what basis improvements could be achieved. Regarding the action of xylanases on lignocellulosic biomass that has not been subjected to prior pretreatment, very little is

known, though some studies on wheat bran and straw that focused on the GH11 xylanase from *Thermobacillus xylanilyticus* (designated Tx-Xyl) have provided insight into the factors that might determine overall enzyme efficiency. Nevertheless, the available information is still sparse, making the prospect of rational engineering rather haphazard.

Alternatively, random approaches coupled to enzyme *in vitro* evolution could be a suitable way to tackle xylanase engineering. So far, the use of such techniques on xylanases has been limited to the improvement of thermostability (Dumon et al. 2008; Miyazaki et al. 2006; Ruller et al. 2008; Stephens et al. 2007; You et al. 2010) and alkaliphilicity (Chen et al. 2001; Inami et al. 2003; Stephens et al. 2009). In these studies, screening methods relied on the use of isolated xylans, such as RBB-xylan and birchwood xylan. However, in a recent study we have developed a new microtitre plate-based screening method that is far more suitable for the study of xylanase action on complex biomass(Song et al. 2010). Therefore, in this paper, we describe the use of this screening procedure in an enzyme engineering project that has focused on the moderately thermostable Tx-Xyl. Using a combination of random mutagenesis and DNA shuffling, we have isolated several Tx-Xyl variants that showed increased activity on wheat straw and improved synergistic action, when used in combination with a commercial cellulase preparation.

II.2 Materials and methods

II.2.1 General materials and regents

Unless otherwise stated, all chemicals were of analytical grade and purchased from Sigma-Aldrich (St. Louis, MO, USA). The T7-promoter based vector pRSETa was purchased from Invitrogen (Cergy Pontoise, France), and the *Escherichia coli* host strains Novablue(DE3) and JM109(DE3) were obtained from Stratagene (La Jolla, CA, USA) and Novagene (Darmstadt, Germany) respectively. All restriction enzymes, T4 DNA ligase, *Taq* DNA polymerase and their corresponding buffers were purchased from New England Biolabs (Beverly, MA, USA). Oligonucleotide primers were synthesized by Eurogentec (Angers, France), and the DNA sequencing is conducted by GATC Inc. (Marseille, France). Sterile 96-well cell culture mictrotiter plate and sealing tapes were purchased from Corning Corp. (NY, USA), and other

analyzing microtiter plates were from Evergreen Scientific (Los Angeles, CA, USA). The low viscosity wheat flour arabinoxylan (LVWAX) was obtained from Megazyme (Wicklow, Ireland), and the birchwood xylan (BWX) was purchased from Sigma-Aldrich (St. Louis, MO, USA).

II.2.2 Mutagenesis procedure and library construction

II.2.2.1 Random mutagenesis by error-prone PCR

Random mutagenesis was carried out by error-prone PCR (epPCR) using an established protocol (Cirino et al. 2003). The adopted template was (first round only) the DNA encoding Tx-Xyl (Swiss-Prot accession number Q14RS0, bearing the substitution N1A) or (in subsequent rounds) Tx-Xyl-AF7 described by Song *et al* (2010). Briefly, the PCR reaction mixture (50 µl) contained 5 ng of template DNA, 0.3 µM of primers epF and epR (see below), 0.2 mM dGTP/ATP (equimolar mixture) and 1 mM dCTP/TTP (equimolar mixture), 7 mM MgCl$_2$, 5 IU *Taq* polymerase and (in the third round of epPCR only) 0.05 mM of MnSO$_4$. Reactions were conducted using the following sequence: 1 cycle at 94°C for 2 min, 30 cycles at 94°C for 1 min, 1 cycle at 42°C for 1 min and 1 cycle at 72°C for 1 min, and finally 1 cycle at 72°C for 5 min. The amplicons were purified using QIAquick PCR Purification Kit (Qiagene, Courtaboeuf, France) and were digested with *EcoR*I and *Nde*I and inserted into a similarly digested pRSETa vector. The ligation mixture was used to transform competent *E.coli* Novablue(DE3) cells.

　　epF: 5'- GGAGATATACAT<u>ATG</u>GCCACG -3'

　　epR: 5'- GGATCAAGCTTCGAATTCTTACC -3'

II.2.2.2 DNA recombination by staggered extension process (StEP)

DNA recombination was carried out using an adapted StEP method (Zhao et al. 1998; Zhao and Zha 2006). The PCR reaction (50 µl) contained 5 ng of total template DNA (equimolar mixture of each parental gene), 0.3 µM of each primer, 0.2 mM of each dNTP, and 5 IU *Taq* polymerase. Reactions were conducted using the following sequence: 1 cycle at 94°C for 2 min; 40 cycles comprising a step at 94°C for 30 s and 1 step at 58°C for 2 s; followed by 40 cycles with 1 step at 94°C for 30 s and 1 step at 56°C for 2 s. Afterwards 20 IU of *Dpn*I was added to PCR reaction, which was incubated at 37°C for 1 hour, before amplicon purification and digestion with *EcoR*I and *Nde*I. Finally, the mutant library was generated by ligating the

digested amplicons to *EcoR*I/*Nde*I-digested pRSET plasmid DNA and transforming the resultant products into competent *E.coli* Novablue(DE3) cells.

II.2.2.3 Site-saturation mutagenesis at amino acid positions 3 and 111

Saturation mutagenesis on residues Tyr3 and Tyr111 of Tx-Xyl was performed using the QuikChange mutagenesis kit (Stratagene, La Jolla, CA). The following mutagenic primers (Eurogentec) were designed using NNK degeneracy (Reetz and Carballeira 2007), according to the recommendations provided in the instruction manual (mismatched bases are underlined; N is A, G, C, or T, K is G or T and M is A or C):

For amino acid position 3:

Y3N_FW: 5'- GATATACATATGGCCACG<u>NNK</u>TGGCAGTATTGGACG -3'

Y3N_REV: 5'- CGTCCAATACTGCCA<u>MNN</u>CGTGGCCATATGTATATC -3'

For amino acid position 111:

Y111N_FW: 5'- C TATCACAGCTGGCGC<u>NNK</u>AACGCACCGTCC ATCGAC -3'

Y111N_REV: 5'- GTCGATGGACGGTGCGTT<u>MNN</u>GCGCCAGCTGTGATAG -3'

After the PCR is performed, a digestion with *Dpn*I removed template DNA, and the product was used to transform *E. coli* Novablue109(DE3) cells.

II.2.2.4 Site-directed mutagenesis

The mutational combinations W109R-Y111H, Y3H-W109R-Y111H, Y3L-W109R-Y111H, S27T, and Y6H were created through site-directed mutagenesis. This was achieved using the QuickChange site-directed mutagenesis kit, according to the manufacturer's instruction. The oligonucleotide primers employed in PCRs are listed in Table II-1S.

II.2.3 Library screening on intact and xylanase treated wheat straw

II.2.3.1 Substrate preparation

Wheat straw (*Triticum aestivum*, cv. Apache) harvested (2007) in France was milled using a blade grinder that procured a fine powder having an average particle size of 0.5 mm. After, the wheat straw powder, designated In-WS, was washed with distilled water (10 volumes), filtered using a Büchner funnel equipped with a Whatman® No.4 filter paper (pore size: 20–25 lm), dried in an oven at 45°C and then sterilized by autoclaving. To prepare xylanase-treated wheat straw (designated Dpl-WS), 20 g In-WS was suspended in 50 mM sodium

acetate buffer, pH 5.8 (containing 0.02% NaN$_3$) containing Tx-Xyl (150 BWX U g^{-1} biomass) and incubated at 60°C for 70 h. Afterwards, the reaction mixture was heated at 95°C for 5 min to inactive the enzyme. The solid residues were recovered by filtration (see above) and dried as before. The sugar composition of both wheat straw substrates was analyzed according to the procedure published by The National Renewable Energy Laboratory (NREL) (Sluiter et al. 2008). The composition is shown in Table II-1.

Table II-1. Characteristics of In-WS and Dpl-WS and short description of the two screening assays.

	In-WS screening	Dpl-WS screening
Substrate properties		
Substrate type	intact (untreated) wheat straw	xylanase-depleted wheat straw
Particle size	0.5 mm in average	
Glucose% (w/w)	44.51±0.08%	45.69±0.94%
Xylose % (w/w)	26.16±0.14%	21.92±0.17%
Arabinose% (w/w)	2.37±0.03%	2.05±0.07%
Ratio of Ara: Xyl	0.091	0.094
Screening conditions		
Weight (mg per microplate)	420-440	385-405
Cell-free extract (CFE) loading	CFE:NaOAc buffer = 1:1, 250 µl well^{-1} in total	
Temperature and time	60°C, 4h	60°C, 16h
Sealing	Aluminum film	Polypropylene film
Evaporation (w/w, %)	1.44±0.16%	0.23±0.05%
Activity assay	Micro-DNS assay	

II.2.3.2 Microtitre plate-based automatized screening

Microtitre plate-based screening of mutant libraries was performed according to the method described by Song *et al* (2010). Briefly, individual *E. coli* transformants were grown in the wells of 96-well microtitre plates and then cells were recovered and lyzed using the combined effect of lysozyme (0.5 g l^{-1}) and freeze – thaw cycling (–80°C and 37°C). The screening of xylanase activity was then achieved using a 4-step protocol, which involved (1) substrate delivery into microtitre plates (2) addition of xylanase-containing cell lysates (3) incubation and (4) measurement of solubilised reducing sugar using a micro-DNS assay. The important experimental details of these steps are summarized in Table II-1. When Dpl-WS was employed in the place of In-WS, the incubation time was extended to 16 h and, consequently, microtitre plates were thermo-sealed using polypropylene film to reduce

evaporation. In all microtitre plate screening, wells containing transformants expressing wild type Tx-Xyl were included as internal controls. These were used to calculate a coefficient of variation (1 CV = $\sigma/\mu \times 100\%$, σ is standard deviation and μ is mean value) of Tx-Xyl activity, which was employed to assess the activity of mutant variants.

II.2.4 Xylanase expression and purification

The production in *E.coli* JM109(DE3) cells and purification of Tx-Xyl and variants thereof was performed according to the previously described procedure (Paës and O'Donohue 2006). Briefly, purification followed a 2-step protocol involving ion-exchange (Q sepharose FF) and then affinity chromatography (Phenyl sepharose) operating on an ÄKTA purification system (GE Healthcare, Uppsala, Sweden). Enzyme conformity and purity were assessed using SDS-PAGE and theoretical extinction coefficients were computed using the ProtParam server (Walker et al. 2005). The concentration of xylanase solutions was determined by measuring UV absorbance at 280 nm and then applying the Lambert-Beer equation.

II.2.5 Evaluation of xylanase-mediated hydrolysis on Dpl-WS and In-WS

To measure xylanase activity using In-WS or Dpl-WS as substrates, a reaction mixture in 50 mM sodium acetate buffer, pH 5.8 was prepared that contained 2 % (w/v) biomass, 0.1 % (w/v) bovine serum albumin (BSA), 0.02 % (w/v) NaN$_3$ and an aliquot (final concentration of 10 nmol enzyme g^{-1} biomass) of Tx-Xyl or a mutant thereof. To analyze the combined effect of xylanase and cellulases on In-WS, reactions were conducted as described above, except that Accellerase 1500 (Genencor, Rochester, NY) (0.2 ml cocktail per g biomass) was added to the reaction mixture and reactions were buffered at pH 5.0. To assess the action of Accellerase 1500 alone, xylanase was omitted.

All hydrolyses were performed at 50°C for 24 h with continuous stirring (250 rpm) in a screwed-capped glass tubes, and then stopped by heating at 95°C for 5 min. For analysis, the reaction mixture was centrifuged (10 000 x *g* for 2 min) and then the supernatant was filtered (PTFE, 0.22 µm), before injection onto a high performance anion exchange chromatography system with pulsed amperometric detection (HPAEC-PAD). For monosaccharide analysis, separation was achieved at 30°C over 25 min on a Dionex

CarboPac PA-1 column (4×200 mm), equipped with its corresponding guard column and equilibrated in 4.5 mM NaOH and running at a flow rate of 1 ml min^{-1}. For the analysis of xylo-oligosaccharides (XOS), a Dionex CarboPac PA-100 column (4×200 mm), equipped with its corresponding guard column and equilibrated in 4.5mM NaOH was employed. Separation of various XOS was achieved by applying a gradient of NaOAc (5 to 85 mM) in 150 mM NaOH over 30 min at 30°C, using a flow rate of 1 ml min^{-1}. Appropriate standards (monosaccharides such as L-arabinose, D-xylose, D-glucose and D-galactose and various XOS displaying DP 2 to 6) at various concentrations (2 - 25 mg l^{-1}) were used to provide quantitative analyses. Finally, the quantitative results from HPAEC analysis (monomeric and oligomeric sugars) were converted into the amount of soluble monosaccharide equivalents (designated ''average solubilised weight''), and the percentage conversion was calculated as follows, either in terms of xylose or glucose:

$$\text{Conversion \%}_{tot.\ N} = \frac{\text{average solubilized } N}{\text{theoretical total } N} \times 100\% \ (w/w)$$

''N'' represents xylose or glucose, and the ''theoretical total N'' is the total amount of sugar N present in the initial straw sample (Table II-1).

II.2.6 Determination of kinetic parameters

To measure the kinetic parameters of Tx-Xyl and its mutants, BWX and LVWAX were used as substrates at eight different concentrations (0 – 12 g l^{-1}). Hydrolysis reactions (1 ml) were performed at 60°C in NaOAc, pH 5.8 using approximately 4.5 and 3.5 nM of xylanase for BWX and LVWAX assays respectively. During the course of the reaction, aliquots (100 µl) were removed at 3-min intervals, and immediately mixed with an equal volume of 3,5-dinitrosalicylic acid (DNS) regent to stop the reaction. The quantity of solubilised reducing sugars present in samples was assessed by the DNS assay(Miller 1959). Finally, results were analyzed using SigmaPlot V10.0, which generated values for k_{cat} and K_M. Taking into account the heterogeneous nature of the substrates computed K_M values are apparent values having units of g. l^{-1}.

II.2.7 Thermostability assay

To measure the thermostability of the xylanases used in this study, enzyme solutions (100 mM in 10 mM Tris·HCl buffer, pH8.0) were incubated at 50 and 60°C for up to 6 h. At intervals, aliquots were removed and used to measure residual xylanase activity on BWX (at

5 g l^{-1}) at 60°C using the DNS method to quantify solubilised reducing sugars. One unit (1 U BWX) of xylanase activity was defined as the amount of xylanase required to release 1 µmol of equivalent xylose per minute from BWX. Enzyme half life (t$_{1/2}$) was deduced by fitting the curve of ln(residual activity) = kt where t is the time and k is the slope, and t$_{1/2}$ = k^{-1}ln(0.5) (You et al. 2010).

II.2.8 Determination of melting temperature by Differential Scanning Fluorimetry (DSF)

A sample in 20 mM Tris-HCl buffer, pH 8.0 was prepared that contained 100 mM NaCl, SYPRO® Orange (Invitrogen, final concentration 10X), and an aliquot (final concentration of 6.75 µM) of Tx-Xyl or mutant xylanases thereof. Negative controls containing either Sypro or xylanase alone were analyzed in parallel. A CFX96 Real-Time PCR Detection System (Bio-Rad) was used as a thermal cycler and the fluorescence emission was detected using the Texas Red channel (λ_{exc}= 560 - 590 nm, λ_{em} = 675 - 690 nm). The PCR plate containing the test samples (20 µl per well) was subjected to a temperature range from 20°C to 99.6°C with increments of 0.3°C every 3 seconds. The apparent melting temperature (T$_m$) was calculated by the Bio-Rad CFX Manager software.

II.3 Results

II.3.1 Screening of randomly mutagenized xylanase librairies

The different steps of the engineering strategy are summarized in Figure II-2. The initial phase of this work involved the use of epPCR to generate random biodiversity. In preliminary work, we observed that more than 10 base mutations/kb produced >70% inactive clones. Therefore, a progressive strategy employing three successive rounds of epPCR was preferred, with moderate mutational charge (5-7 base mutations/kb) at each stage. The results of screening at each round are summarized in Table II-2. Regarding the first round of screening, this work has already been reported by Song *et al* (2010). Although the best mutant from this first round, designated Tx-Xyl-AF7, displays a wild type amino acid sequence, its DNA sequence contains two mutations (at nucleotiode positions 27 and 516) that cause 2-fold higher expression of the recombinant enzyme. Therefore, the sequence encoding Tx-Xyl-AF7 was used as the template for the second round of epPCR.
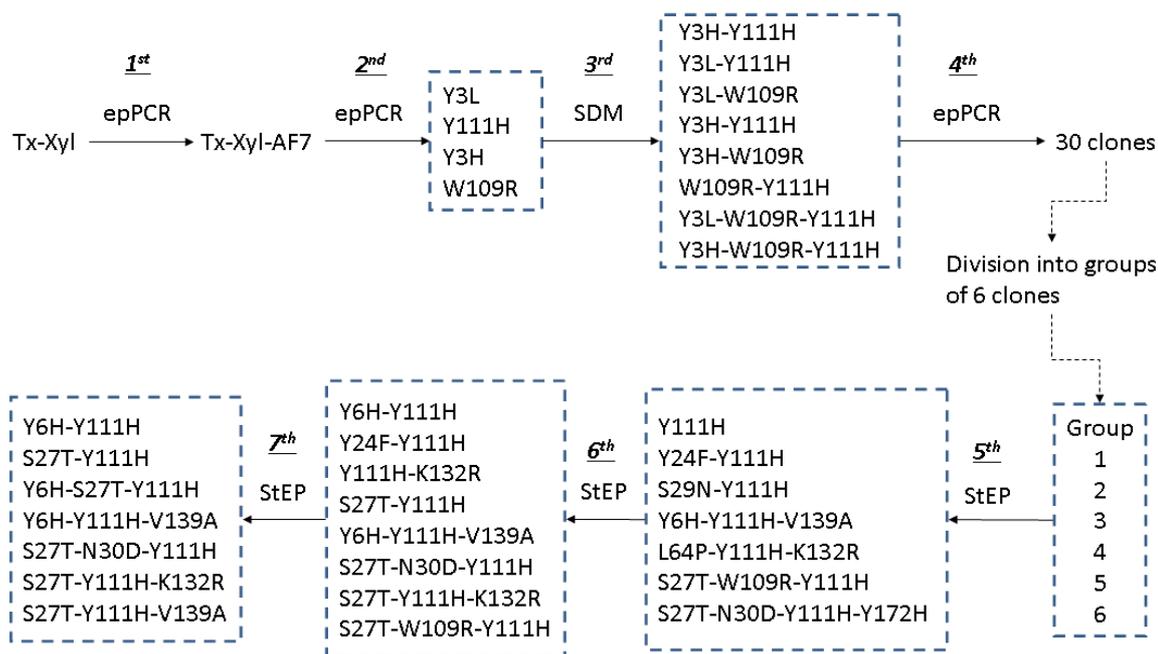
Figure II-2. Flowchart of *in vitro* evolutional process. The terms epPCR, SDM and StEP are abbreviations of error-prone PCR, site-directed mutagenesis and staggered extension process respectively. The best-performing mutants selected as parental input for the next round of evolution are boxed, and mutants are named after the point mutations that characterize them.

DNA sequence analysis of 10 library clones, taken from the 2nd generation library, revealed an average mutation rate of 5.4 base substitutions/kb and a transition/transversion ratio of approximately 1.4, indicating that the mutations were relatively unbiased in this respect. A total of 4 333 clones were screened on In-WS, and the four most active clones (>4CV) were selected, using the activity of Tx-Xyl-AF7-bearing clones as the base case for comparison. DNA sequencing revealed that all four clones were characterized by single amino acid changes. Two clones were mutated at position 3 (Y3L and Y3H), while two others were mutated at independent, but neighbouring locations (W109R and Y111H).

Examination of the 3D structure of Tx-Xyl, reveals that Y3 lies in the distal glycon part of the active site cleft, while W109 and Y111 are situated nearby and in the thumb region respectively, thus all three residues are potentially important for enzyme function. For this reason, at this stage in the experiment it was decided to focus on these mutations for the

creation of further mutant libraries. However, to ensure that all of the possible permutations would be present in the 3rd generation, recombination was achieved using site-directed mutagenesis. Consequently, five double mutants (Y3L-W109R, Y3L-Y111H, Y3H-W109R, Y3H-Y111H and W109R-Y111H) and two triple mutants (Y3L-W109R-Y111H and Y3H-W109R-Y111H) were created. Together with the other four original single-mutants, these were used as parental templates for the next round of epPCR, which led to the creation of a 4th generation.

Table II-2. Summary of directed evolution for improvement of Tx-Xyl xylanase activity

Generation	Library type	Screening substrate	No. of variants screened	CV of wild-type control *	% of clones with improved activity					No. of hits selected
					>4CV	>5CV	>6CV	>7CV	>8CV	
1st	epPCR on wild-type gene	In-WS	264	11.1±1.3%	0.4%	0.4%	-	-	-	1
2nd	epPCR on Tx-Xyl-AF7 gene	In-WS	4333	18.1±5.4%	0.1%	-	-	-	-	4
3rd	Site-directed mutagenesis †	-	-	-						11
4th	epPCR using 11 parental genes	Dpl-WS	4300	10.9±2.2%	1.2%	0.6%	-	-	-	30
5th	Recombination	Dpl-WS	3840 (≈ 2500) ‡	8.1±0.6%	1.4%	6.0%	2.1%	0.8%	0.1%	7
6th	Recombination	Dpl-WS	864 (1847) ‡	10.2%	9.3%	2.8%	0.9%	0.2%	-	8
7th	Recombination	Dpl-WS	864 (127) ‡	11.3%	19.5%	7.5%	2.4%	0.5%	0.2%	7

* CV = coefficient of variation. Wild type Tx-Xyl original coding was only used in the 1st generation. For 2nd generations, variant Tx-Xyl-AF7 sequence (silent mutations T27C and C516T) is used as template.

† Site-directed mutagenesis used to create mutants Y3L-W109R, Y3L-Y111H, Y3H-W109R, Y3H-Y111H, W109R-Y111H, Y3L-W109R-Y111H and Y3H-W109R-Y111H.

‡ Values between brackets are the number of theoretical mutation combinations, which only depends on the mutation number (designated n) in the template genes and is calculated by $\sum_{i=1}^{n} C_n^i$.

For the 5th generation, since only the templates for group 1 were sequenced, the total number of theoretical combinations was calculated by multiplying by 5 the number of theoretical combinations in group 1.

To efficiently challenge clones present in the 4th library, the microtitre plate assay was modified by replacing In-WS with Dpl-WS. The principle behind this was to select clones that produce enzymes that can actually hydrolyze arabinoxylans that are inaccessible or resistant

to wild type xylanase. The key features and performance descriptors of this modified assay are summarized in Table II-1. Overall, the CV value for individual wells of Tx-Xyl-AF7 control varied between 8 to 11%, indicating that this screen was sufficiently reliable for library screening.

DNA sequence analysis of a randomly picked sample of 4th generation library clones revealed an average mutation rate of 7.2 nucleotide substitutions/Kb. Likewise, functional screening using the modified Dpl-WS assay indicated that 0.6% of screened clones presented activities that were significantly higher (>5CV) than the mean value of the activity of Tx-Xyl-AF7 clones. Therefore, the top 30 clones were isolated and used for subsequent rounds of DNA recombination.

II.3.2 Optimization of mutant xylanases using DNA recombination

To further increment the functional fitness of the enzymes expressed by the candidate clones obtained from random mutagenesis, the StEP DNA shuffling approach (Zhao and Zha 2006) was adopted, because it offers a much simpler procedure than classical DNA shuffling (Chaparro-Riggers et al. 2007; Zhao and Zha 2006). This method was used to successively create 5th, 6th and 7th generation libraries. To appreciate the impact of the iterative use of StEP on overall library fitness, Figure II-3 shows the relative performance of 4th to 6th generation libraries. At each generational increment, library fitness increased in accordance with the expectations (Cirino et al. 2003; Zhao et al. 1998; Zhao and Zha 2006). The results of statistical analyses performed on the three successive libraries (5th, 6th and 7th generations) that were created using this method are summarized in Table II-2.

For the initial round of DNA shuffling, 30 clones were used as parental input. Since we considered that 30 templates was quite a high number to handle, these were randomly separated into 5 groups of 6 clones and DNA shuffling was performed on each group. After DNA shuffling, the 5 libraries were submitted to screening using the modified Dpl-WS assay. This step allowed the selection of 7 hits whose activity were significantly higher (>7CV) than the mean value of the activity of Tx-Xyl-AF7 clones. DNA sequencing revealed that these 7 clones contained 11 point mutations that give rise to new amino acid substitutions, and that constituted Y111H and 6 mutational combinations (Figure II-2). As before, the 7 mutants

were used as parental input for two further rounds (6th and 7th) of DNA shuffling. However, of the 11 point mutations present in above 7 parental clones the locations of Y24F, S27T, S29N and N30D were extremely close, meaning that the likelihood of homologous recombination would be low, thus reducing the actual number of mutational combinations when compared to the theoretical number. Taking this into account, a much smaller sample of library clones was screened after the 6th round of the *in vitro* evolution process.
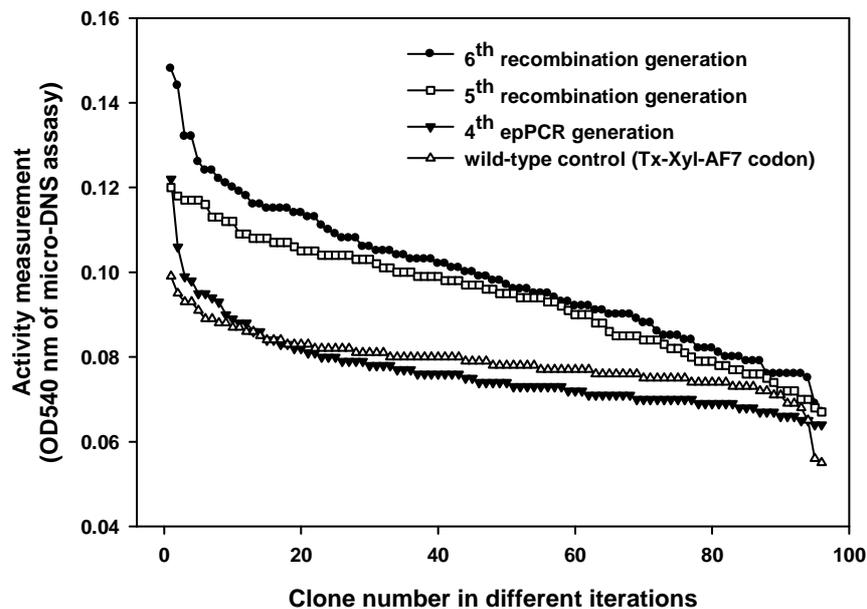
Figure II-3. Evolutional acceleration among consecutive iterations screening on Dpl-WS. The X-axis is 96 clones in one microtitre plate which is randomly selected from wild-type control (△, using Tx-Xyl-AF7 coding sequence), 4th random mutagenesis library (▼), 5th recombinant library (□) and 6th recombinant library (●). Y-axis indicates the activity value of corresponding clone in the screening. The same batch of Dpl-WS substrate was used for the 4 experiments.

After the creation of the 7th generation library, the experiment was stopped, because DNA sequencing of the highest performing 7th generation clones showed that 5 mutational combinations out of a total of 7 had already identified in the 6th generation (Figure II-2). This observation suggested that the evolutionary itinerary had almost reached an end, with very little new biodiversity being introduced.

Of the 7 best performing 7th generation clones, Y6H-Y111H and Y6H-S27T-Y111H displayed the highest activity increase (>8 CV) in the screening, compared to that of wild-type control (Tx-Xyl-AF7). In addition, among the 6 amino acid substitutions that were detected in clones obtained from DNA shuffling, Y111H was present in every template and the frequency of Y6H and S27T increased from the 5th generation to the 7th generation (Table II-3). Consequently, we decided to focus on clones containing these three amino acid changes for enzyme production and characterization. Overall mutants that were retained for characterization included Y6H-Y111H, S27T-Y111H and Y6H-S27T-Y111H from the 7th generation screening and the single mutants Y111H, Y6H and S27T.

Table II-3. Mutation frequency in the 5th – 7th iterations of recombination library.

Generation	Y6H	Y24F	S27T	S29N	N30D	L64P	W109R	Y111H	K132R	V139A	Y172H
5th	22.2%	11.1%	22.2%	11.1%	11.1%	22.2%	11.1%	100%	22.2%	22.2%	11.1%
6th	28.6%	14.3%	28.6%	-	14.3%	-	-	100%	14.3%	14.3%	-
7th	44.4%	-	55.6%	-	11.1%	-	-	100%	22.2%	22.2%	-

II.3.3 Saturation mutagenesis at position 3 and 111

Among the 2nd generation clones, selected for higher activity on In-WS, two amino acid positions, 3 and 111, were pinpointed as potentially interesting locations. Therefore, in addition to the use of Y3H and Y111H as parental templates for further random mutagenesis and DNA shuffling, site-saturation mutagenesis was performed to investigate the importance of these two residues on the activity against recalcitrant AX in wheat straw (i.e. Dpl-WS). In each case an adequately large library was created and 288 clones were screened using the modified Dpl-WS assay, which was sufficient from a statistical point of view to ensure that all permutations were present (Georgescu et al. 2003). Additionally, a random sample of each library was submitted to DNA sequence analysis in order to control the success of the experiment.

Figure II-4 shows the results of the screening of the two site-saturation libraries. Overall, the Y111N (N represents any amino acid) library provides a larger population of improved clones, though both libraries contain a small minority of clones that display activities that are above the value of μ+4σ of wild-type control. Three highest performing clones were selected from

each library and analyzed by DNA sequencing. All three clones from the Y3N library displayed the same Y3W mutation, whereas two clones from the Y111N library were phenotypically and genotypically identical (encoding the mutation Y111S) and one displayed an Y111T mutation. In view of these results, three individual clones encoding Y3W, Y111S and Y111T were retained for further characterization.
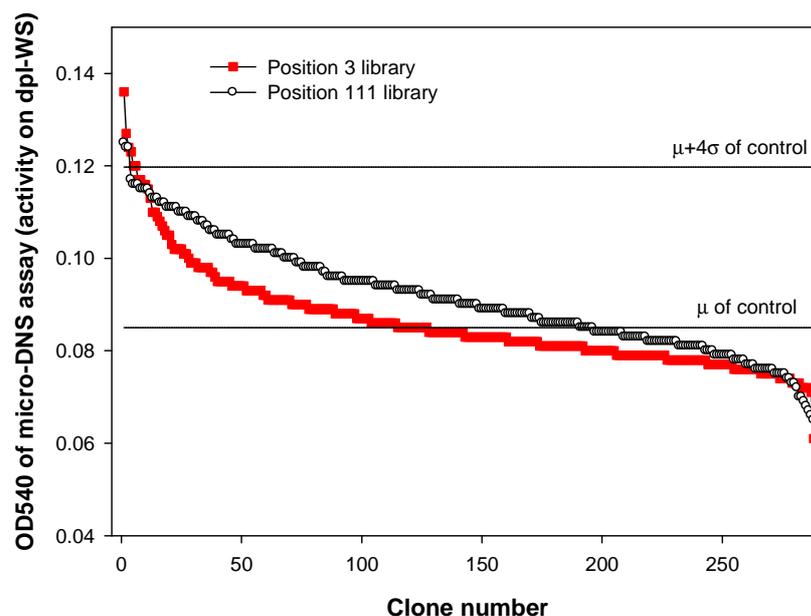
Figure II-4. Dpl-WS screening of site-saturation libraries targeted on position 3 (■) and position 111 (○). The X-ray is the numbers for 288 clones in each library, and the Y-axis represents the corresponding activity measurement of each clone towards Dpl-WS. Two solid lines represent the mean value (μ) and mean value + 4 ×standard deviation (μ + 4σ) of the wild-type (Tx-Xyl-AF7 code) activities screened in the same condition.

II.3.4 Characterization of key properties of the Tx-xyl mutants

Since the screening of mutant enzyme libraries obeys the maxim "you get what you screen for", the mutants selected in this work were only improved with respect to the hydrolysis of wheat straw. Hence, other important properties such as thermostability could have been negatively affected. Consequently Table II-4 summarizes the evaluation of thermostability that was performed on each purified mutant. Although the thermostability of some mutants at 60°C was clearly affected (e.g. that of Y6H and Y6H-Y111H), all of the enzymes were sufficiently stable to be able to measure kinetic properties without any major modifications

to the protocols that were routinely used to characterize wild type Tx-Xyl. It is also noteworthy that all of the mutants were highly stable at 50°C, since measured activity remained stable over a 6h incubation period.

Table II-4. Thermostability of Tx-Xyl and the mutants thereof. The melting temperature (T_m) was determined using differential scanning fluorimetry (DSF) and the half-life ($t_{1/2}$) was defined period necessary for the initial activity to be reduced by 50% at 60°C.

Mutant	T_m (°C)	$t_{1/2}$ at 60°C (h)
Tx-Xyl	75.9	5.4
Y6H	72.9	2.6
S27T	76.4	6.4
Y111H	75.1	3.9
Y6H-Y111H	72.7	2.7
S27T-Y111H	75.4	6.4
Y3W	73.1	3.2
Y111S	75.1	3.6
Y111T	74.9	5.0

Each of the mutants was characterized with regard to its ability to hydrolyze BWX and LVWAX. According to another recent study(Song et al. Manuscript in preparation), BWX is devoid of α-L-arabinosyl substitutions, and LVWAX displays an A/X ratio of 0.54. Concerning wild-type Tx-Xyl, its turnover number and specificity constant were higher for LVWAX, though the apparent K_M value was lower on BWX. This tendency was also displayed by the majority of the mutants (Table II-5). Nevertheless, the specificity constants for the single mutants Y3W, Y111S, Y111T, Y6H and S27T were almost identical for both substrates. Regarding the apparent values of K_M, all of the mutants displayed improved affinity for BWX, but this was not the case for LVWAX. Notably, Y111H was the mutant that displayed the best affinity for BWX, while its affinity for LVWAX was unaltered. Other the hand, the rate constant for Y111H-mediated hydrolysis of BWX was lowered when compared to that of the wild type enzyme, but was improved on LVWAX. Intriguingly, the opposite was true for Y111T, for which the value of *k_{cat}* was 48% greater than that of Tx-Xyl on BWX, but identical to that of Tx-Xyl on LVWAX. When Y111H was combined with other mutations (e.g. S27T-

Y111H or Y6H-Y111H), its influence on the specificity constant appeared to be dominant, annulling the improved activity on BWX, displayed by the single mutants S27T and Y6H.

Table II-5. Kinetic parameters of Tx-Xyl wild-type and mutants. Kinetic parameters were determined for hydrolysis reactions using either birchwood xylan (BWX, no arabinosyl substitution) or low viscosity wheat arabinoxylan (LVWAX, arabinose/xylose ratio of 0.54) respectively.

Mutant	Kinetic parameters						SR[‡]
	BWX			LVWAX			
	k_{cat} (s^{-1})	K_M[†] (g L^{-1})	k_{cat}/K_M[†] (s^{-1} g^{-1}. L)	k_{cat} (s^{-1})	K_M[†] (g L^{-1})	k_{cat}/K_M[†] (s^{-1} g^{-1}. L)	
Tx-Xyl	610.5±19.6	2.54±0.19	242.1	1699.4±95.9	5.10±0.09	333.1	0.67
Y6H	806.1±61.2	2.37±0.27	340.5	2081.6±16.4	5.73±0.09	363.0	0.94
S27T	742.9±22.4	1.93±0.15	376.5	1936.0±19.2	4.81±0.04	402.5	0.94
Y111H	449.3±23.2	1.54±0.14	292.0	1889.0±72.9	5.01±0.07	376.7	0.78
Y6H-Y111H	433.1±12.5	1.91±0.11	226.8	1834.7±75.4	5.33±0.35	345.9	0.66
S27T-Y111H	535.8±30.2	1.72±0.21	311.5	1906.1±4.4	4.39±0.02	434.1	0.72
Y3W	704.5±29.8	2.11±0.14	333.4	1743.0±26.5	5.51±0.02	316.3	1.05
Y111S	758.8±15.9	2.12±0.04	358.6	1755.0±106.3	4.75±0.14	369.4	0.97
Y111T	905.2±17.2	2.35±0.15	369.4	1740.4±41.8	4.81±0.06	361.4	1.02

[†] the heterogeneous nature of the substrate allows the determination of a relevant K_M in stead of a true value for K_M

[‡] SR: ratio of the specificity constants (BWX/LVWAX)

II.3.5 Assessment of the impact of Tx-Xyl mutants on wheat straw

To further evaluate the altered properties of the different mutants, their activities on the original wheat straw samples (In-WS and Dpl-WS) were examined. Reactions were performed using pure preparations of wild-type and mutant xylanases either alone or in the combination with Accellerase 1500 (a cellulase cocktail). The results of HPAEC-PAD analyses performed on the reaction supernatants are shown in Figure II-5.A and B, which present as the conversion of total xylose and glucose (i.e. %$_{tot. xyl}$ and %$_{tot. glu}$, w/w) in the straw residues. The soluble sugar yields are summarized in Table II-2S and 3S.

The hydrolysis of Dpl-WS revealed that all of the mutants could release further amounts of soluble xylose equivalents and that their performance was superior to that of wild type Tx-Xyl. The mutants S27T and Y111T produced the most outstanding results, because these could release 2.3-fold and 2.1-fold more xylose equivalents from Dpl-WS than Tx-Xyl. The

lowest performers were Y111H and Y3W, which yielded 35% and 46% more xylose equivalents respectively (Figure II-5.A). However, it should be noted that even the best variant S27T could only release 2.5%$_{tot. xyl}$ of Dpl-WS (5.5 g xylose per kg wheat straw), which is a witness to the recalcitrance of this substrate.

For the hydrolysis of In-WS (pH 5.8), wild type Tx-Xyl released 43.7 g equivalent xylose per kg wheat straw. This represents 4.4% of the dry weight and 16.7% of total xylan (16.7%$_{tot. xyl}$) content. Similar results were obtained for the mutants Y6H, Y6H-Y111H, Y6H-S27T-Y111H and Y3W, but five other mutants yielded higher amounts (18.6 – 20.4%$_{tot. xyl}$) of soluble xylose equivalents, with the best mutant being Y111T (Figure II-5.A).

The five mutants, displaying improved activity on In-WS, were further selected to investigate synergy with cellulases on In-WS, operating at the optimum pH for Accellerase (pH 5.0). Likewise, suitable control reactions at pH 5.0 were performed using only mutant xylanases, or wild type Tx-Xyl. All controls revealed that the different xylanases displayed reduced hydrolytic capacity, compared to their activity at pH 5.8 (Figure II-5.A). According to its manufacturer, Accellerase 1500 principally contains endoglucanase and β-glucosidase activities. In our trials Accellerase alone, was able to solubilise 7.3%$_{tot. xyl}$ and 18.9%$_{tot. glu}$ In-WS (Figure II-5.B). However, in combination with xylanases higher yields of xylose and glucose were measured, which were greater than the sum of the yields of Accellerase or the xylanases alone, clearly revealing the synergistic interactions between the enzyme participants. The mixture of wild type Tx-Xyl and Accellerase solubilised 24.5%$_{tot. xyl}$ and 23.6%$_{tot. glu}$ of In-WS (Figure II-5.B). However, significantly the different mutants were able to improve on this performance, solubilizing 27.4 – 29.0%$_{tot. xyl}$ and 24.9 – 26.4%$_{tot. glu}$ from In-WS. Among the mutants, Y111S provided the highest synergistic effect on xylan solubilization, while Y111T procured the most advantageous effect on glucan solubilization.

(A)

(B)
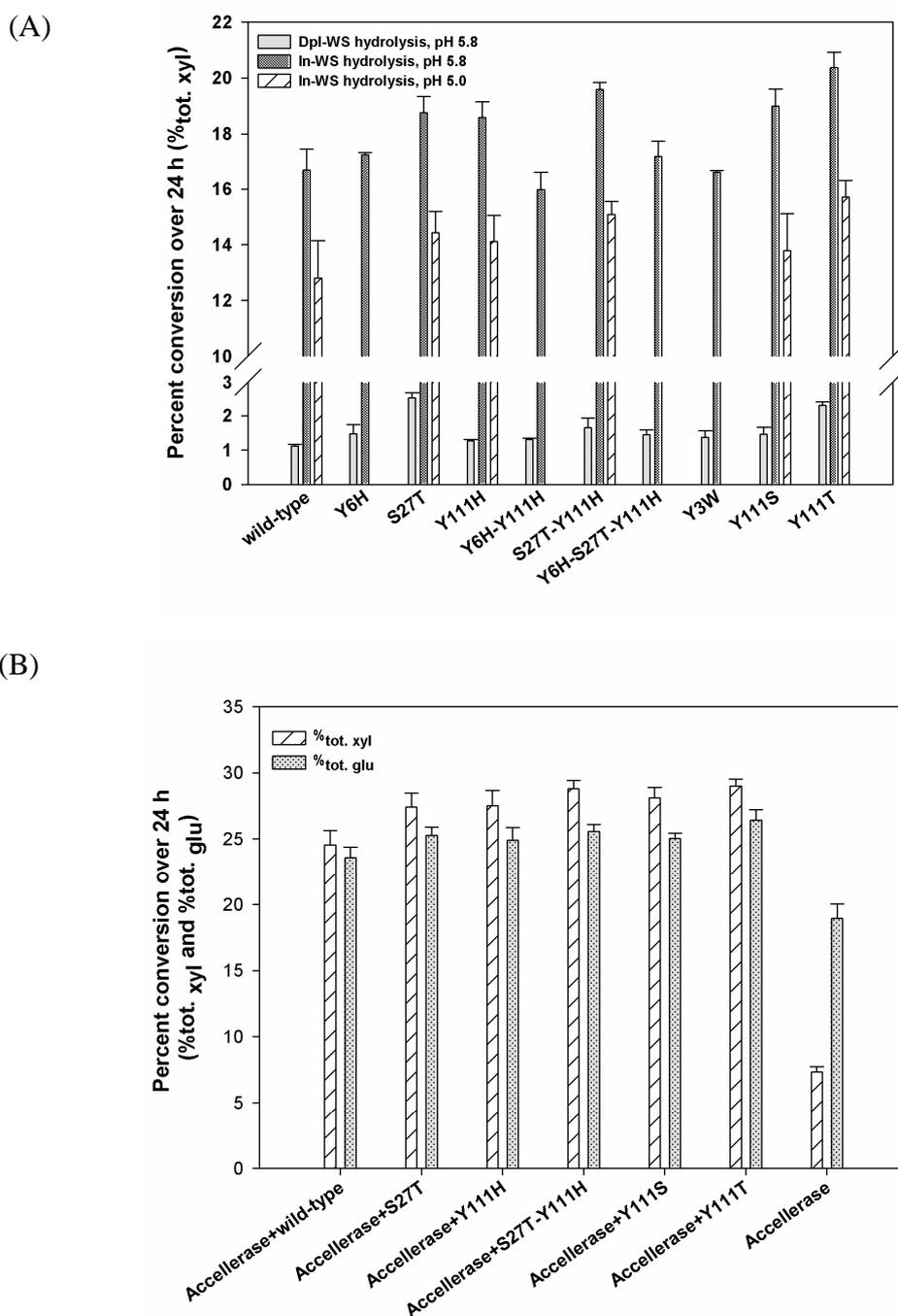
Figure II-5. (A) Percent conversion of total xylose in the Tx-Xyl wild-type/mutant–mediated hydrolyses on Dpl-WS (pH 5.8) and In-WS (pH 5.0 and 5.8). (B) Percent conversion of total xylose and glucose in the In-WS hydrolyses (pH 5.0), performed by the Accellerase 1500 alone and in combination with Tx-Xyl wild-type/mutant xylanases. In both diagrams, the X-axis indicates the enzyme(s) employed in the corresponding hydrolysis reaction. The $\%_{tot.\ xyl}$ and $\%_{tot.\ glu}$ are calculated as described in the methods.

II.4 Discussion

II.4.1 Is enzyme engineering a useful strategy to improve biomass deconstruction?

Artificial enzyme evolution, relying on *in vitro* random mutagenesis and DNA recombination techniques, is a powerful strategy to pinpoint functional determinants and to rapidly improve enzyme fitness with regard to a variety of physical or biochemical properties (Arnold and Moore 1997; Arnold and Volkov 1999; Johannes and Zhao 2006; Walsh 2001). However, the need for an appropriate screen is vital. In this work, we relied on a previously described screening method, which allowed us to address a highly ambitious target, which was the isolation of enzymes that display higher activity on raw biomass. To our knowledge, no such enzyme engineering has yet been attempted, mainly because biomass-degrading enzymes are improved for their activity on artificially isolated biodiversities or pretreated biomass, wherein the total chemical and structural complexities of complex biomass are omitted, or only rich in cellulose and lignin and hemicelluloses are very minor components (Lantz et al. 2010; Nakazawa et al. 2009; Zhang et al. 2006) .

Therefore, the underlying rationale of our approach was to investigate to what extent the fitness of a xylanase, or for that matter any other biomass-degrading enzyme, can be independently improved for hydrolysis of complex biomass, without interfering with the structural and chemical complexity of the substrate. Likewise, we hoped to provide a novel angle to the understanding of the factors that govern the enzymatic deconstruction of raw biomass.

Our previous study revealed that the Tx-Xyl-mediated hydrolysis of wheat straw is a complex reaction that cannot be modelled using Michaelis-Menten kinetics and does not reach completion even at high enzyme loading and long time periods (Song et al. 2010; Zilliox and Debeire 1998). To achieve the first phase of the reaction requires quite long incubation times (approximately 8 h), thus screening using raw wheat straw (i.e. In-WS) provides a means to find variants that display improved initial catalytic rates, which can result either from the improvement of intrinsic catalytic properties of the xylanase, or from an increase in

enzyme production. On the other hand, the use of In-WS is not appropriate to isolate xylanases that will surpass the sugar solubilization yield of the wild type Tx-Xyl. For this purpose, it is more appropriate to use Tx-Xyl-pretreated wheat straw (i.e. Dpl-WS), which should provide a means to identify enzyme variants that can accelerate the latter phase of the reaction and better surmount the obstacles that prevent further action by Tx-Xyl. Therefore, in the strategy developed here, both screening approaches were applied, first in an attempt to accelerate the reaction and second to improve the overall impact of xylanase action on wheat straw.

Overall, all of the qualitative indicators that are presented here show that the enzyme evolution approach was successful in increasing the fitness of Tx-Xyl for biomass hydrolysis. At each step, clones with ever increasing activity could be selected and the ultimate analysis of the best clones revealed that several could actually better hydrolyze wheat straw, especially when their action was coupled to a cellulase cocktail. Nevertheless, unsurprisingly the overall impact of the improvements was modest, but these results need to be considered in the light of current knowledge.

Two recent studies (Hansen et al. 2010; Kristensen et al. 2008) have attempted to relate enzyme action on wheat straw to changes at the ultrastructural level. These authors have shown that a mild hydrothermal pretreatment (185°C, 10 min) releases approximately 34% of available xylans (i.e. approximately 8.2% of the initial DW), which appear to come from the pith that lines the central lumen of wheat straw. Further treatment of the sample with a cellulase cocktail released glucose and xylose from cellulose microfibrils and xylans respectively, apparently present in the parenchyma cells that form the cortex. However, enzymatic degradation was impotent on lignified cells (e.g. sclerenchyma cells). In our experiment, total xylans in wheat straw represent approximately 26% DW and Tx-Xyl can release 16.7% of these (i.e. 4.4% DW). The mutant Y111T is able to solubilise approximately 21.9%$_{tot. xyl}$ or 5.3% DW over a 24 h period. Taken together, our results reveal that the hydrolysis of wheat straw using Tx-Xyl variants procures solubilization yields that are inferior, but not dissimilar, to those obtained using mild hydrothermal treatment and thus allows us to speculate that the same structures are affected, i.e. the pith and parenchyma cells.

The failure of Tx-Xyl, or variants thereof, to further solubilise xylans is probably not linked to intrinsic catalytic potency or to substrate selectivity of Tx-Xyl and its mutants, but rather to the inaccessibility of the substrate. Indeed, coupling of wild type Tx-Xyl to that of a cellulase cocktail clearly revealed a certain degree of synergy, releasing approximately 24% of the theoretical yield of sugars, Significantly, mutants generated in this work amplified this synergy and achieved higher levels of sugar solubilization, indicating that the enzymatic removal of cellulose exposes xylan and vice versa. Possibly, the improved action of the mutants allows a slightly more profound degradation of the parenchyma cells that form the cortex of wheat straw. However, the results of this study indicate that enzyme engineering alone cannot overcome the limits imposed by the lignin barrier, which is progressively exposed by the peeling action of the xylanse/cellulases cocktail.

II.4.2 Structure-function relationships revealed in this study

One of the remarkable findings in this study is the identification of a relatively small number of mutations. After six rounds of combined mutagenesis and DNA shuffling, seven mutants possessing a total of six point mutations were identified. Among these mutations, three emerge (amino acids 6, 27 and 111) as important positions, because of their reoccurrence in the seven mutants. In addition, another three mutants (Y3W, Y111S and Y111T) were isolated from site-saturation mutagenesis libraries, in which amino acids 3 and 111 respectively were targeted. Tyr3 and Tyr6 are located at the B2 β-strand in the N-terminal region of Tx-Xyl, whereas Ser27 forms part of the "knuckles" region of fingers and Tyr111 is located on the thumb (Figure II-6.A). The examination of the different combinations that were obtained reveals that generally these mutations did not provide additive benefits. For example, regarding the mutants Y6H-Y111H, S27T-Y111H and Y6H-S27T-Y111H, the two point mutation variants Y6H and S27T displayed greater hydrolytic potency on Dpl-WS than any of these combinations. Similarly, S27T displayed the highest catalytic efficiency towards the two soluble xylan substrates, BWX and LVWAX. Therefore, it appears legitimate to consider the impacts of the different mutations independently.

The findings presented here concerning the reduced thermostability of mutants displaying substitutions at positions 6 (Y6H) and/or 111 (Y111H) clearly provide support for the existence of hydrophobic patches that might mediate the oligomerization, and thus the

thermostabilization, of Tx-Xyl in solution. According to Harris *et al* (1997), Tyr6 and Tyr111 are surface exposed aromatic amino acids that along with 9 other aromatic residues participate in the formation of intermolecular "sticky patches" that form the basis for thermostability in Tx-Xyl. Nevertheless, it is also important to note that not all mutations at position 111 produced the same effect. Notably, the mutant Y111T displayed thermostability very close to that of the wild type Tx-Xyl. Interestingly, the mutant S27T actually increased thermostability, which agrees with a trend among GH11 xylanases that correlates thermostability with an increased Thr : Ser ratio (Hakulinen et al. 2003; Park et al. 2009).

Among the six mutants bearing single substitutions, S27T, Y111H, Y111S and Y111T displayed improved hydrolysis of In-WS and synergy with the cellulase cocktail. On the other hand the selection of the mutants Y6H and Y3W in our assay was more surprising, because these did not appear to improve wheat straw hydrolysis, although their specificity towards BWX was clearly altered and Y6H displayed the highest k_{cat} value on both BWX and LVWAX. The mutants S27T, Y111S and Y111T also showed increased specificity towards BWX, indicating that all single site mutants selected in our assay had acquired an improved ability to hydrolyze less substituted xylans, displaying a Ara: Xyl ratio that is comparable to that of wheat straw xylan (Ara : Xyl ratio of 0.091). Curiously, the only exception to this trend was the double mutant Y6H-Y111H, which displayed unaltered specificity on In-WS, when compared to wild type Tx-Xyl.

The amino acid Ser27 is located in a region that has been identified as a secondary binding site (SBS) in the GH11 xylanases from *Bacillus circulans* (Ludwiczek et al. 2007) and *Bacillus subtilis* (Cuyvers et al. 2011; Vandermarliere et al. 2008). Tx-Xyl shares 73% amino acid identity with the xylanase from *B. circulans* xylanase*, and this figure increases to 81% when one just considers the SBS determinants, suggesting that a functional SBS might be present in Tx-Xyl (Figure II-6.B). In this context, it is noteworthy that Ser27 is located in a relatively deep part of a surface groove in Tx-Xyl that is linked to a shallower region via Ser25, and that surface grooves are potential ligand binding sites (Bhinge et al. 2004). Therefore, one can speculate that Ser27 constitutes an element of a SBS in Tx-Xyl. Functionally, it is proposed that the SBS in certain GH11 xylanases interacts with three or four xylosyl units via hydrogen bonds and Van der Waals interactions, and possibly improves binding of xylan polymers in

the active site cleft (Ludwiczek et al. 2007). The mutation of Ser27 to Thr certainly leads to a localized increase in hydrophobicity, which is probably favorable for xylan binding to the putative SBS. Indeed, experimental evidence supports this, because the mutant S27T significantly reduced the Michaelis constant for the hydrolysis of BWX and, to a lesser extent, for LVWAX. In this respect, it is also noteworthy that among the other mutations identified during the directed evolution process (Table II-3), figure S29N, N30D and V139A are also in the vicinity of the putative SBS region in Tx-Xyl. Therefore, a complementary study of these mutations could be an interesting way forward to better define the Tx-Xyl SBS and understand its effect on the enzyme activity.

The thumb loop is known to be of prime functional importance in GH11 xylanases. The open and closing of this loop almost certainly plays a key role in substrate selectivity and binding (Muilu et al. 1998; Murakami et al. 2005; Paës et al. 2007; Vieira et al. 2009) and product release (Pollet et al. 2009); Paës et al, manuscript in preparation) respectively. Regarding substrate binding, the conserved tip of the thumb, composed of the motif Pro-Ser-Ile (position 114 − 116 in Tx-Xyl), is involved in binding of xylosyl residues at the -1 and -2 subsites via hydrogen bonds (Havukainen et al. 1996; Sabini et al. 1999; Vandermarliere et al. 2008). Tyr111 and its opposing neighbour Thr121 are located at the base of the loop where they control the movement of this structure (Paës et al. 2007; Pollet et al. 2009). The mutation of Tyr111 to either His, Ser or Thr reduces the spatial occupancy at position 111 (Figure II-6.C), although this is less so for His, and thus probably renders the loop more mobile and more inclined to fold downwards and inwards towards the -1 and -2 subsites. The overall effects of these changes would be improved catalytic turnover and possibly improved binding affinity, both of which are observed for the mutants Y111S and Y111T.

Regarding the loop movement, the mutation of Tyr6 is also worth consideration. The relatively conservative substitution of this residue by a slightly less bulky histidine clearly improved the enzyme turnover on both BWX and LVWAX, but had a slightly negative effect of substrate affinity in the case of LVWAX. This implies that Tyr6 might influence the movement of the loop, although a direct interaction is impossible. Nevertheless, Trp7 forms part of the -2 subsite site and faces Pro114 and Ile116, which form the thumb tip. Slight adjustments in the position of Trp7 could facilitate the open-close movement of the thumb

loop, with the risk of disturbing the high energy interaction between this residue and the -2 xylosyl moiety.
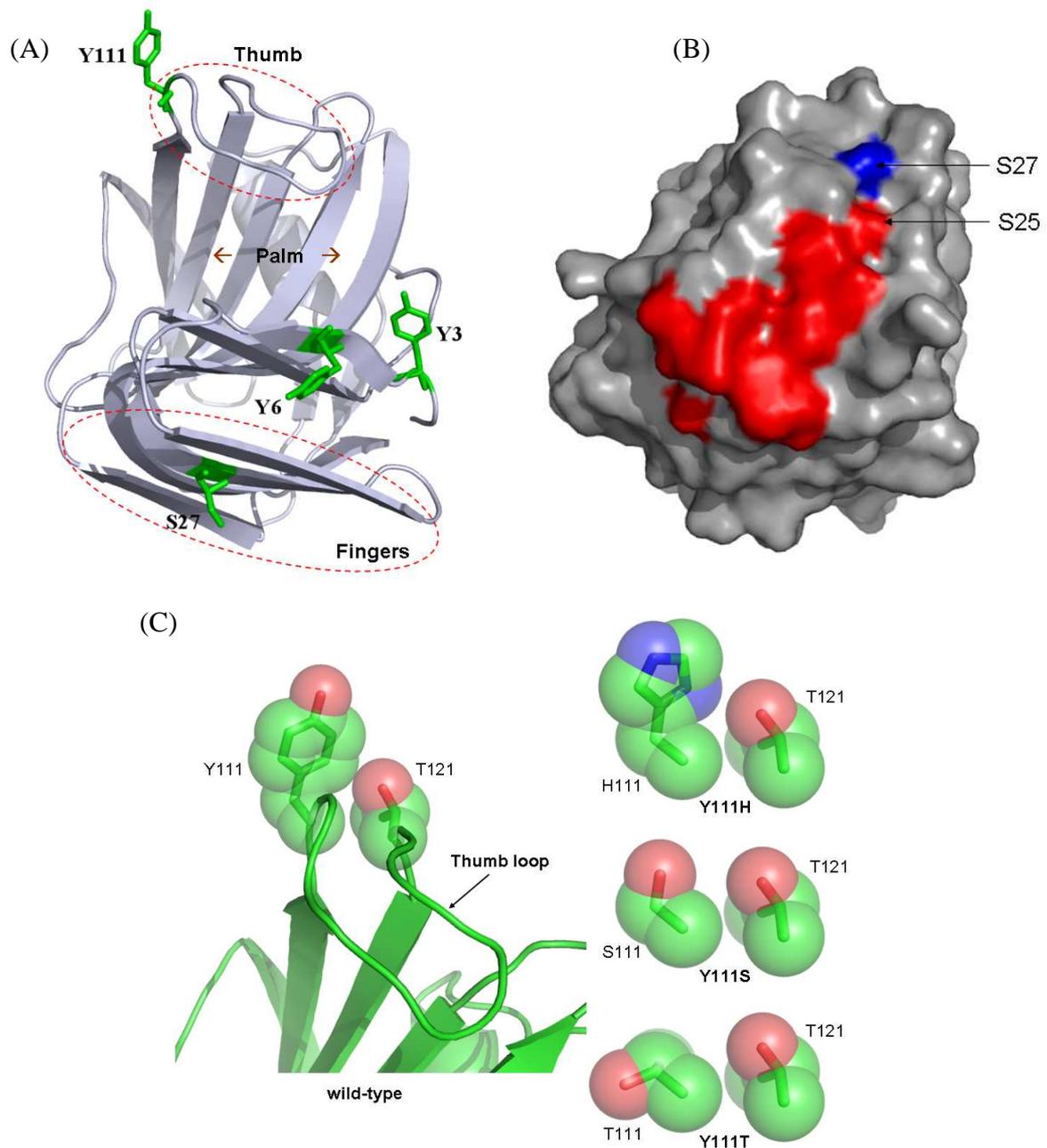
Figure II-6. (A) Four mutated positions (green) in the tertiary structure of Tx-Xyl xylanase. The mutated residues identified from directed evolution are focused on Y3, Y6, S27 and Y111. **(B) Potential secondary binding site (red) including residue S27 (blue) in the surface representation of Tx-Xyl xylanase**. The residues S25 and S27 are indicated by arrows. **(C) Spatial sphere view of the side chains at positions 111 and 121 in the Tx-Xyl wild-type and variants Y111H, Y111S and Y111T.** All the figures are prepared by Pymol software (DeLano 2002).

Finally it is noteworthy that many of the mutations that were identified in this study involved the loss of aromatic side chains. Often, the nonproductive binding by lignin is cited as a major cause of enzyme inefficiency on lignocellulosic biomass (Berlin et al. 2005; Chang and Holtzapple 2000; Tu et al. 2009; Zhu et al. 2008). In an earlier study, it was shown that wild type Tx-Xyl was strongly absorbed by both wheat straw and isolated wheat straw lignin (Zilliox and Debeire 1998). In a more recent study (Boukari, 2011, in press), it has been shown that phenolic acids can act as noncompetitive multisite inhibitors of Tx-Xyl that might provoke conformational alterations of the enzyme. Therefore, it is tempting to speculate that the elimination of surface exposed aromatic amino acid side chains might lower such inhibitory effects.

In conclusion, using a random mutagenesis and directed evolution approach we have been able to generate a number of mutants whose behavior is globally coherent with the screening assay that was employed. Several mutants display improved (albeit modest) hydrolytic activity on wheat straw and shown increased synergy with cellulase, though none are sufficiently potent able to overcome the lignin barrier, which inevitably blocks the way to further hydrolysis of polysaccharides.

II.5 References

Arnold FH, Moore JC. 1997. Optimizing industrial enzymes by directed evolution. Adv Biochem Eng Biotechnol 58:1-14.

Arnold FH, Volkov AA. 1999. Directed evolution of biocatalysts. Curr Opin Chem Biol 3(1):54-9.

Berlin A, Gilkes N, Kurabi A, Bura R, Tu M, Kilburn D, Saddler J. 2005. Weak lignin-binding enzymes. Applied Biochemistry and Biotechnology 121(1):163-170.

Berrin JG, Juge N. 2008. Factors affecting xylanase functionality in the degradation of arabinoxylans. Biotechnol Lett 30(7):1139-50.

Bhinge A, Chakrabarti P, Uthanumallian K, Bajaj K, Chakraborty K, Varadarajan R. 2004. Accurate Detection of Protein:Ligand Binding Sites Using Molecular Dynamics Simulations. Structure 12(11):1989-1999.

Chang V, Holtzapple M. 2000. Fundamental factors affecting biomass enzymatic reactivity. Applied Biochemistry and Biotechnology 84-86(1):5-37.

Chaparro-Riggers JF, Loo BL, Polizzi KM, Gibbs PR, Tang XS, Nelson MJ, Bommarius AS. 2007. Revealing biases inherent in recombination protocols. BMC Biotechnol 7:77.

Chen YL, Tang TY, Cheng KJ. 2001. Directed evolution to produce an alkalophilic variant from a Neocallimastix patriciarum xylanase. Can J Microbiol 47(12):1088-94.

Cirino PC, Mayer KM, Umeno D. 2003. Generating Mutant Libraries Using Error-Prone PCR. In: Arnold FH, Georgiou G, editors. Directed Evolution Library Creation: Humana Press. p 3-9.

Collins T, Gerday C, Feller G. 2005. Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiology Reviews 29(1):3.

Cuyvers S, Dornez E, Rezaei MN, Pollet A, Delcour JA, Courtin CM. 2011. Secondary substrate binding strongly affects activity and binding affinity of Bacillus subtilis and Aspergillus niger GH11 xylanases. FEBS Journal 278(7):1098-1111.

DeLano WL. 2002. The PyMOL Molecular Graphics System. Version Version 1.2r3pre. DeLano Scientific LLC, San Carlos, CA, USA. : http://www.pymol.org.

Dumon C, Varvak A, Wall MA, Flint JE, Lewis RJ, Lakey JH, Morland C, Luginbuhl P, Healey S, Todaro T and others. 2008. Engineering hyperthermostability into a GH11 xylanase is mediated by subtle changes to protein structure. J Biol Chem 283(33):22557-64.

Fraser-Reid BO, Tatsuta K, Thiem J, Glasser WG. 2008. Cellulose and Associated Heteropolysaccharides. Glycoscience: Springer Berlin Heidelberg. p 1473-1512.

Georgescu R, Bandara G, Sun L. 2003. Saturation Mutagenesis. p 75-83.

Hakulinen N, Ossi T, Janne J, Matti L, Juha R. 2003. Three-dimensional structures of thermophilic β-1,4-xylanases from Chaetomium thermophilum and Nonomuraea flexuosa. European Journal of Biochemistry 270(7):1399-1412.

Hansen MAT, Kristensen JB, Felby C, Jogensen H. 2010. Pretreatment and enzymatic hydrolysis of wheat straw (Triticum aestivum L.) - The impact of lignin relocation and plant tissues on enzymatic accessibility. Bioresource Technology 102(3):2804-2811.

Harris GW, Pickersgill RW, Connerton I, Debeire P, Touzel JP, Breton C, Perez S. 1997. Structural basis of the properties of an industrially relevant thermophilic xylanase. Proteins 29(1):77-86.

Havukainen R, Törrönen A, Laitinen T, Rouvinen J. 1996. Covalent Binding of Three Epoxyalkyl Xylosides to the Active Site of endo-1,4-Xylanase II from Trichoderma reesei. Biochemistry 35(29):9617-9624.

Hornsby PR, Hinrichsen E, Tarverdi K. 1997a. Preparation and properties of polypropylene composites reinforced with wheat and flax straw fibres: Part I Fibre characterization. Journal of Materials Science 32(2):443-449.

Inami M, Morokuma C, Sugio A, Tamanoi H, Yatsunami R, Nakamura S. 2003. Directed evolution of xylanase J from alkaliphilic Bacillus sp. strain 41M-1: restore of alkaliphily of a mutant with an acidic pH optimum. Nucleic Acids Res Suppl(3):315-6.

Johannes TW, Zhao H. 2006. Directed evolution of enzymes and biosynthetic pathways. Curr Opin Microbiol 9(3):261-7.

Kristensen JB, Thygesen LG, Felby C, Jorgensen H, Elder T. 2008. Cell-wall structural changes in wheat straw pretreated for bioethanol production. Biotechnology For Biofuels 1.

Kulkarni N, Shendye A, Rao M. 1999. Molecular and biotechnological aspects of xylanases. FEMS Microbiol Rev 23(4):411-56.

Kumar R, Wyman CE. 2009b. Effect of xylanase supplementation of cellulase on digestion of corn stover solids prepared by leading pretreatment technologies. Bioresour Technol 100(18):4203-13.

Lantz S, Goedegebuur F, Hommes R, Kaper T, Kelemen B, Mitchinson C, Wallace L, Stahlberg J, Larenas E. 2010. Hypocrea jecorina CEL6A protein engineering. Biotechnology for Biofuels 3(1):20.

Ludwiczek ML, Heller M, Kantner T, McIntosh LP. 2007. A secondary xylan-binding site enhances the catalytic activity of a single-domain family 11 glycoside hydrolase. J Mol Biol 373(2):337-54.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66(3):506-77, table of contents.

Lynd LR, Zyl WHv, McBride JE, Laser M. 2005. Consolidated bioprocessing of cellulosic biomass: an update. Current Opinion in Biotechnology 16(5):577.

Miller GL. 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Analytical Chemistry 31(3):426.

Miyazaki K, Takenouchi M, Kondo H, Noro N, Suzuki M, Tsuda S. 2006. Thermal stabilization of Bacillus subtilis family-11 xylanase by directed evolution. J Biol Chem 281(15):10236-42.

Muilu J, Törrönen A, Peräkylä M, Rouvinen J. 1998. Functional conformational changes of endo-1,4-xylanase II from Trichoderma reesei: a molecular dynamics study. Proteins 31(4):434-44.

Murakami MT, Arni RK, Vieira DS, Degrève L, Ruller R, Ward RJ. 2005. Correlation of temperature induced conformation change with optimum catalytic activity in the recombinant G/11 xylanase A from Bacillus subtilis strain 168 (1A1). FEBS Letters 579(28):6505-6510.

Nakazawa H, Okada K, Onodera T, Ogasawara W, Okada H, Morikawa Y. 2009. Directed evolution of endoglucanase III (Cel12A) from Trichoderma reesei. Applied Microbiology and Biotechnology 83(4):649.

Paës G, O'Donohue MJ. 2006. Engineering increased thermostability in the thermostable GH-11 xylanase from Thermobacillus xylanilyticus. J Biotechnol 125(3):338-50.

Paës G, Tran V, Takahashi M, Boukari I, O'Donohue MJ. 2007. New insights into the role of the thumb-like loop in GH-11 xylanases. Protein Eng Des Sel 20(1):15-23.

Park SH, Park HY, Sohng JK, Lee HC, Liou K, Yoon YJ, Kim BG. 2009. Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening. Biotechnol Bioeng 102(4):988-94.

Pollet A, Vandermarliere E, Lammertyn J, Strelkov SV, Delcour JA, Courtin CM. 2009. Crystallographic and activity-based evidence for thumb flexibility and its relevance in glycoside hydrolase family 11 xylanases. Proteins-Structure Function and Bioinformatics 77(2):395-403.

Purmonen M, Valjakka J, Takkinen K, Laitinen T, Rouvinen J. 2007. Molecular dynamics studies on the thermostability of family 11 xylanases. Protein Engineering, Design and Selection:gzm056.

Reetz MT, Carballeira JD. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nat. Protocols 2(4):891-903.

Ruller R, Deliberto L, Ferreira TL, Ward RJ. 2008. Thermostable variants of the recombinant xylanase A from Bacillus subtilis produced by directed evolution show reduced heat capacity changes. Proteins 70(4):1280-93.

Sabini E, Sulzenbacher G, Dauter M, Dauter Z, Jorgensen PL, Schulein M, Dupont C, Davies GJ, Wilson KS. 1999. Catalysis and specificity in enzymatic glycoside hydrolysis: a 2,5B

conformation for the glycosyl-enzyme intermediate revealed by the structure of the Bacillus agaradhaerens family 11 xylanase. Chem Biol 6(7):483-92.

Scheller HV, Ulvskov P. 2010. Hemicelluloses. Annual Review of Plant Biology 61(1):263-289.

Shallom D, Shoham Y. 2003. Microbial hemicellulases. Curr Opin Microbiol 6(3):219-28.

Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D. 2008. Determination of Structural Carbohydrates and Lignin in Biomass. National Renewable Energy Laboratory (NREL) Analytical Procedures:http://www.nrel.gov/biomass/analytical_procedures.html.

Song L, Dumon C, Siguier B, Kulminskaya A, Bozonnet S, O'Donohue MJ. Manuscript in preparation. Impact of a N-terminal extension on the stability and activity of the GH11 xylanase from *Thermobacillus xylanilyticus*.

Song L, Laguerre S, Dumon C, Bozonnet S, O'Donohue MJ. 2010. A high-throughput screening system for the evaluation of biomass-hydrolyzing glycoside hydrolases. Bioresource Technology 101(21):8237-8243.

Stephens DE, Rumbold K, Permaul K, Prior BA, Singh S. 2007. Directed evolution of the thermostable xylanase from Thermomyces lanuginosus. Journal of Biotechnology 127(3):348-354.

Stephens DE, Singh S, Permaul K. 2009. Error-prone PCR of a fungal xylanase for improvement of its alkaline and thermal stability. FEMS Microbiol Lett.

Törrönen A, Harkki A, Rouvinen J. 1994. Three-dimensional structure of endo-1,4-beta-xylanase II from Trichoderma reesei: two conformational states in the active site. EMBO J 13(11):2493-501.

Tu M, Pan X, Saddler JN. 2009. Adsorption of Cellulase on Cellulolytic Enzyme Lignin from Lodgepole Pine. Journal of Agricultural and Food Chemistry 57(17):7771-7778.

Vandermarliere E, Bourgois TM, Rombouts S, Van Campenhout S, Volckaert G, Strelkov SV, Delcour JA, Rabijns A, Courtin CM. 2008. Crystallographic analysis shows substrate binding at the -3 to +1 active-site subsites and at the surface of glycoside hydrolase family 11 endo-1,4-beta-xylanases. Biochem J 410(1):71-9.

Vieira DS, Degrève L, Ward RJ. 2009. Characterization of temperature dependent and substrate-binding cleft movements in Bacillus circulans family 11 xylanase: A molecular dynamics investigation. Biochimica et Biophysica Acta (BBA) - General Subjects 1790(10):1301-1306.

Walker JM, Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook: Humana Press. p 571-607.

Walsh C. 2001. Enabling the chemistry of life. Nature 409(6817):226-231.

Wyman CE. 2007. What is (and is not) vital to advancing cellulosic ethanol. Trends in Biotechnology 25(4):153.

You C, Huang Q, Xue H, Xu Y, Lu H. 2010. Potential hydrophobic interaction between two cysteines in interior hydrophobic region improves thermostability of a family 11 xylanase from Neocallimastix Patriciarum. Biotechnology and Bioengineering 105(5):861-870.

Zhang PYH, Himmel ME, Mielenz JR. 2006. Outlook for cellulase improvement: Screening and selection strategies. Biotechnology Advances 24(5):452-481.

Zhao HM, Giver L, Shao ZX, Affholter JA, Arnold FH. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nature Biotechnology 16(3):258-261.

Zhao HM, Zha WJ. 2006. In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP). Nature Protocols 1(4):1865-1871.

Zhu L, O'Dwyer JP, Chang VS, Granda CB, Holtzapple MT. 2008. Structural features affecting biomass enzymatic digestibility. Bioresource Technology 99(9):3817-3828.

Zilliox C, Debeire P. 1998. Hydrolysis of wheat straw by a thermostable endoxylanase: Adsorption and kinetic studies. Enzyme and Microbial Technology 22(1):58.

Zimmermann W. 1991. Degradation of Environmental Pollutants by Microorganisms and their Metalloenzymes. Sigel HSaA, editor. New York Marcel Dekker. 357–398 p.

II.6 Supplement Results

Table II-1S. Oligonucleotide primer pairs used for site-directed mutagenesis

Mutation	Primer sequence (5' $\rightarrow$ 3')
W109R and Y111H	CTATCACAGCCGGCGCCACAACGCACCGTCCATC
	GATGGACGGTGCGTTGTGGCGCCGGCTGTGATAG
Y6H	CCACGTACTGGCAGCATTGGACGGACGGC
	GCCGTCCGTCCAATGCTGCCAGTACGTGG
S27T	GCAACTACAGCGTAACCTGGAGCAACAGCGG
	CCGCTGTTGCTCCAGGTTACGCTGTAGTTGC

Table II-2S. Equivalent xylose yield (24 h) in the Tx-Xyl wild-type/mutant mediated hydrolyses of Dpl-WS (pH 5.8) and In-WS (pH 5.0 and 5.8).

Equivalent Xylose yield (g kg^{-1} biomass)						
Xylanase	Dpl-WS, pH 5.8		In-WS			
			pH 5.8		pH 5.0	
	μ	σ	μ	σ	μ	σ
wild-type	2.5	0.1	43.7	2.1	33.5	4.6
Y6H	3.2	0.6	45.1	0.4	ND	
S27T	5.5	0.4	49.1	1.5	37.8	2.0
Y111H	2.8	0.1	48.6	1.5	36.9	2.5
Y6H-Y111H	2.9	0.1	41.8	1.7	ND	
S27T-Y111H	3.6	0.6	51.3	1.2	39.5	1.2
Y6H-S27T-Y111H	3.2	0.3	45.0	1.4	ND	
Y3W	3.0	0.4	43.5	0.2	ND	
Y111S	3.2	0.5	49.7	2.2	36.1	3.4
Y111T	5.0	0.2	53.3	1.4	41.1	1.6

μ and σ: mean value and standard deviation of triplicate measurements; ND: not determined

Table II-3S. Equivalent xylose and glucose yields (24 h) in the In-WS hydrolyses mediated by the enzyme combination of Accellerase 1500 and Tx-Xyl wild-type/mutant.

Enzyme	xylose (g kg^{-1} biomass)		glucose (g kg^{-1} biomass)	
	μ	σ	μ	σ
wild-type+Accellerase	64.2	2.9	105.0	3.6
S27T+Accellerase	71.7	2.7	112.4	2.8
Y111H+Accellerase	72.0	3.0	110.8	4.3
S27T-Y111H+Accellerase	75.3	1.7	113.7	2.4
Y111S+Accellerase	73.5	2.1	111.4	1.8
Y111T+Accellerase	75.8	1.4	117.5	3.6
Accellerase	19.2	1.0	84.3	4.9

μ and σ: mean value and standard deviation of triplicate measurements

ARTICLE III

IMPACT OF AN N-TERMINAL EXTENSION ON THE STABILITY AND ACTIVITY OF THE GH11 XYLANASE FROM *THERMOBACILLUS XYLANILYTICUS*

The content of this article is relatively independent from Article I and II. First, we describe how to fuse two additional β-strands to the N-terminal of Tx-Xyl. For the resulting hybrid xylanase, the changes of enzymatic properties and activities on different substrates are studied and compared to parental Tx-Xyl. Finally, according to the modelled structure of hybrid xylanase, the potential roles of N-terminal region influencing substrate selectivity are hypothesized. In addition, it should be noted that the analysis data of sugar composition of soluble xylans (i.e. BWX and LVWAX) and wheat straw residues (i.e. In-WS and Dpl-WS) presented in this article were used in the whole thesis study.

Impact of an N-terminal extension on the stability and activity of the GH11 xylanase from Thermobacillus xylanilyticus

Letian Song[a,b,c], Claire Dumon[a,b,c], Béatrice Siguier[a,b,c], Isabelle André[a,b,c], Anna Kulminskaya[d], Sophie Bozonnet[a,b,c], Michael J. O'Donohue[a,b,c,*]

[a] *Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France*

[b] *INRA, UMR792, F-31400 Toulouse, France*

[c] *CNRS, UMR5504, F-31400 Toulouse, France*

[d] *Petersburg Nuclear Physics Institute, Russian Academy of Science, Molecular and Radiation Biology Division, Gatchina, 188350, St. Petersburg, Russia.*

Abstract

The 17 N-terminal amino acids of the GH11 xylanase from *Neocallimastix patriciarum* (Np-Xyl) have been grafted onto the N-terminal extremity of the unusually short GH11xylanase from *Thermobacillus xylanilyticus* (Tx-Xyl), creating a hybrid enzyme denoted Tx-Xyl-NTfus. Remarkably, this neo-xylanase was produced in *Escherichia coli* as a stable, soluble protein that displayed pH (6.2) and temperature optima (approximately 67°C) that were close to those of the parental Tx-Xyl. Measurement of the thermostability (at 60°C and 70°C) indicated that the hybrid xylanase Tx-Xyl-NTfus was less thermostable than Tx-Xyl, and this was confirmed by the measurement of its T_m value, which was 5°C lower than that of Tx-Xyl. Kinetic assays using two *o*NP-xylo-oligosaccharides (DP2 and 3) indicated that the longer N-terminal region of Tx-Xyl-NTfus does not procure more extensive substrate binding, but these experiments did provide evidence for a putative -3 subsite in both Tx-Xyl and the hybrid xylanase. The activities of Tx-Xyl-NTfus and parental Tx-Xyl on birchwood xylan were quite comparable, but the hybrid xylanase displayed higher activity on the more substituted low viscosity wheat arabinoxylan and, notably, on milled wheat straw. The combined action of Tx-Xyl-NTfus and the cellulolytic cocktail Accellerase 1500 on wheat straw procured yields of soluble glucose (24 % theoretical) and xylose (23 % theoretical) that were higher than those obtained when Tx-Xyl was combined with Accellerase 1500. Overall, this study revealed that the extension of the N-terminal region of Tx-Xyl leads to unexpected alterations in substrate selectivity and activity, which might be linked to alterations involving a hitherto unidentified secondary binding site.

Keywords

Abbreviation

A/X ratio: arabinose to xylose ratio ; BWX: birchwood xylan; Dpl-WS: xylanase-depleted wheat straw; H bond: hydrogen bonds; In-WS: intact (untreated) wheat straw; LVWAX: low viscosity wheat arabinoxylan; Np-Xyl: *Neocallimastix patriciarum* xylanase; *o*NP-X2: o-nitrophenyl-β-D-xylobioside; *o*NP-X3: o-nitrophenyl-β-D-xylotrioside; Tx-Xyl: *Thermobacillus xylanilyticus* xylanase; Tx-Xyl-NTfus: N-terminal fused hybrid xylanase; XOS: xylo-oligosaccharides.

III.1 Introduction

Endo-β-1,4-xylanases (EC 3.2.1.8) that group in family 11 (GH11) of the CAZy database are true xylanases, because they only hydrolyze β-1,4 xylosidic linkages in xylans (Berrin and Juge 2008; Collins et al. 2005). In the past, xylanases from family GH11 have been empirically selected for a wide variety of industrial applications, and are currently in use in sectors such as paper pulping and food and animal feed processing (Collins et al. 2005; Kulkarni et al. 1999). Likewise, now that biorefining has become a priority R&D target, the role of xylanases in the conversion of lignocellulosic biomass into platform intermediates, such as glucose and xylose, is increasingly recognized. Recent studies have underlined the importance of xylanases (and other hemicellulases) in cellulolytic cocktails for the hydrolysis of residual hemicelluloses, which remain in the biomass after pretreatment (Gao et al. 2010; Kumar and Wyman 2009b; Rémond et al. 2010; Wyman 2007), and used in integrated or consolidated biomass processing which require complete arsenals of highly active and efficient biomass degrading enzymes (Lynd et al., 2005). Therefore, in order to meet tomorrow's needs in terms of robust, efficient xylanases, it is necessary to acquire a better understanding of structure-function relationships, which is a prerequisite for enzyme design and engineering.

The secondary structure of GH11 xylanases is composed of two large anti-parallel β-pleated sheets and one α-helix. β-sheet A is composed of a maximum of six β-strands denoted A1 to A6, whereas β-sheet B is composed of nine β-strands, which are named B1 to B9. The N-terminal part of GH11 is variable in length. In cases in which the N-terminal region is shorter, β-strand A1 is absent, being replaced by a loop (Havukainen et al. 1996; Törrönen and Rouvinen 1995). Regarding tertiary structure, GH11 xylanases display compact β-jelly roll architecture, which is common to members of the CAZy clan GH-C (i.e. families GH11 and GH12). Conveniently, the overall structure of GH11 xylanases has been likened to a partially closed right hand, and accordingly the different structural elements have been identified using anthropomorphic terms such as fingers, palm and thumb (Törrönen et al. 1994). The fingers are formed by β-sheet A and part of β-sheet B, while the palm is made up of the unique α-helix and the twisted β-sheet B. The thumb corresponds to the loop that connects strands B7 and B8 and finally, the 'cord' is a long loop that joins B6 and B9 (Purmonen et al.

2007). The active site is located within a cleft that is the major feature of the palm. The cleft is partially surrounded by the fingers and the thumb, and is blocked at one extremity by the cord (Ludwiczek et al. 2007; Purmonen et al. 2007). The catalytic dyad, which is composed of two conserved glutamate residues, is located at the centre of the cleft and is surrounded by a number of other conserved amino acids that are important for substrate binding (Collins et al. 2005; Jeffries 1996).

The N-terminal region of GH11 xylanases, which is located on the opposite side of the protein with respect to the cord, is generally defined as the region that extends from the N-terminal amino acid through to β-strand B3, and hence includes five or six strands. Numerous studies have shown that the N-terminal region is related to enzyme stability and recent molecular dynamics data have indicated that unfolding initiates there (Hakulinen et al. 2003; Purmonen et al. 2007; Ruller et al. 2008). To improve the thermostability of various xylanases, several groups have focused on the N-terminal region, introducing therein disulphide bonds (Fenel et al. 2004; Paës and O'Donohue 2006) or arginine rich sequences (Sung 2007), or by replacing all (Shibuya et al. 2000; Sun et al. 2005) or part of the N-terminal sequence of mesophilic GH11 xylanases (Zhang et al. 2010) with that of a thermophilic counterpart. However, the role of the N-terminal region in xylanase activity has been the focus of much less attention, despite the fact that it forms part of the active site cleft. Nevertheless, having engineered two disulphide bonds into the GH11 xylanase from *Thermobacillus xylanilyticus* (denoted Tx-Xyl), with one located in the N-terminal region, Paës and O'Donohue (2006) reported that the resultant mutant enzyme released more soluble sugars from destarched wheat bran than the parental enzyme. Similarly, a study of the GH11 xylanase from *Neocallimastix patriciarum* (denoted Np-Xyl) has revealed that this enzyme, which displays an unusually extended N-terminal region (one of the longest reported among GH11 members), is characterized by elevated catalytic efficiency on oat spelt xylan (k_{cat}/K_M = 1400 s^{-1} mM^{-1}), which is tentatively attributed to the fact that the N-terminal extension provides the xylanase with an extra subsite (Vardakou et al. 2008).

Unlike Np-Xyl, the moderately thermostable Tx-Xyl is characterized by an exceptionally short N-terminal region, meaning that these two enzymes represent opposite extremes of family GH11 with respect to this structural feature. Therefore, in an attempt to gain some new

insight into the functional role of the N-terminal region of GH11 xylanases and to improve the catalytic potency of Tx-Xyl we have created a hybrid enzyme by adding the A1 and B1 strands of Np-Xyl to the N-terminal extremity of Tx-Xyl. In this report, we recount our findings, in particular with regard to enzymatic activity of the hybrid enzyme.

III.2 Materials and methods

III.2.1 Plasmid construction and bacterial strain

The recombinant plasmids used in the study were as follows: pECXYL-R2 contains the coding sequence of Tx-Xyl in vector pRSETa (Paës and O'Donohue 2006); pNP1 encodes the mature form of Np-Xyl in vector pET22b (Vardakou et al. 2008); the construction of expression vector pNPTX5 encodes hybrid xylanase Tx-Xyl-NTfus was described below. The *E. coli* strain JM109(DE3) (Stratagene, USA) was used for protein expression.

III.2.2 N-terminal fusion of Tx-Xyl xylanase

An 81 base-length oligonucleotide (primer PL1, shown below) including the coding sequence of A1 and B1 strands of Np-Xyl (shown in bold as below) was synthesized, based on *E. coli* codon preference (K12 type strain), using the online optimizer server (http://genomes.urv.es/OPTIMIZER/). The fusion work was achieved through a two-step PCR. In the first PCR, the reaction mixture (total volume of 50 µl) contained 10 ng of pECXYL-R2 as template, 250 µM of primers PL1 and P2 (shown below), 200 µM of each dNTP and 1 IU of Phusion polymerase (NEB Inc., USA). The amplification reaction was conducted using the following sequence: 1 cycle at 98°C for 1 min, 25 cycles of [98°C for 10 s, 50°C for 30 s and 72°C for 15 s] and finally 1 cycle of 10 min at 72°C.

PL1: 5'-

GTGGCAT<u>ATG</u>**GCCTTCACCGTTGGTAACGGTCAGAACCAGCACAAAGGTGTTA**

    ACGACGGTACCTACTGGCAGTATTGGAC-3'

PS1: 5'- GTGG*C*AT<u>ATG</u>GCCTTCACC-3'

P2: 5'- GGATCAAGCTTCGAATTCTTACC -3'

The second PCR, which amplified the final DNA product, used the products of the first PCR as template, with PS1 and P2 oligonucleotides as forward and reverse primers respectively (shown above). The same PCR reaction sequence as above was employed. The final PCR

product was purified using the QIAquick PCR purification kit (Qiagen, Germany), then double-digested by restriction enzymes *Nde*I and *EcoR*I (NEB Inc., USA) at 37°C. The insert was recovered from Sybr-safe agarose gel (Qiagen, Germany) and purified again, then cloned into the similarly digested pRSETa vector.

III.2.3 Growth condition and xylanase purification

JM109(DE3) cells bearing the recombinant plasmids were cultured and expressed in LB medium at 37°C as described previously (Paës and O'Donohue 2006). The proteins, Tx-Xyl and Tx-Xyl-NTfus, were purified using the routine two-step chromatographic procedure that employs Q-sepharose fast flow followed by Phenyl Sepharose (GE healthcare, USA) (Paës et al. 2007), working on an ÄKTA FPLC purification system (GE healthcare, Uppsala, Sweden). The purification of Np-Xyl was performed according to described previously (Vardakou et al. 2008). The purified enzymes were adjudged homogeneous by SDS-PAGE. 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 (Walker et al. 2005).

III.2.4 Thermoactivity and thermostability assays

Thermoactivity assays were performed at various temperatures, in the range of 50 – 75°C. For Tx-Xyl-NTfus, activity at 60°C was explored in a pH range from 5.0 to 7.5. All the experiments were performed in triplicate, if not indicated otherwise. Specific activity was determined on birchwood xylan (5 g l^{-1}, Sigma-Aldrich, USA) using the 3,5-dinitrosalicylic acid (DNS) method as previously described (Song et al. 2010).

Thermostability assays were performed by incubating a xylanase solution (100 mM) in 10 mM Tris-HCl, pH 8.0 buffer at 50°C, 60°C and 70°C for up to 6 h. At regular time intervals, residual xylanase activity was quantified at 60°C. Xylanase half life ($t_{1/2}$) was deduced by fitting the curve with the following equation: ln(residual activity) = kt where t is the time and k is the slope, and $t_{1/2} = k^{-1}$ln(0.5) (You et al. 2010).

III.2.5 Differential Scanning Fluorimetry (DSF)

CFX96 Real-Time PCR Detection System (Bio-Rad) was used as a thermal cycler and the fluorescence emission was detected using Texas Red channel (λ_{exc}= 560 - 590 nm, λ_{em} = 675 -

690 nm). The PCR plate containing the test samples (in triplicate) of 20 μl per well were prepared by mixing SYPRO® Orange (Invitrogen, final concentration 10X) with protein (6.75 μM) in 20 mM Tris-HCl, 100 mM NaCl, pH 8.0 buffer. Negative controls with either Sypro or proteins alone were analyzed in parallel. The DSF assay was conducted using a temperature ramp from 20°C to 99.6°C, with increments of 0.3°C every 3 seconds. The apparent melting temperature (T_m) was calculated using the Bio-Rad CFX Manager software.

III.2.6 Kinetic assay

Kinetic parameters were derived from reactions using birchwood xylan (BWX) and low viscosity wheat arabinoxylan (LVWAX, Megazyme Inc., Ireland) as substrates. Eight different concentrations were selected in the range of 0.5 – 12 g l^{-1} for each substrate. Initial velocities were measured by following the rates of appearance of reducing sugar with the DNS method (Miller 1959). The kinetic parameters (k_{cat} and apparent K_M) were calculated using SigmaPlot V10.0, with non linear regression using the algorithm one site saturation.

III.2.7 -3 subsite mapping

Two end-labeled xylo-oligosaccharides (XOS) – o-nitrophenyl-β-D-xylobioside (oNP-X2) and o-nitrophenyl-β-D-xylotrioside (oNP-X3) synthesized as previously described (Eneyskaya et al. 2003) were used to investigate substrate binding in the -3 subsites of Tx-Xyl and Tx-Xyl-NTfus. Using a previously described method (Matsui et al. 1991; Suganuma et al. 1978), the binding affinity of glycon subsites was calculated using equation (1). It was assumed that both Tx-Xyl and Tx-Xyl-NTfus release oNP as the main product from oNP-X2 and oNP-X3 (Pollet et al. 2010) and that the oNP group functionally substitutes for the xylosyl moiety that binds in subsite +1. Therefore, assuming i = 3 and n = 4, equation (1) can be transformed into equation (2), which is suitable to calculate the binding energy of -3 subsite.

$$\Delta G_{-i} = -\frac{RT}{4183}\ln\frac{\left(\frac{k_{cat}}{K_m}\times BCF(X_{n-i})\right)_{(X_n)}}{\left(\frac{k_{cat}}{K_m}\times BCF(X_{n-i})\right)_{(X_{n-1})}} \qquad (Kcal \cdot mol^{-1}) \qquad ---- (1)$$

$$\Delta G_{-3} = -\frac{RT}{4183}\ln\frac{\left(\frac{k_{cat}}{K_m}\right)_{(oNP-X3)}}{\left(\frac{k_{cat}}{K_m}\right)_{(oNP-X2)}} \qquad (Kcal \cdot mol^{-1}) \qquad --------------------- (2)$$

$\triangle G_i$: Gibbs free energy (binding energy) of subsite *i*

k_{cat}/K_M of X_n: specificity constant for reducing end-labeled XOS with DP n

BCF(X_{n-i}) of X_n: bond cleavage frequency for (n-*i*)-mer end-labeled product from all

hydrolysis products of DP n end-labeled XOS

R: universal gas constant (8.314 J mol^{-1} K^{-1})

T: Temperature, in °K (273.15 + degree Celsius)

The kinetic assays for the hydrolysis of *o*NP-X2 and *o*NP-X3 were determined at six different substrate concentrations in the range of 1 – 8 mM and 0.5 – 6 mM respectively. The final concentrations of both xylanases were 35 nM and 15 nM for reactions with oNP-X2 and oNP-X3 respectively. The hydrolysis reaction was conducted in a quartz cuvette at 60°C, in 500 µl of 50 mM NaOAc buffer, pH 5.8. Liberated *o*NP was measured by continuous detection (Cary 100 spectrophotometer, GE healthcare) at 380 nm and quantified using a standard curve of free oNP (0.1 – 1 mM). The k_{cat} and t K_M values were deduced by SigmaPlot V10.0.

III.2.8 Evaluation of xylanase-mediated hydrolysis on Dpl-WS and In-WS

Finely milled (0.5 mm diameter in average) wheat straw (*Triticum aestivum*, cv. Apache, France), denoted In-WS, was prepared as previously described (Song et al. 2010). Xylanase-depleted wheat straw (Dpl-WS) was prepared by incubating In-WS with Tx-Xyl (150 BWX U g^{-1} biomass) for 72 hours at 60°C in 50 mM sodium acetate buffer, pH5.8 (containing 0.02% NaN$_3$) . After, the solid residue was recovered by filtration (Whatman® No.4 filter paper), washed and dried at 45°C.

To measure xylanase activity using In-WS or Dpl-WS as substrates, a reaction mixture in 50 mM NaOAc buffer, pH 5.8 was prepared that contained 2 % (w/v) biomass, 0.1 % (w/v) bovine serum albumin (BSA), 0.02 % (w/v) NaN$_3$ and an aliquot (final concentration of 10 nmol enzyme g^{-1} biomass) of Tx-Xyl or Tx-Xyl-NTfus thereof. Hydrolysis was performed at 50°C for 24 h with continuous stirring (250 rpm) in a screwed-capped glass tube, and then arrested by heating at 95°C for 5 min. To analyze the combined effect of xylanase and cellulases on In-WS, reactions were conducted as described above, except that Accellerase 1500 (Genencor, Rochester, NY) (0.2 ml cocktail per g^{-1} biomass) was added to the reaction mixture and reactions were buffered at pH 5.0 (i.e. the optimal pH for the Accellerase 1500

cocktail). Control reactions (buffered at pH 5.0) using either Accellerase 1500 or xylanase alone were also performed. For the analysis of monosaccharides, the supernatants were adjusted to 2 M H_2SO_4 and incubated for 2 hours at 95°C. The carbohydrate content in hydrolysis samples was determined by high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) using a Dionex ICS 3000 chromatography system (Sunnyvale, CA, USA). Monosaccharides were separated on a Dionex CarboPac PA-1 column, working at a concentration of 4.5 mM NaOH and a flow rate of 1 ml min^{-1} at 30°C over 25 min. Determination of xylo-oligosaccharides (XOS) was achieved on a Dionex CarboPac PA-100 column thermostated at 30°C, using a linear gradient of NaOAc (5 to 85 mM NaOAc) over 30 min, in a 150 mM NaOH solution, at a flow rate of 1 ml min^{-1}. Appropriate standards such as L-arabinose, D-xylose, D-glucose and D-galactose and XOS displaying DP 2 to 6 between 2 to 25 mg l^{-1} were used to provide quantitative analyses.

The amount of monosaccharide or polymers released during hydrolysis was expressed as a percentage of theoretical yields according to the following equation:

$$\text{Conversion \%}_{tot.\,N} = \frac{\text{average solubilized } N}{\text{theoretical total } N} \times 100\% \ (w/w)$$

''N'' represents any monosaccharide (e.g. xylose, glucose) in the straw, and the ''theoretical total N'' is the total amount of sugar N in the straw which is derived from composition analysis.

III.2.9 Compositional analysis of substrates

The sugar composition of BWX, LVWAX, In-WS and Dpl-WS were performed according to the standard method proposed by NREL (Sluiter et al. 2008). Monosaccharide separation and quantification was performed using a Dionex CarboPac PA-1 column as described above.

III.2.10 Modelling of hybrid xylanase

A model of the Tx-Xyl-NTfus was constructed using the 3D structures coordinates of Tx-Xyl (Harris et al., 1997) and Np-Xyl (PDB code 2C1F) as starting points. The modelling work was carried out using the CFF91 force field implemented within the DISCOVER module of the Insight II software suite (Accelrys, San Diego, CA, USA). For minimization, the CFF91 cross terms, a harmonic bond potential, and a dielectric of 1.0 were used. An initial minimization with a restraint on the protein backbone was performed using a steep descent algorithm, followed by conjugated gradient minimization steps, until the maximum RMS value was less

than 0.1. In subsequent step the system was fully relaxed. Additionally, the Tx-Xyl crystal model was optimized using the same procedure. A model of the hybrid Tx-Xyl-NTfus in complex with the xylohexaose was constructed based on the previously developed Tx-Xyl-xylohexaose (Paës et al., 2007), and optimized using the same procedure described above. Structural visualization and superposition were performed using PyMol software (DeLano, 2002).

III.3 Results

III.3.1 Design and production of the N-terminal modified xylanase, Tx-Xyl-NTfus

Examination of an alignment of mature (i.e. without signal peptide) amino acid sequences of 516 GH11 xylanases (ID PF00457 in Pfam database), including those of Tx-Xyl and Np-Xyl sequences, clearly illustrates that both of these enzymes are atypical, with Tx-Xyl displaying an exceptionally short N-terminal region, while Np-Xyl possesses an extended N-terminal extremity (Figure III-1). Similarly, the superposition of the 3D structures of Tx-Xyl and Np-Xyl further highlights these variations and reveals that the N-terminal extension in Np-Xyl constitutes two extra N-terminal β-strands (A1 and B1) (Figure III-2.A).
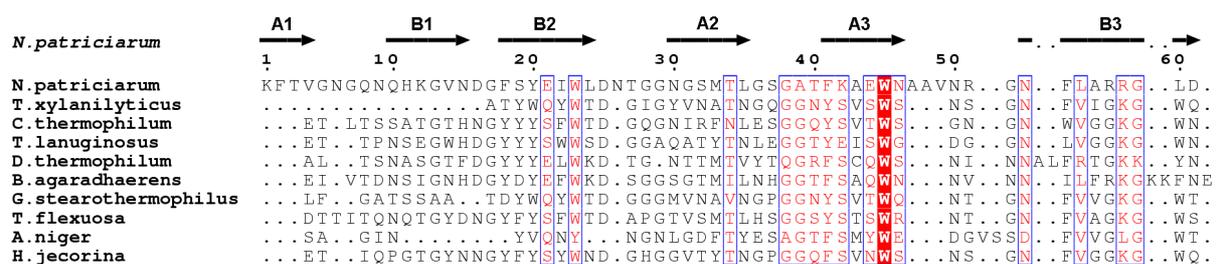
Figure III-1. Partial alignment of the N-terminal sequences of GH11 xylanases. All protein sequences were obtained from Swiss-Prot database of which the signal peptides are eliminated. The Np-Xyl and Tx-Xyl are represented by their organism names of *N. patriciarum* and *T. xylanilyticus* respectively. The positions of secondary structure elements are indicated on the top according to the structure of Np-Xyl (PDB 2C1F). The red colour letters represent highly conserved residues. Alignment is prepared by ESPript 2.2 utility (Gouet et al. 2003)

The structural superposition of Tx-Xyl and Np-Xyl was used to design a fusion protein, Tx-Xyl-NTfus that is composed of the Tx-Xyl amino acid sequence combined with the N-terminal extension from Np-Xyl, which represents approximately 17 amino acids (51 nucleotides). Briefly, the DNA sequence encoding the first 18 amino acids of Np-Xyl (Met1 to Gly18) was linked to the Tx-Xyl encoding DNA sequence, starting at codon 2 (encoding Thr2) (Figure III-2.B). Regarding the 51 bp DNA extension, compared to the wild type sequence two changes were made. First, in order to ensure the correct processing of the resultant recombinant fusion protein by the *E. coli* methionyl aminopeptidase, the second codon (encoding Lys2 in Np-Xyl) was altered to encode alanine and, second, the overall codon usage was optimized for expression in *E. coli*.

(A)

(B)

```
                    ββββ  LLLLL  βββββ  LLL  βββββββ
Np-Xyl/1       M KFTV GNGQN QHKGV NDG FSYEIWL 25
Tx-Xyl/1       - ---- ----- ----- --- A TYWQYWT 8
Tx-Xyl-NTfus/1 -  AFTV GNGQN QHKGV NDG TYWQYWT 24
```
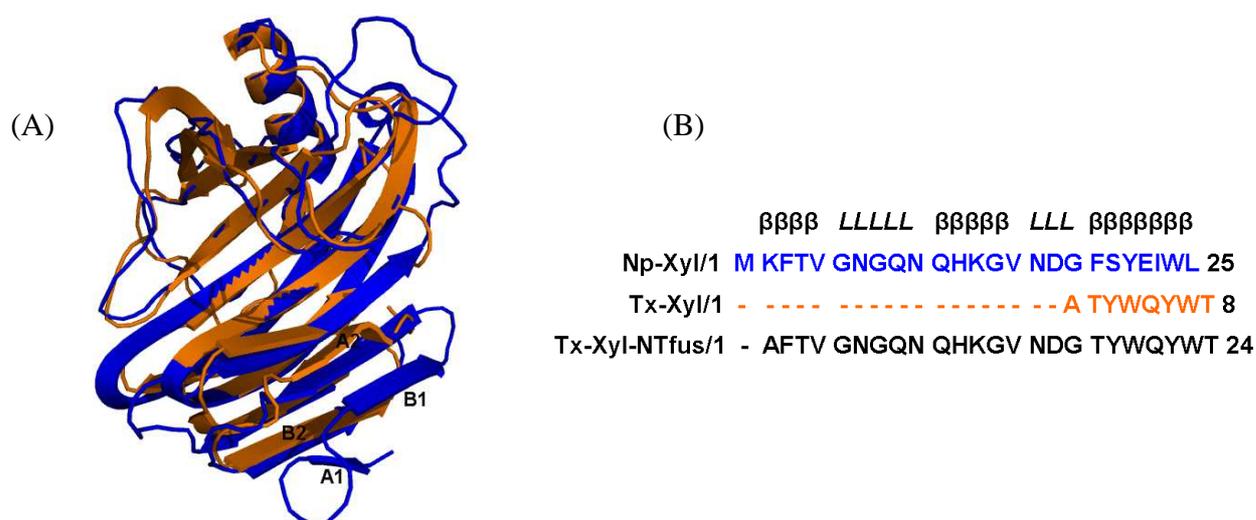
Figure III-2. **(A) Superposition of 3D structures of Tx-Xyl (orange) and of Np-Xyl (blue).** The four N-terminal β-strands from A1 to B2 are indicated, with only Np-Xyl showing A1 and B1 strands; **(B) Alignment of N-terminal sequences among Np-Xyl, Tx-Xyl, and Tx-XylNTfus.** Secondary structure motifs of Np-Xyl sequence are indicated, where β and *L* represent strands and loops respectively.

After construction, cloning and expression of Tx-Xyl-NTfus in *E. coli*, it was found that the protein was soluble and could be purified using the standard protocol, previously established for wild type Tx-Xyl. Analysis of the purified Tx-Xyl-NTfus by SDS-PAGE clearly indicated an increase in molecular weight compared to Tx-Xyl, indicating that the fusion was successful (Figure III-3).
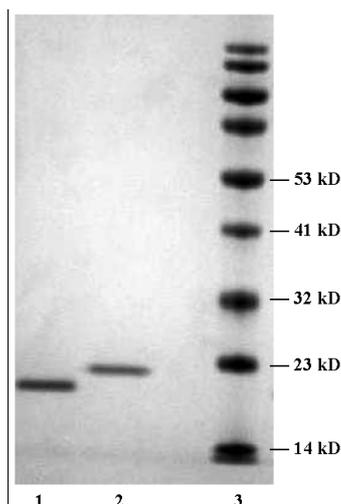
Figure III-3. SDS-PAGE gel electrophoresis of Tx-Xyl (*Mw* = 20.6 kD, lane 1), Tx-Xyl-NTfus (*Mw* = 22.3 kD, lane 2) and protein ladder (lane 3).

III.3.2 Characterization of Tx-Xyl-NTfus

The specific activity of Tx-Xyl-NTfus at 60°C on birchwood xylan was 1170 U/mg protein, which is approximately 80% that of Tx-Xyl (Table III-1). Measurement of the pH optimum of Tx-Xyl-NTfus revealed that this enzyme was optimally active at pH 6.2 and displayed quite stable activity (3.5% variation) over 0.8 pH units, from pH 5.8 – 6.5. Similarly, rather like the parental enzyme, Tx-Xyl-NTfus displayed good pH activity over the range pH 5.0 to 7.5, with residual activities being 90% (pH 5.0) and 85% (pH 7.5) of that measured at pH 6.2.

Table III-1. Enzyme characteristics of Tx-Xyl and Tx-XylNTfus.

Enzyme	ε (M^{-1} cm^{-1})	*Mw* (kDa)	SA (U mg^{-1})	pH$_{opt}$	T$_{opt}$ (°C)	T$_m$ (°C)	60°C t$_{1/2}$ (hour)	70°C t$_{1/2}$ (hour)
Tx-Xyl	102790	20.65	1450	5.8 - 6.0	~ 67	75.9	5.4	0.32
Tx-Xyl-NTfus	102790	22.30	1127	~ 6.2	~ 67	70.9	4.1	0.16

ε: Extinction coefficient; SA: specific activity measured on 5 g l^{-1} birchwood xylan at 60°C; T$_m$: melting temperature; t$_{1/2}$: half-life

Concerning thermoactivity, Tx-Xyl-NTfus displayed a temperature dependency that is very similar to that of Tx-Xyl, with a T$_{opt}$ °C situated at approximately 67°C (Figure III-4.A). Over the temperature range 50 – 60°C, activity increased progressively, with measured specific activity at 50°C being 60% of the value determined at 60°C. Above 60°C, the specific activity rose rapidly until highest specific activity was reached in the range 65 to 70°C. However,

above 70°C, activity was sharply decreased, with the decline being apparently faster than that observed for Tx-Xyl. Above 75°C, the activity of xylanase could not be measured, because of rapid inactivation.
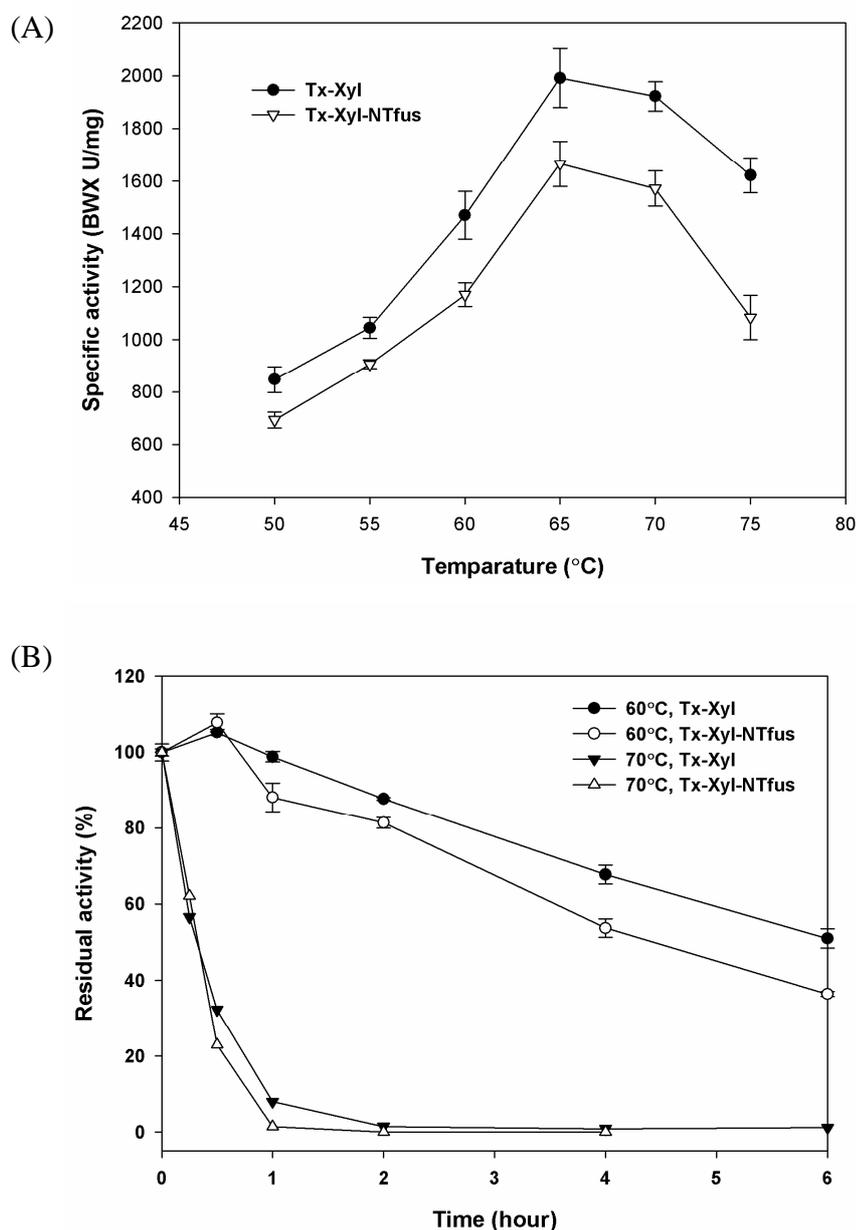
Figure III-4. Thermoactivity (A) and thermostability at 60°C and 70°C (B) of Tx-Xyl and Tx-Xyl-NTfus.

Regarding the thermostability of Tx-Xyl-NTfus and Tx-Xyl, both xylanases were highly stable at 50°C, retaining 100% activity after 6-hour incubation. At 60°C, the thermostability of both enzymes appeared to be concentration dependant (data not shown). At higher protein concentration (100 μM in stock solution), Tx-Xyl showed better thermostability than Tx-Xyl-

NTfus (Figure III-4.B), and the half-lives were determined as 5.4 and 4.1 hours respectively (Table III-1). At 70°C, both enzymes were subject to rapid inactivation (> 90% activity after 1 h) (Figure III-4.B), and half-lives were estimated to be 0.16 h (Tx-Xyl-NTfus) and 0.32 h (Tx-Xyl) respectively. Overall, Tx-Xyl-NTfus appeared to be less thermostable than Tx-Xyl. This postulate was corroborated by results from a protein melting experiment performed by differential scanning fluorimetry, which indicated that the T_m value of Tx-Xyl-NTfus was 70.9°C, while that of Tx-Xyl was 75.9°C (Table III-1).

Taking into account the different results, it was decided to conduct further experiments using a reaction temperature of 60°C, because this temperature represents a good compromise between thermoactivity and thermostability of Tx-XylNTfus.

III.3.3 Determination of kinetic parameters using soluble xylans

The activity of Tx-Xyl-NTfus was investigated using two soluble polymeric xylan substrates, whose composition is markedly distinct with respect to the amount of arabinose content (Table III-2). The compositional analysis of the birchwood xylan (BWX) used in this study did not reveal any arabinose after acid hydrolysis, whereas the low viscosity wheat arabinoxylan (LVWAX) is highly substituted by arabinose with an A/X ratio of 0.54 (Table III-2).

Table III-2. Sugar composition of the different xylanase substrates (% of dry matter, w/w).

Substrate	Glucose%	Xylose%	Arabinose%	Galactose%	A/X ratio
BWX	2.99±0.41%	97.01±0.41%	-	-	-
LVWAX	0.95±0.02%	64.13±0.42%	34.92±0.43%	-	0.543
In-WS	44.51±0.08%	26.16±0.14%	2.37±0.03%	0.44±0.06%	0.091
Dpl-WS	45.69±0.94%	21.92±0.17%	2.05±0.07%	0.46±0.05%	0.093

BWX: birchwood xylan; LVWAX: low viscosity wheat arabinoxylan; In-WS: intact wheat straw; Dpl-WS: xylanase-depleted wheat straw; A/X ratio: Arabinose to Xylose ratio

When acting on BWX, both Tx-Xyl and Tx-Xyl-NTfus displayed very similar Michaelis-Menten constants (k_{cat} and K_M app) (Table III-3). Compared to Np-Xyl, both enzymes displayed a lower turnover rate, but very similar values for K_M app. However, in the case of LVWAX, Tx-

Xyl-NTfus displayed a slightly higher turnover rate (14%) than Tx-Xyl, although the Michaelis constant remained highly similar (Table III-3).

Table III-3. Kinetic parameters of Tx-Xyl and Tx-Xyl-NTfus

	Tx-Xyl *			Tx-Xyl-NTfus *			Np-Xyl †		
	k_{cat} (s^{-1})‡	K_M (g.l^{-1})‡	k_{cat}/K_M (l.s^{-1}.g^{-1})‡	k_{cat} (s^{-1})‡	K_M (g.l^{-1})‡	k_{cat}/K_M (l.s^{-1}.g^{-1})‡	k_{cat} (s^{-1})	K_M (g.l^{-1})	k_{cat}/K_M (l.s^{-1}.g^{-1})
BWX	610.5±19.6	2.5±0.2	240.4	573.7±2.9	2.3±0.1	245.2	939.2±51.2	2.5±0.3	374.2
LVWAX	1699.4±95.9	5.1±0.1	333.1	1944.3±111.5	5.0±0.3	388.4		ND	
*o*NP-X2	74.4±3.4	10.1±0.2	7.4	51.1±0.1	8.7±0.5	5.9		ND	
*o*NP-X3	141.2±19.3	5.2±1.0	27.3	124.2±10.0	5.7±0.9	21.7		ND	

* Measured in 50mM NaOAc buffer, pH 5.8, at 60°C

† Measured in 50mM K_2HPO_4, 12mM citric acid, pH 6.5 at 37°C

‡ Units for polysaccharides are K_M in g.l^{-1} and k_{cat}/K_M in l.s^{-1}.g^{-1}, for *o*NP substrates units are expressed in mM for K_M and in s^{-1} mM^{-1} for k_{cat}/K_M.

*o*NP-X2: *o*-nitrophenyl-β-D-xylobioside; *o*NP-X3: *o*-nitrophenyl-β-D-xylotrioside; ND: not determined

III.3.4 Probing the presence of a -3 subsite in Tx-Xyl and Tx-Xyl-NTfus

To probe for extended substrate binding, two synthetic substrates, *o*NP-X2 and *o*NP-X3, were employed. For both Tx-Xyl and Tx-Xyl-NTfus, the specificity constant (k_{cat}/K_M) for oNP-X3 was 3.7-fold greater than for *o*NP-X2, though the overall performance of Tx-Xyl-NTfus was slightly inferior to that of Tx-Xyl (Table III-3). Calculation of the additional binding energy revealed a value of -0.86 kcal mol^{-1} for both xylanases, indicating a similar weak contribution of the -3 subsite.

III.3.5 Hydrolysis of wheat straw and synergy with commercial cellulases

Regarding the initial wheat straw samples, In-WS and Dpl-WS, compositional analysis (Table III-2) revealed that Dpl-WS was characterized by less xylose (4.3% dry weight) and arabinose (0.3% dry weight) than In-WS, but the Xyl/Ara ratio in both samples was highly similar.

To assess the catalytic potency of Tx-Xyl and Tx-Xyl-NTfus on In-WS and Dpl-WS, reactions were performed in two conditions (pH 5.0 or pH 5.8) using xylanases alone or in combination

with a commercial cellulase cocktail, Accellerase 1500 (Figure III-5). Analyses showed that Tx-Xyl-NTfus solubilised more xylose than Tx-Xyl in both pH conditions, and as expected neither of the xylanases released glucose. Tx-Xyl solubilised 12.8% $_{tot. xyl}$ and 16.7% $_{tot. xyl}$ of In-WS at pH 5.0 and 5.8 respectively, whereas Tx-Xyl-NTfus released 15.0% $_{tot. xyl}$ (pH 5.0) and 19.2% $_{tot. xyl}$ (pH 5.8) (Figure III-5). Monosaccharide sugar analysis of hydrolytic products (obtained for reactions performed at pH 5.8 only) revealed that Tx-Xyl-NTfus also released slightly more arabinose (Xyl/Ara = 8.97±0.13) than Tx-Xyl (Xyl/Ara = 9.70±0.21). Moreover, analysis of similar reactions using Dpl-WS as substrate showed that Tx-Xyl-NTfus released 29.1% more xylose yield than Tx-Xyl (Table III-4).

Table III-4. Yields of soluble xylose and glucose equivalents in xylanase-mediated hydrolyses of In-WS or Dpl-WS.

Enzymes employed in hydrolysis	Sugar yield (g.kg^{-1} biomass)					
	Dpl-WS		In-WS			
	pH 5.8		pH 5.8		pH 5.0	
	xylose*	glucose	xylose*	glucose	Xylose*	glucose
Tx-Xyl	2.5±0.1	0	43.7±2.1	0	33.5±4.6	0
Tx-Xyl-NTfus	3.2±0.1	0	50.1±3.8	0	39.2±3.7	0
Accellerase	ND		ND		19.2±1.1	84.3±4.7
Tx-Xyl + Accellerase	ND		ND		64.2±2.9	105.0±3.4
Tx-Xyl-NTfus + Accellerase	ND		ND		73.3±5.4	110.5±3.2

* The amount of xylose in hydrolysis product equals to the number of xylose detected in HPAEC-PAD plus the weight of equivalent xylose converted from detected xylo-oligosaccharides; ND: not determined

When an enzyme mixture of Tx-Xyl and Accellerase was used to hydrolyze In-WS, 24.5% $_{tot. xyl}$ and 23.6% $_{tot. glu}$ were released into the aqueous reaction medium (Figure III-5), indicating that both arabinoxylans and β-1,4 glucans had been targeted. Regarding glucose, this was always detected as a monosaccharide, probably because Accellerase 1500 displays β-glucosidase activity. When a similar experiment was performed using Tx-Xyl-NTfus, the amount of xylose and glucose that was solubilised increased, reaching values of 28.0% $_{tot. xyl}$ and 24.8% $_{tot. glu}$ respectively. In comparison, when Accellerase 1500 alone was used, 7.3% $_{tot. xyl}$ and 18.9% $_{tot. glu}$ were solubilised respectively.
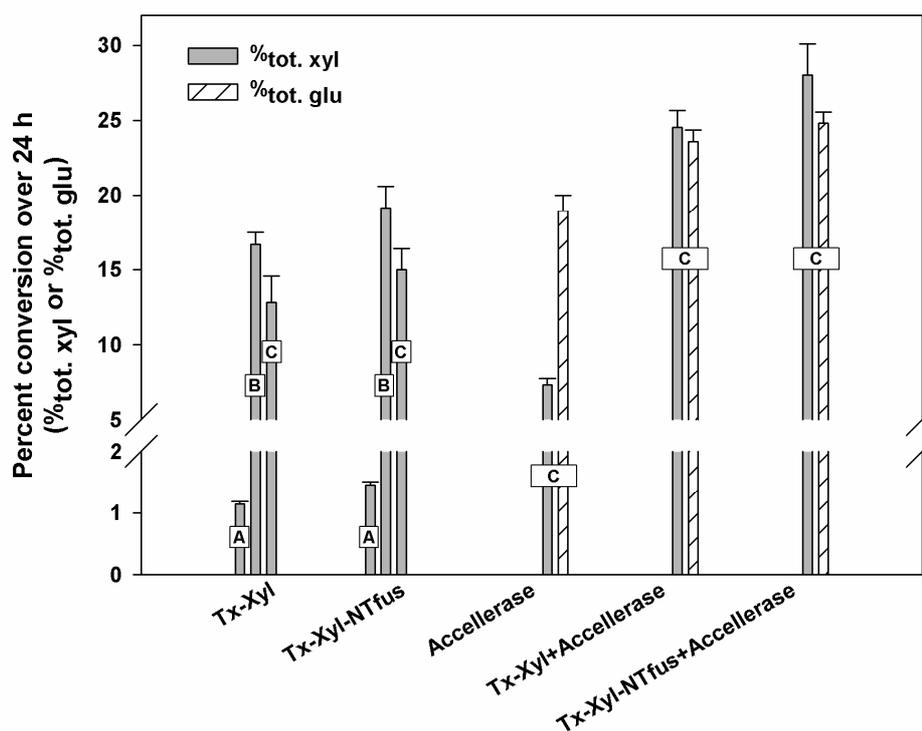
Figure III-5. Hydrolysis of In-WS and Dpl-WS using various enzyme combinations. The nature of the enzyme combination used is indicated on the X-axis. The letters A – C refer to reaction conditions: A is the hydrolysis of Dpl-WS at pH 5.8, B is the hydrolysis of In-WS at pH 5.8 and C is the hydrolysis of In-WS at pH 5.0.

II.3.6 Model of Tx-Xyl-NTfus

Since the 3D structures of Np-Xyl (PDB code: 2C1F) and Tx-Xyl (Harris et al., 1997)are available, it was possible to construct a 3D model of Tx-Xyl-NTfus. Briefly the β-strands A1 and B1 from Np-Xyl were excised *in silico* and pasted on to the N-terminal extremity of Tx-Xyl and residue 2 was mutated *in silico*, thus procuring a model of Tx-Xyl-NTfus. As expected, the energy minimized Tx-Xyl-NTfus was globally similar to that of the parental enzymes, with the N-terminal region displaying a conformation and position highly similar to that found in Np-Xyl (Figure III-6.A and B). Nevertheless, one major difference in the N-terminal region was the absence of β-strand B1, which was replaced by a long loop (denoted L1) (Figure III-6.A).
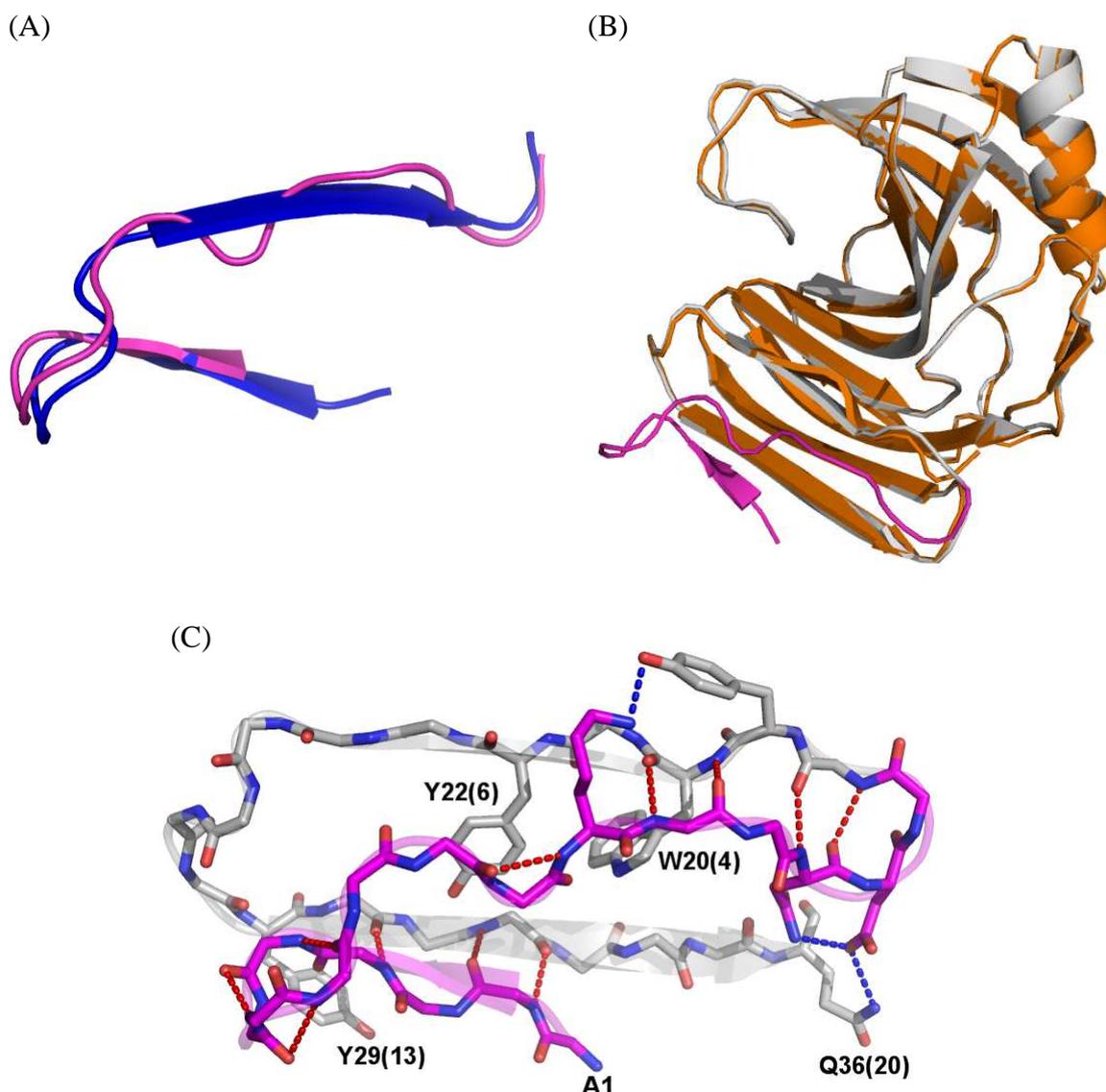
Figure III-6. 3D model of Tx-XylNTfus. (A) Superposition of the N-terminal regions of Tx-Xyl-NTfus (violet) and Np-Xyl (blue.). (B) Overall superposition of the 3D structures of Tx-Xyl-NTfus (grey) and Tx-Xyl (orange). The N-terminal extension in Tx-Xyl-NTfus is shown in violet. (C) Putative H-bonding in the N-terminal region (Ala1 – Gln36) of Tx-Xyl-NTfus. The H bonds formed by main chains and side chains are indicated by red and blue dashed lines respectively. The aromatic rings of Trp20(4) and Tyr22(6) (Tx-Xyl numbering in brackets) create steric hindrance in the vicinity of the loop L1. The Tyr22(6) and Tyr29(13) are identified as 'sticky patches' residues in original Tx-Xyl structure

Regarding β-strand A1 in Tx-Xyl-NTfus (from Phe2 to Val4), this appears to be suitably configured to allow the formation of three hydrogen (H) bonds with the anti-parallel β-strand A2. In contrast, only one H bond is feasible between the loop L1 and strand B2, mainly because of steric hindrances linked to the presence of Trp20 and Tyr22, which are not

present in Np-Xyl. Moreover, this lack of H-bonding explains why the region that composes L1 cannot adopt β-sheet conformation. However, three possible backbone H bonds were identified in the β-turn that links L1 to B2, in addition to the other three side chain H bonds that involve Lys12 - Tyr19, Asn15 - Gln36 and Asp16 - Gln36 (Figure III-6.C), these interactions strongly contribute to the stabilization of conformation adopted by L1.

III.4 Discussion

The rationale behind this work is based on the fact that the GH11 xylanase from *N. patriciarum* was previously reported to be highly active on oat spelt xylan. The authors suggested this elevated activity might be correlated to the presence of an extensive active site, comprised of six subsites, from -3 to +3, and significant substrate binding in the -3 subsite (-2.1 kcal/mol) (Vardakou et al. 2008). On the other hand, Tx-Xyl is an exceptionally short enzyme among GH11 members. Compared to the majority of GH11 enzymes, Tx-Xyl appears to be truncated at the N-terminal extremity and lacks at least one β-strand. Inevitably, this implies that the active site is shorter, and makes the existence of a -3 subsite uncertain (Paës et al. 2007). Therefore, to probe the functional consequences of this difference, an N-terminal extended version of Tx-Xyl has been created, using the N-terminal sequence of Np-Xyl.

The initial construction of Tx-Xyl-NTfus met with immediate success, leading us to assume that the addition of 17 amino acids was structurally coherent and that overall side-chain packing is adequate for stability. This result is gratifyingly surprising, because it is well-established that the N-terminal region of GH11 xylanases is a "hot spot" for protein unfolding (Hakulinen et al., 2003; Purmonen et al., 2007). Therefore, the fact that Tx-Xyl-NTfus is only moderately less stable than Tx-Xyl can be considered to be a success. Nevertheless, it is interesting to consider the factors that might be responsible for the loss of stability. Firstly, it is important to underline the fact that Np-Xyl is not thermostable, thus one might assume that the N-terminal region of this enzyme is not adapted to higher temperatures. Indeed, the examination of the 3D structure of Tx-Xyl-NTfus model reveals that β-strand A1 is only loosely associated with L1, and that the main interactions between these appear to be mediated by His11 (numbering of Tx-Xyl-Ntfus sequence), which can

potentially establish electrostatic interactions with Phe3 and Val5. However, there does not appear to be any extensive hydrophobic interactions, disulphide bonds or salt bridges, the latter being significant features in protein stabilization (Kumar et al., 2000; Vogt et al., 1997). In addition, in Tx-Xyl-NTfus, the newly introduced N-terminal region fails to adopt β-strand conformation and thus does not extend the original β-sheet, a factor that might contribute its lowered thermostability, because in GH11 xylanases the two rather rigid β-sheets tend to stabilize the protein structure (Hakulinen et al., 2003).

Previously, it was suggested that the thermostability of Tx-Xyl is protein concentration dependant and determined by inter-molecular hydrophobic protein-protein interactions, which mediate the formation of protein oligomers in solution (Harris et al., 1997). Based on crystallographic observations it was hypothesized that surface exposed aromatic side-chains form sticky patches, which allow pair-wise association of xylanase monomers. Among these residues, Tyr6 and 13 are located on β-strands B2 and A2, which correspond to Tyr22 and Tyr29 respectively in Tx-Xyl-NTfus. Clearly, in the 3D structure of Tx-Xyl-NTfus model, Tyr22 is completely buried inside the structure and Tyr29 presents an altered exposure (Figure III-6.C), and thus a modified ability to engage in thermostabilizing intermolecular hydrophobic interactions.

The enzyme engineering presented in this work failed in its primary objective. At the outset, it was thought that the addition of two extra β-strands to Tx-Xyl might create a -3 subsite or reinforce an existing one. However, this is apparently not the case, because both the wild type and the extended enzymes display a similar -3 subsite binding energy. Nevertheless, this study has revealed a moderately favourable energy for the putaive-3 subsite that results in higher activity when *o*NP-xylotriose is used as the substrate in place of *o*NP-xylobiose. Therefore, this represents the first experimental evidence for the existence of a weak -3 subsite in Tx-Xyl, which is coherent with our previous prediction based on substrate docking (Paës et al. 2007).

Presumably the failure to reinforce the -3 subsite is due to an imperfect mimicking of the Np-Xyl -3 subsite. According to Vardakou et al. (2008), three amino acid side chains (Gln11, Ile151 and Tyr194 might contribute to substrate binding in the distal glycon subsite via water

mediated hydrogen bonds. In Tx-Xyl-NTfus, the equivalent of these residues are Gln11, Ile132 (116) and Ser178(162) (original Tx-Xyl numbering in brackets). Ile 132 is located at the tip of the thumb and its position is not affected by the N-terminal fusion, and Gln11, contained within the N-terminal extension, is also correctly positioned in the model structure of Tx-Xyl-NTfus. However, Ser178, though capable of entering into a H bond network, is relatively distant (8.55 Å) from the endocyclic oxygen of the -3 xylosyl moiety, whereas to make a water-mediated contact the distance between these should be no greater than approximately 5.1 Å (assuming an average H bond length of 1.97 Å and a maximum dimension for H_2O of 1.18 Å). Encouragingly, the *in silico* mutation of Ser 178 to tyrosine reduces the overall distance to 5.5 Å, which suggests that a simple site-directed mutagenesis experiment could be sufficient to complete the creation of a -3 subsite. This experiment is currently underway.

Intriguingly, though a -3 subsite was not reinforced, the activity and selectivity of Tx-Xyl-NTfus was altered. Although both enzymes exhibited an almost equivalent $k_{cat}/K_{M\ app}$ value on birchwood xylan, on the more arabinose-substituted LVWAX, Tx-Xyl-NTfus revealed a 17% increase in catalytic efficiency compared to the parental enzyme. Moreover, on highly complex biomass such as wheat straw a similar difference was observed, since Tx-Xyl-NTfus released more soluble sugars that Tx-Xyl, indicating that it was better able to hydrolyze a wider range of arabinoxylans. Interestingly, this result recalls a previous observation made by Paës and O'Donohue (2006) that was linked to the introduction of disulphide bonds into Tx-Xyl. These authors reported that the modified enzyme, which notably contained a disulphide bond linking the N- and C- extremities, could better hydrolyze destarched wheat bran. Likewise, Moers *et al* (2007) have shown that mutation of the C-terminal Trp185 in the GH11 xylanase from *Bacillus circulans* alters specificity towards wheat water-unextractable arabinoxylans (Xyl/Ara = 1.96). More recently, Vandermarliere *et al* (2008) have linked this latter observation to the existence of a secondary substrate binding site, which could play a role in selectivity. Nevertheless, in the case of Tx-Xyl-NTfus, the careful comparison of its structure with that of Tx-Xyl indicates that the residues that are homologous to Trp185 of the *B. circulans* xylanase (i.e. Trp198 in Tx-Xyl-NTfus and Trp182 in Tx-Xyl) occupy identical positions, though the surrounding environment is different. In this respect, the most prominent change concerns a shallow groove that is located to one side of Trp182 in Tx-Xyl.

Accounting for the fact that homologous grooves in other GH11 xylanases (Ludwiczek et al., 2007) have been described as SBS, it is likely that the aforementioned groove in Tx-Xyl is also a SBS. In this regard, the fact that this groove becomes deeper and wider in Tx-Xyl-NTfus allows us to speculate that ligand binding in the SBS is improved Figure III-7. Furthermore, examination of the surface of Tx-Xyl-NTfus in the vicinity of Trp198(182) reveals an adjacent residue, Thr18(2) (Figure III-7), whose homologues in other GH11 xylanases have been designated as SBS determinants (Ludwiczek et al., 2007). In Tx-Xyl-NTfus, the proximity of Trp198 and Thr18 might also lead to greater binding in the SBS. Overall, we hypothesize that the presence of an N-terminal extension in Tx-Xyl-NTfus alters a putative SBS, which in turn improves ligand binding and alters substrate selectivity.
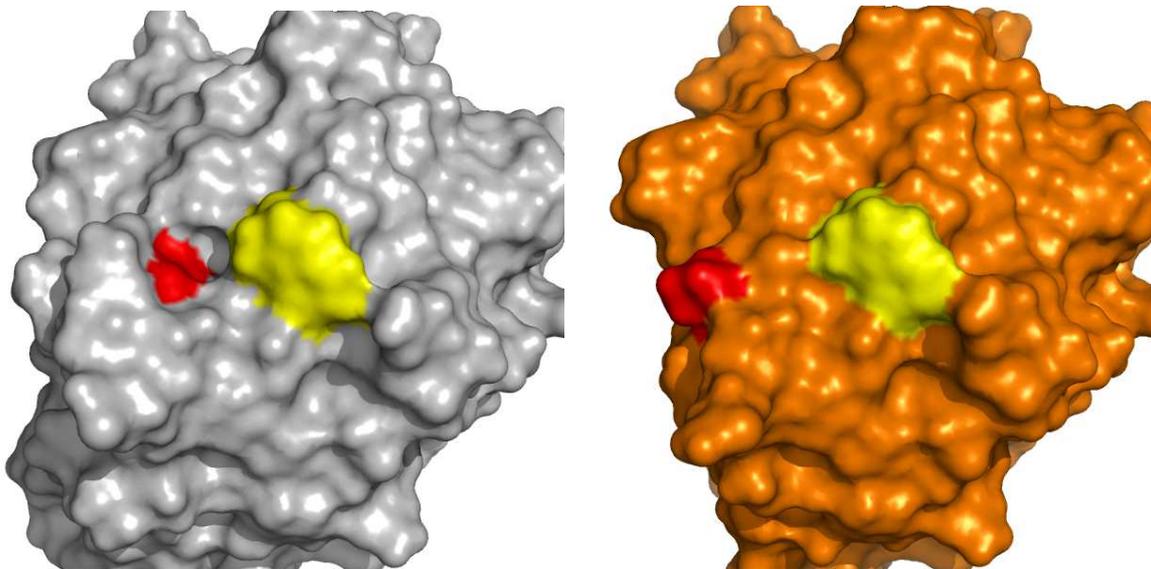
Figure III-7. Protein surface comparison between Tx-Xyl-NTfus (left, grey) and Tx-Xyl (right, orange). The relevant residues Trp198 (182) and Thr18 (2) are indicated in yellow and red respectively.

In conclusion, the straightforward addition of 17 amino acids to the N-terminal of Tx-Xyl did not lead to the creation of a more extensive catalytic cleft. However, this experiment clearly demonstrated that the creation of hybrid enzymes via the N-terminal addition of peptides capable of adopting a coherent structure is a viable method for the engineering of the properties of GH11 xylanases. Moreover, this work gives clues to how one might improve the activity of GH11 xylanases on complex substrates and increase synergy with cellulase cocktails. The next step in this strategy is to further engineer the distal glycon region of the

active site and optimize the hybrid enzyme, perhaps through the targeted optimization of the range between β strands A1 and B2.

III.5 References

Berrin JG, Juge N. 2008. Factors affecting xylanase functionality in the degradation of arabinoxylans. Biotechnol Lett 30(7):1139-50.

Collins T, Gerday C, Feller G. 2005. Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiology Reviews 29(1):3.

DeLano WL. 2002. The PyMOL Molecular Graphics System. Version Version 1.2r3pre. DeLano Scientific LLC, San Carlos, CA, USA. : http://www.pymol.org.

Eneyskaya EV, Brumer H, 3rd, Backinowsky LV, Ivanen DR, Kulminskaya AA, Shabalin KA, Neustroev KN. 2003. Enzymatic synthesis of beta-xylanase substrates: transglycosylation reactions of the beta-xylosidase from Aspergillus sp. Carbohydr Res 338(4):313-25.

Fenel F, Leisola M, Janis J, Turunen O. 2004. A de novo designed N-terminal disulphide bridge stabilizes the Trichoderma reesei endo-1,4-[beta]-xylanase II. Journal of Biotechnology 108(2):137-143.

Gao D, Chundawat SPS, Krishnan C, Balan V, Dale BE. 2010. Mixture optimization of six core glycosyl hydrolases for maximizing saccharification of ammonia fiber expansion (AFEX) pretreated corn stover. Bioresource Technology 101(8):2770-2781.

Gouet P, Robert X, Courcelle E. 2003. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Research 31(13):3320-3323.

Hakulinen N, Ossi T, Janne J, Matti L, Juha R. 2003. Three-dimensional structures of thermophilic β-1,4-xylanases from Chaetomium thermophilum and Nonomuraea flexuosa. European Journal of Biochemistry 270(7):1399-1412.

Harris GW, Pickersgill RW, Connerton I, Debeire P, Touzel JP, Breton C, Perez S. 1997. Structural basis of the properties of an industrially relevant thermophilic xylanase. Proteins 29(1):77-86.

Havukainen R, Törrönen A, Laitinen T, Rouvinen J. 1996. Covalent Binding of Three Epoxyalkyl Xylosides to the Active Site of endo-1,4-Xylanase II from Trichoderma reesei. Biochemistry 35(29):9617-9624.

Jeffries TW. 1996. Biochemistry and genetics of microbial xylanases. Current Opinion in Biotechnology 7(3):337-342.

Kulkarni N, Shendye A, Rao M. 1999. Molecular and biotechnological aspects of xylanases. FEMS Microbiol Rev 23(4):411-56.

Kumar R, Wyman CE. 2009b. Effect of xylanase supplementation of cellulase on digestion of corn stover solids prepared by leading pretreatment technologies. Bioresour Technol 100(18):4203-13.

Kumar S, Tsai CJ, Nussinov R. 2000. Factors enhancing protein thermostability. Protein Eng 13(3):179-91.

Ludwiczek ML, Heller M, Kantner T, McIntosh LP. 2007. A secondary xylan-binding site enhances the catalytic activity of a single-domain family 11 glycoside hydrolase. J Mol Biol 373(2):337-54.

Matsui I, Ishikawa K, Matsui E, Miyairi S, Fukui S, Honda K. 1991. Subsite Structure of Saccharomycopsis Î±-Amylase Secreted from Saccharomyces cerevisiae. Journal of Biochemistry 109(4):566-569.

Miller GL. 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Analytical Chemistry 31(3):426.

Moers K, Bourgois T, Rombouts S, Beliën T, Van Campenhout S, Volckaert G, Robben J, Brijs K, Delcour JA, Courtin CM. 2007. Alteration of Bacillus subtilis XynA endoxylanase substrate selectivity by site-directed mutagenesis. Enzyme and Microbial Technology 41(1-2):85.

Paës G, O'Donohue MJ. 2006. Engineering increased thermostability in the thermostable GH-11 xylanase from Thermobacillus xylanilyticus. J Biotechnol 125(3):338-50.

Paës G, Tran V, Takahashi M, Boukari I, O'Donohue MJ. 2007. New insights into the role of the thumb-like loop in GH-11 xylanases. Protein Eng Des Sel 20(1):15-23.

Pollet A, Lagaert S, Eneyskaya E, Kulminskaya A, Delcour JA, Courtin CM. 2010. Mutagenesis and subsite mapping underpin the importance for substrate specificity of the aglycon subsites of glycoside hydrolase family 11 xylanases. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics 1804(4):977.

Purmonen M, Valjakka J, Takkinen K, Laitinen T, Rouvinen J. 2007. Molecular dynamics studies on the thermostability of family 11 xylanases. Protein Engineering, Design and Selection:gzm056.

Rémond C, Aubry N, Cronier D, Noel S, Martel F, Roge B, Rakotoarivonina H, Debeire P, Chabbert B. 2010. Combination of ammonia and xylanase pretreatments: Impact on enzymatic xylan and cellulose recovery from wheat straw. Bioresource Technology 101(17):6712-6717.

Ruller R, Deliberto L, Ferreira TL, Ward RJ. 2008. Thermostable variants of the recombinant xylanase A from Bacillus subtilis produced by directed evolution show reduced heat capacity changes. Proteins 70(4):1280-93.

Shibuya H, Kaneko S, Hayashi K. 2000. Enhancement of the thermostability and hydrolytic activity of xylanase by random gene shuffling. Biochem J 349(Pt 2):651-6.

Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D. 2008. Determination of Structural Carbohydrates and Lignin in Biomass. National Renewable Energy Laboratory (NREL) Analytical Procedures:http://www.nrel.gov/biomass/analytical_procedures.html.

Song L, Laguerre S, Dumon C, Bozonnet S, O'Donohue MJ. 2010. A high-throughput screening system for the evaluation of biomass-hydrolyzing glycoside hydrolases. Bioresource Technology 101(21):8237-8243.

Suganuma T, Matsuno R, Ohnishi M, Hiromi K. 1978. A Study of the Mechanism of Action of Taka-Amylase A on Linear Oligosaccharides by Product Analysis and Computer Simulation. Journal of Biochemistry 84(2):293-316.

Sun J-Y, Liu M-Q, Xu Y-L, Xu Z-R, Pan L, Gao H. 2005. Improvement of the thermostability and catalytic activity of a mesophilic family 11 xylanase by N-terminus replacement. Protein Expression And Purification 42(1):122-130.

Sung W, L.; 2007. MODIFICATION OF XYLANASES TO INCREASE THERMOPHILICITY, THERMOSTABILITY AND ALKALOPHILICITY, Patent, Application No. PCT/CA2006/001192

Törrönen A, Harkki A, Rouvinen J. 1994. Three-dimensional structure of endo-1,4-beta-xylanase II from Trichoderma reesei: two conformational states in the active site. EMBO J 13(11):2493-501.

Törrönen A, Rouvinen J. 1995. Structural comparison of two major endo-1,4-xylanases from Trichoderma reesei. Biochemistry 34(3):847-56.

Vandermarliere E, Bourgois TM, Rombouts S, Van Campenhout S, Volckaert G, Strelkov SV, Delcour JA, Rabijns A, Courtin CM. 2008. Crystallographic analysis shows substrate binding at the -3 to +1 active-site subsites and at the surface of glycoside hydrolase family 11 endo-1,4-beta-xylanases. Biochem J 410(1):71-9.

Vardakou M, Dumon C, Murray JW, Christakopoulos P, Weiner DP, Juge N, Lewis RJ, Gilbert HJ, Flint JE. 2008. Understanding the structural basis for substrate and inhibitor recognition in eukaryotic GH11 xylanases. J Mol Biol 375(5):1293-305.

Vogt G, Woell S, Argos P. 1997. Protein thermal stability, hydrogen bonds, and ion pairs. J Mol Biol 269(4):631-43.

Walker JM, Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook: Humana Press. p 571-607.

Wyman CE. 2007. What is (and is not) vital to advancing cellulosic ethanol. Trends in Biotechnology 25(4):153.

You C, Huang Q, Xue H, Xu Y, Lu H. 2010. Potential hydrophobic interaction between two cysteines in interior hydrophobic region improves thermostability of a family 11 xylanase from Neocallimastix Patriciarum. Biotechnology and Bioengineering 105(5):861-870.

Zhang S, Zhang K, Chen X, Chu X, Sun F, Dong Z. 2010. Five mutations in N-terminus confer thermostability on mesophilic xylanase. Biochemical and Biophysical Research Communications 395:200-206.

GENERAL CONCLUSION AND

FUTURE WORK

This doctoral thesis deals with two key challenges, which are the rational design and the random engineering of catalytic properties in GH11 xylanases. From a fundamental point of view, this work aimed to provide original structure-function data that will serve in the future for more precise engineering. On the other hand, from an applicative point of view, a major part of this study was geared towards the engineering of useful properties for biorefining purposes into GH11 xylanases. In both cases, the results obtained are encouraging and a certain number of conclusions and perspectives can be drawn from these, thus forming a rich collection of knowledge that provides the basis for future experimentation.

How has this thesis contributed to state of the art in screening?

The discovery and engineering of biomass-degrading enzymes using techniques such as functional metagenomics or *in vitro* enzyme evolution, is fraught with difficulties. This is because biomass-degrading enzymes act on an incredibly complex substrate that is difficult to deploy in screening protocols. Nevertheless, in order to account for the truism ''you get what you screen for" (Schmidt-Dannert and Arnold, 1999), screening using biomass appears to be a necessity if one wishes to efficiently detect variants that display targeted properties on lignocellulosic material (Wilson, 2009; Zhang et al., 2006). From this point of view, in this doctoral work, we chose to embrace this challenge and develop a screen that would allow the detection of xylanase variants that display improved activity in wheat straw.

To develop a wheat straw-based screen was not simple, because this substrate is insoluble. Therefore, the first challenge was linked to reproducible substrate delivery into the wells of microtitre plates. Ideally, we wanted to avoid the purchase of sophisticated apparatus and procedures that involved pipetting slurries using manually prepared pipette tips. Fortuitously, early discussions with Dr. Jean-Marc Nicaud (CNRS-INRA Micalis), allowed us to identify a rather simple device, a Microscreen® Column Loader, which is designed to deliver powder into microtitre plates.

Overall the procedure that we devised is simple and robust, displaying a tolerably low error rate that provides a means to detect variants that display 15% more activity than the reference xylanase sample. The relative simplicity of the technique allowed a screening rate

of approximately 2000 clones per week, which is reasonable for medium sized enzyme engineering projects. Interestingly, the method is adaptable and so we expect that it can be employed in the future using other biomass sources and other enzymes, such as cellulases.

In this work, we only targeted a xylanase and we chose to probe the limits of enzyme engineering with regard to the improvement of a single enzyme on biomass. Nevertheless, the screening method that we developed could have been used in a multi-enzyme improvement strategy or in a single enzyme improvement strategy, but including the use of partner enzymes such as cellulases. In this case, cellulases could be easily added into the reaction mixture at the start of the hydrolysis reaction. Clearly, in the light of our results, this could be a promising avenue for future work.

Was the experimental strategy adapted towards the aim of finding more potent biomass-degrading xylanases?

To develop a xylanase able to solubilise more xylan from wheat straw, we used a combinatorial approach that associated both random mutagenesis, *in vitro* enzyme evolution and site-saturation mutagenesis. In the initial phase of this project, In-WS (intact wheat straw) was used as the substrate for two main reasons: (i) it was expected that enzyme engineering could accelerate the initial phase of the reaction and thus considerably reduce enzyme loadings and (ii) it was anticipated that certain variants would significantly increase the overall yield of hydrolysis. In fact, the analysis of mutants from the 1st and 2nd generation libraries revealed that neither aim was clearly attained, because instead we achieved increased protein expression (the case of Tx-Xyl-AF7) and just small increases in xylan solubilisation. Therefore, after the 2nd generation, we decided to use Dpl-WS assay, thus attempting to avoid the detection of variants that were simply better expressed and rather isolate high-performance mutants that go beyond the normal hydrolysis endpoint.

Overall, the analyses of our screening data revealed that the fitness of library clones increased in an incremental way from one generation to the next, in accordance with generic trends that have already been described for such experiments (Cirino et al., 2003; Zhao and Zha, 2006). Ultimately, nine of the best-performing mutants were identified, that

pinpointed four mutated positions, involving amino acids 3 (Y3W), 6 (Y6H), 27 (S27T) and 111 (Y111H, Y111S, Y111T). Compared to the wild-type Tx-Xyl, all of the mutants showed increased hydrolytic activity towards recalcitrant wheat straw (i.e. Dpl-WS), which was coherent with the screening assay. Therefore, overall we can consider that the strategy that was devised was appropriate and that the initial aim of the work was achieved. Nevertheless, even the best mutants, S27T and Y111T, only improved performance by a modest increment, since xylan solubilisation using In-WS as substrate only increased from 16.7%$_{tot. xyl}$ for Tx-Xyl to 18.6 – 20.4%$_{tot. xyl}$ for the mutant enzymes.

The reasons for the modest improvements observed in this work are ineluctably linked to the complexity of the substrate. When acting on wheat straw, Tx-Xyl is confronted with a multitude of challenges, including substrate accessibility, substrate chemical complexity and various types of inhibition mediated by different biomass components. Regarding accessibility, this is particularly flagrant when one considers the results obtained when reactions were performed in the presence of the cellulase cocktail, Accellerase 1500. Here, the action of the xylanases was enhanced by cellulase action and, reciprocally, cellulase action was enhanced by the presence of xylanase. This indicates that the degradation of biomass is a complex process that involves the 'peeling off' of alternating 'layers' of xylan and cellulose. In principle, the xylan that was solubilised in our experiments probably arose from the pith lining of the lumen and perhaps parenchyma cells, which possess mainly primary (unlignified) cell walls. Clearly, the more lignified structures were recalcitrant, which is unsurprising since no pretreatment (apart from milling) had been operated. To further investigate this, it would be interesting to perform ultra-structural studies aimed at specifically tracking the impact of xylanases on wheat straw. Ideally, to achieve this, one should identify a non-destructive technique that would allow semi-continuous assessment of xylanase action. In this respect, electron microscopy is unsuitable, because it requires complex sample preparation and can alter cell structures.

To go further with xylanase engineering, our results indicate that it will probably be beneficial to co-engineer several enzymes or, at least, perform screening in the presence of cellulases. For co-engineering of several enzymes, the actual number of genes to engineer could be limited by focusing, for example, on just a xylanase and an endoglucanase. The

other essential activities, such as cellobiohydrolase and glucosidase, could be simply incorporated into the hydrolysis reaction.

Were any new structure-function data obtained from the random mutagenesis / *in vitro* enzyme evolution experiments?

One of the interesting facts revealed by the analysis of our data was the independent nature of the mutations that were created. Although multiple mutations were obtained, in most cases these did not appear to provide additive effects, which is a little disappointing for a multi-generation experiment, where one would expect an accumulation of beneficial effects. Nevertheless from a purely analytical point of view, the lack of additivity simplifies the conclusions.

Does Tx-Xyl possess a SBS?

In recent years, two teams have provided evidence for secondary binding site (SBS) in three GH11 xylanases (Ludwiczek et al., 2007; Vandermarliere et al., 2008) from different microbial origins. In two xylanases from the evolutionarily-related species *B. circulans* and *B. subtilis*, the amino acids that define the SBS are highly conserved. Interestingly, the comparison of the sequence of Tx-Xyl with that of the *B. circulans* xylanase, reveals a global amino acid identity of 73% and, specifically regarding the SBS determinants, an identity of 81%. Only the two N-terminal SBS residues in *B. circulans* xylanase are absent in the Tx-Xyl sequence, basically because Tx-Xyl displays a shorter N-terminal region. Therefore, taking into account these observations, it is probable that Tx-Xyl also possesses a SBS, although so far this has not been shown experimentally.

Interestingly, in this work we have generated a mutation, S27T, whose improved hydrolytic performance might constitute a first piece of experimental evidence for a SBS in Tx-Xyl. Residue 27 lies on the putative SBS pathway and is probably one of the SBS determinants. The mutation of serine to threonine will logically increase local hydrophobicity, and thus the mutation S27T is probably beneficial for secondary xylan binding. To probe this hypothesis further, it could be interesting to perform gel retardation studies using xylan and an inactive variant of S27T.

What did we learn about thermostability?

Regarding thermostability, our results also provide some support for a previous hypothesis that links the thermostability of Tx-Xyl to the presence of numerous surface-exposed aromatic residues. According to this hypothesis, surface-exposed aromatic residues mediate intermolecular associations between xylanase monomers, thus producing oligomerization in solution. Among the surface-exposed residues that supposedly participate in the hydrophobic sticky patches, are amino acids 6 and 111, both of which were mutated in this study. As predicted, the mutation of Y6 and Y111 led to lower thermostability, although an exception to this was the mutation Y111T. Taken together, this work provides experimental support for the sticky patch hypothesis, although the isolated case of Y111T requires further consideration.

Intriguingly, the mutant S27T was actually more thermostable than the parental Tx-Xyl. It is gratifying to note that this is in perfect agreement with a trend among GH11 xylanases that was first underlined by Hakulinen *et al* (2003). Therefore, this result adds further credence to this postulate and could serve in future experiments, where one could perform a 'Threonine scanning' approach, creating a library where all of the different surface exposed serine residues are substituted one by one with threonine. Moreover, the link between surface-exposed threonine residues and improved thermostability might explain the behaviour of the mutant Y111T. In this case, the loss of an aromatic residue that participates in a sticky patch might be counterbalanced by an increase in the surface Thr/Ser ratio.

Do mutations at position 111 alter the movement of thumb loop?

The mutation of Tyr111 is particularly interesting, because this residue is located in the functionally important thumb loop, which is involved in both substrate binding and catalysis in GH11 xylanases. Being at the base of the loop, it has been proposed that residue 111 will have a strong impact on thumb movement, the precise details of which are still unclear. In our work, mutations that reduce spatial occupancy at position 111 were selected as improved biocatalysts, and we concluded that reduced spatial occupancy is probably the driver for smoother thumb movement and faster turnover rates. Interestingly, the alignment of 222 different GH11 sequences, including those of enzymes of both bacterial and fungal origin, reveals that the most frequent occupants of position 111 (or its equivalent in other

GH11 xylanases) are Tyr, Val and Thr (Fig. a), the latter being among the substitutions that we created.
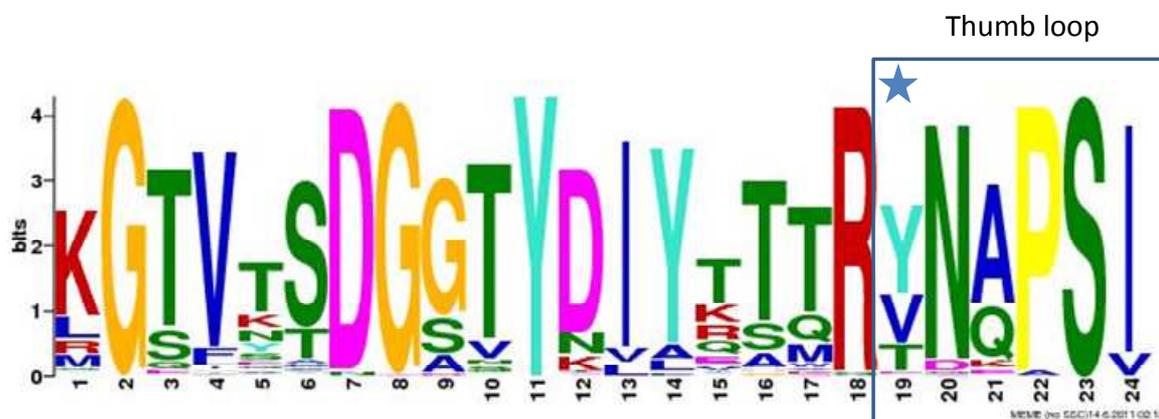
Fig. a: A MEME motif (generated at http://meme.nbcr.net/) showing the frequency of occurrence of various amino acids in GH11 xylanases. The motif is derived from an alignment of 222 GH11 xylanase sequences. The amino acids involved in the thumb loop are indicated by a blue box and Tyr111 (Tx-Xyl numbering) is marked with a blue star.

Beyond the fact that the mutation of Y111 could lead to a more efficient thumb movement, it is also interesting to speculate that the loss of an aromatic moiety at position 111 might lower the inhibition of Tx-Xyl by phenolic molecules and, perhaps, even lignin. Indeed, a recent study has shown that Tx-Xyl is inhibited by phenolic compounds that act in a non-competitive mode, which would be consistent with an interaction at the base of the thumb loop. Binding of aromatic molecules to the thumb could slow down or prevent its movement, thus preventing catalysis. To test such a hypothesis, it will be interesting to test the mutants Y111H, T and S for inhibition by phenolic compounds and perhaps attempt to ascertain their propensity for lignin binding.

Is it possible to create -3 subsite binding in Tx-Xyl?

In attempt to probe and/or, perhaps, create -3 subsite biding in Tx-Xyl, we have extended the N-terminal of Tx-Xyl with two β-strands, representing 17 amino acids, that come from

the GH11 xylanase produced by the fungus *N. patriciarum* xylanase (Np-Xyl). Although this experiment was technically straightforward and the approach quite routine, the result was nevertheless exceptional, because predicting the production of a stable and active hybrid enzyme is still a feat that currently remains out of the scientist's reach.

In contrast, regarding the aims of this work, the creation of hybrid Tx-Xyl-NTfus did not produce the expected improvement in -3 subsite binding. Indeed, Tx-Xyl-NTfus appeared to display similar binding to the wild type Tx-Xyl. However, this work did allow us to show that Tx-Xyl does weakly bind a xylosyl moiety in the -3 subsite, a fact that had, up until now, not been established experimentally. Moreover, in the light of our experimental results and analyses performed *a posteriori* via molecular modelling, it is possible to offer a rational explanation for the failure of the experiment. In Np-Xyl, it has been proposed that three amino acids, Gln 11, Ile 151 and Tyr 194, whose equivalents in Tx-Xyl-NTfus are Gln 11, Ile 132 and Ser 178, are involved in indirect interactions with the -3 xylosil group. Regarding Tyr 194 (Np-Xyl), this amino acid supposedly forms an indirect H bond with the endocyclic oxygen of the -3 xylosyl moiety via a water molecule. Examination of the 3D model of Tx-Xyl-NTfus reveals that Ser 178 is too far away to perform this role, because the distance that separates its hydroxyl group oxygen from the sugar endocyclic oxygen is 8.55 Å, whereas a water-mediated contact can be no longer than approximately 5.1 Å (assuming that the average H bond length is 1.97 Å and that the longest dimension of H_2O is 1.18 Å). Interestingly however, modelling also reveals that the replacement of Ser178 by a tyrosine would reduce this distance to about 5.5 Å, thus making a water-mediated H bond much more plausible. Therefore, this experiment should definitely be attempted in the near future in order to further investigate the possibility of increasing -3 subsite binding.

In conclusion, the establishment of a reliable and robust high-throughput screening approach using crude lignocellulosic biomass as substrate was the prerequisite to a vast series of combinatorial engineering experiments performed on *Thermobacillus xylanilyticus* GH11 xylanase. In combination with more rational approach, such as the design of a chimeric protein, our work led to mutant enzymes with improved activity towards the hydrolysis of recalcitrant AXs in wheat straw. Through this, we have gained useful insights into detailed structure-activity relationships for our xylanase, but yet, this work, depending

tightly on a quite straightforward protein engineering strategy, could not lead to overcome significantly the biomass architectural complexity. To reach this objective, future works will need to thoroughly investigate the function of 'hot spot' positions identified in this work, and to enlarge screening methodology via the development of multi-enzymatic approaches.

Cirino, P.C., Mayer, K.M., Umeno, D., 2003. Generating Mutant Libraries Using Error-Prone PCR. in: F.H. Arnold, G. Georgiou (Eds.), Directed Evolution Library Creation. Humana Press, pp. 3-9.

Hakulinen, N., Ossi, T., Janne, J., Matti, L., Juha, R., 2003. Three-dimensional structures of thermophilic β-1,4-xylanases from Chaetomium thermophilum and Nonomuraea flexuosa. *European Journal of Biochemistry*, 270, 1399-1412.

Ludwiczek, M.L., Heller, M., Kantner, T., McIntosh, L.P., 2007. A secondary xylan-binding site enhances the catalytic activity of a single-domain family 11 glycoside hydrolase. *J Mol Biol*, 373, 337-54.

Schmidt-Dannert, C., Arnold, F.H., 1999. Directed evolution of industrial enzymes. *Trends in Biotechnology*, 17, 135.

Vandermarliere, E., Bourgois, T.M., Rombouts, S., Van Campenhout, S., Volckaert, G., Strelkov, S.V., Delcour, J.A., Rabijns, A., Courtin, C.M., 2008. Crystallographic analysis shows substrate binding at the -3 to +1 active-site subsites and at the surface of glycoside hydrolase family 11 endo-1,4-beta-xylanases. *Biochem J*, 410, 71-9.

Wilson, D.B., 2009. Cellulases and biofuels. *Curr Opin Biotechnol*, 20, 295-9.

Zhang, P.Y.H., Himmel, M.E., Mielenz, J.R., 2006. Outlook for cellulase improvement: Screening and selection strategies. *Biotechnology Advances*, 24, 452-481.

Zhao, H.M., Zha, W.J., 2006. In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP). *Nature Protocols*, 1, 1865-1871.

ANNEX

ARTICLE MANUSCRIPT:

PROGRESS AND FUTURE PROSPECTS FOR PENTOSE-SPECIFIC BIOCATALYSTS IN BIOREFINING

Progress and future prospects for pentose-specific biocatalysts in biorefining

Claire Dumon[a,b,c], Letian Song[a,b,c], Sophie Bozonnet[a,b,c], Régis Fauré[a,b,c], Michael J. O'Donohue[a,b,c,*]

[a] *Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France*

[b] *INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France*

[c] *CNRS, UMR5504, F-31400 Toulouse, France*

running title : Pentose-specific biocatalysts

*corresponding author: michael.odonohue@insa-toulouse.fr

1 Abstract

2 Due to the fact that cellulose represents up to 60% of the dry weight of plant-derived biomass,

3 so far R&D efforts have mainly focused on the extraction and conversion of cellulose into

4 added value products. Consequently, abundant heteroxylans have been somewhat neglected in

5 current biomass-to-fuel concepts. The "glucocentric" approach to biorefining means that

6 extraction technologies are sub-optimal and non-specific with regard to pentose sugars, the

7 development of hemicellulases as biorefining enzymes has been slow and pentose-specific

8 conversion technologies for the production of value-added products are relatively scarce.

9 Nevertheless, xylan-related biocatalysis has continued to make steady progress in many areas,

10 including the discovery and characterization of a wide range of hemicellulases, which are

11 important enzymes for biomass hydrolysis. Similarly, the development of high-performance

12 ethanol producing yeast has focused for many years on the recruitment of pentose isomerases

13 or, alternatively, pentose reductases and pentitol dehydrogenases. However, similar efforts are

14 being made to develop microorganisms for alternative bioconversion processes that are

15 widening the range of chemicals that can be derived from pentoses. Finally, progress is also

16 being made in the area of glycosynthesis, which is opening new prospects for the use of

17 pentose sugars as building blocks for engineered pentosides, which will have quite different

18 applications, such as non-ionic surfactants or prebiotic food/feed ingredients. This review

19 provides an overview of these different development areas and discusses future prospects for

20 discovery and impact on biorefining.

21

22 Keywords: Pentoses; Enzymes; Biorefining; Bioconversion

23 **Introduction**

24 Although industrial, or white biotechnology, is well established in some specific industrial

25 sectors [1], the idea of its wider deployment, especially in the fields of fuels and chemicals, is

26 a recent one. Nevertheless, industrial biotechnology is now high on policy agendas, because it

27 is considered to be one of the most promising routes towards the hoped-for sustainability of

28 Man's future industrial activities. In the so-called bio-based economy, it is expected that

29 industrial biotechnology will play a key role in the development of biorefining, a term which

30 describes the integrated conversion of plant biomass into a wide array of fuels, chemicals and

31 materials [2-4]. Already, the production of first generation bioethanol is a bio-based industrial

32 process, relying on enzymes to produce glucose syrups from starch [5,6] and on yeast-

33 mediated fermentation of glucose for the production of ethanol.

34 Unlike starch-to-ethanol processing, which has benefited from the considerable experience

35 gained from the food industry, the biotechnological conversion of lignocellulosic biomass is a

36 fledging activity that is complicated by the intrinsic structural and chemical complexity of the

37 biomass feedstock and the diversity of the intermediate platform compounds, which include

38 D-glucose (and other hexoses), pentoses, lignins and derivatives thereof. Therefore, current

39 and future developments in industrial biotechnology need to account for this complexity,

40 developing a wide variety of tools, some of which should be versatile and others specific for

41 certain biomass components. Cellulose extraction and deconstruction have been the focus of

42 major research efforts over the last few decades, mainly because its component sugar, D-

43 glucose, is a sought after commodity that can be used in a wide variety of well-established

44 bioconversion processes. Nevertheless, in the economically important graminaceous plants

45 such as cereals and non-food crops (e.g. Switchgrass, Miscanthus), hemicelluloses are also

46 major components, representing up to 30% of the dry weight of harvested biomass [7-11].

47 However, less emphasis has been accorded to hemicelluloses, notably arabinoxylans, which

48   are highly abundant plant polysaccharides that yield D-xylose and, to a lesser extent, L-

49   arabinose as sugar intermediates. Although current uses for these sugars are limited, partly

50   due to the fact that high quality D-xylose and L-arabinose are not yet produced as commodity

51   chemicals, their future use as platform intermediates will be necessary in order to ensure the

52   sustainability of lignocellulosic biomass value chains and to avoid excessive non-food use of

53   D-glucose. Likewise, less attention has been paid to the development of pentose-specific

54   bioconversion processes aimed at the production of tailored products, despite the fact that the

55   transformation of pentoses is not only economically necessary, but useful in terms of product

56   diversification. Therefore, in this review we concentrate on the state of the art in this area,

57   first describing some of the key enzymes that are used for the extraction of pentose-based

58   polymers from biomass and then focusing on the current and future tools that can be used to

59   convert pentose sugars into useful products.

60   **Arabinoxylan-degrading enzymes**

61   Xylans, composed of β-1,4-linked D-xylosyl subunits, are the most abundant xylose-based

62   plant polymers and the second most abundant repository of biomass-based carbon after

63   cellulose. Xylans are widely distributed among dicotyledons and in grassy species, though

64   their amount (10-30% dry weight) and structures are variable from species to species, and

65   notably between these two major groups. In dicotyledons, xylans are generally decorated with

66   α-1,2-linked D-glucuronosyl and 4-*O*-methyl-D-glucuronosyl groups, whereas xylans from

67   grasses are frequently modified by α-1,2- and/or α-1,3-linked L-arabinosyl residues [10,12].

68   However, these descriptions are necessarily simplistic, because xylans are irregular structures

69   displaying a wide diversity of main-chain modifying residues (Fig. 1). Therefore, the array of

70   enzymes that are active on xylans is equally diverse.

71   In addition to the chemical and structural complexity described above, most heteroxylans,

72   especially those from dicotyledon secondary cell walls, are modified to various extents by

73   acetyl groups that are generally linked to main-chain D-xylosyl moieties via *O*-3 or *O*-2

74   linkages [7]. Likewise, ferulate esters are a major feature of grass heteroxylans [13].

75   However, in this case ferulic acid is ester-linked to side-chain L-arabinofuranosyl moieties,

76   commonly via the *O*-5 position. The presence of ferulic acid in heteroxylans is important for

77   overall cell wall integrity and resistance, because studies have revealed that in grass cell walls

78   ferulate esters crosslink cell wall heteroxylans through the formation of intermolecular

79   diferulic dimers [14-16]. Similarly, the oxidative crosslinking of ferulates to lignin contributes

80   to cell wall cohesion, covalently linking hemicelluloses to lignins.

81

82   *Xylanases and xylosidases*

83   Endo-1,4-β-xylanases (EC 3.2.1.8), or more commonly xylanases, randomly hydrolyze the

84   backbone of β-1,4-linked heteroxylans and produce xylo-oligosaccharides. Some rarer

85   xylanases possess endo-1,3-β-xylanase activity [17], catalyzing the hydrolysis of β-1,3-xylan,

86   which is typically found in marine algal species [18,19]. Examination of the CAZy

87   classification [20-23], which defines enzyme families based on structural and amino acid

88   sequence similarity, reveals that xylanases are mainly found in GH families 5, 8, 10 and 11

89   (Table 1). Nevertheless, the two major xylanase families are GH10 and 11 families, which

90   have been extensively investigated due to their industrial interest [24,25]. While GH10 and

91   GH11 enzymes share a similar retaining mechanism [26,27], they are quite different from a

92   structural point of view. A representative structure of GH10 members is that of the GH10

93   xylanase of *Thermoascus aurantiacus* [28]. This displays a catalytic domain that shows

94    $(\beta/\alpha)_8$-barrel architecture, which can be associated with other functional domains, such as

95 carbohydrate binding modules. The molecular weight of the catalytic domain of GH10

96 xylanases is usually ≥ 30 kDa and the pI value is often low [24,25]. In contrast, GH11

97 xylanases are characterized by a smaller catalytic domain (Mw ≥ 20 kDa) and higher pI

98 values. A large number of structures are available for GH11 members [29,30]. These show

99 that the catalytic domain of GH11 xylanases displays a typical β-jelly roll fold and an overall

100 structure that has been likened to a partially closed right hand [31].

101 Beyond differences in structure, pH optima or thermostability, it is important to note that

102 xylanases also display quite different substrate specificities [32]. This fact is well illustrated

103 by GH10 and 11 xylanases, enzymes that basically cleave xylosidic bonds using the same

104 catalytic mechanism and via a catalytic dyad composed of two glutamate residues. However,

105 GH10 xylanases are generally considered to have wider substrate selectivity than GH11

106 xylanases, especially regarding decorated xylo-oligosaccharides (Fig. 2). Studies of certain

107 family members has indicated that GH10 xylanases can accommodate α-1,2-subtituted D-

108 xylosyl residues at the +1 and −3 subsites, although this is much less likely to occur at the −1

109 and −2 subsites. Likewise, structural and experimental data indicate that GH10 xylanases are

110 likely to accommodate α-1,3-subtituted D-xylosyl residues at the -2, +1 and +2 subsites

111 [32,33]. Generally, this is not true in the case of GH11 xylanases, which show a clear

112 preference for less substituted regions, with α-1,2- and α-1,3-substituted D-xylosyl moieties

113 being excluded from the -2, -1, +1 (α-1,2) and −1 and +1 (α-1,3) subsites respectively [30].

114 As a consequence, GH10 xylanases tend to release shorter linear xylo-oligosaccharides (and

115 in some cases xylose) of the type XX, A^{3}X or U^{4m2}XX (nomenclature based on Fauré *et al.*,

116 http://axonym.cermav.cnrs.fr) [34], whereas GH11 xylanases will typically produce XX,

117 XXX, XA^{3}X, XA^{3}XX and XU^{4m2}XX [32,35-37].

118 Exo-1,4-β-xylosidases (EC 3.2.1.37), or generally speaking β-xylosidases, hydrolyze short

119 xylo-oligosaccharides and produce xylose, thus complementing the action of xylanases on

120   xylans. The CAZy classification reveals that β-xylosidases belong to families GH3, 30, 39, 43,

121   52, 54, 116 and 120 (Table 1). Regarding structures, models based on crystallographic data

122   are available for many of the families (except GH52, 116 and 120), although the known

123   structures do not necessarily relate to family members that posses β-xylosidase activity.

124   Families GH3, 30 and 39 are characterized by catalytic domains that display a $(\beta/\alpha)_8$ or TIM-

125   barrel fold, with GH30 and 39 belonging to the same clan GH-A [38,39]. The catalytic

126   domains of GH43 members display a β propeller fold [40], while those in GH54 possess a β-

127   sandwich, or β-jelly fold, that is reminiscent of the β jelly fold of enzymes in clan GH-B (i.e.

128   GH7 and GH16 members) [41]. It is noteworthy that most β-xylosidases operate via a

129   retaining mechanism, though GH43 enzymes are inverting enzymes and the mechanism

130   associated with GH120 members is not yet known. This information is important with regard

131   to technological applications, because some retaining enzymes have been shown to possess

132   marked transglycosylation activity, which can be disadvantageous when the goal is hydrolysis

133   and the production of monomeric syrups [42-44].

134   Regarding the technological deployment of xylanases and xylosidases, the former are widely

135   utilized in paper, food and feed industries [45] and both categories of enzymes are

136   increasingly recognized as being important components of enzyme cocktails for the

137   biorefining of lignocellulosic biomass [46,47]. This is particularly true for the refining of

138   cereal grain residues that are rich in hemicelluloses [48,49] and for lignocellulosic biomass

139   that has undergone pretreatment at high alkaline pH (e.g. AFEX, ARP etc). In this latter case,

140   much of the xylans remain in the pretreated fraction, which means that xylan-hydrolyzing

141   enzymes are required in cellulolytic cocktails in order to achieve optimal sugar yields [50-53].

142   Unfortunately, many commercially available cellulolytic cocktails lack adequate β-

143   glucosidase and xylanase activities [54]. Moreover, it is clear that simply boosting xylanase

144   activity in cocktails is not sufficient, because xylo-oligosaccharides produced by xylanases

145   appear to be strong inhibitors of cellulases [52,55]. Therefore, cocktails should also be

146   supplemented with robust xylosidase activity [53].

147

148   *Arabinofuranosidases*

149   α-L-Arabinofuranosidases (EC 3.2.1.55) are mainly exo-acting enzymes that release arabinose

150   through the cleavage of the α-1,2 and α-1,3, or α-1,5 bonds that link L-arabinosyl side-chain

151   decorations to the main chain of polysaccharides such as arabinoxylan, arabinogalactan, and

152   arabinan [47,56]. Those that act specifically on arabinoxylan are sometimes referred to as

153   arabinoxylan arabinofuranohydrolases (AXH) [57]. In the CAZy database, enzymes

154   displaying arabinofuranosidase activity are present in families GH3, 43, 51, 54 and 62 (Table

155   1). The enzymes in families GH3, 51 and 54 are retaining enzymes, while those in GH43 and

156   62 are inverting enzymes. Regarding linkage specificity, many arabinofuranosidases that act

157   on arabinoxylans appear to be active on either α-1,2, α-1,3, although preferences in some

158   enzymes have been observed [58,59].

159   The action of α-L-arabinofuranosidases (Abfs) is particularly important in the deconstruction

160   of biomass, because L-arabinosyl side chains can both hinder the action of xylanases [35] and

161   act as substrate specificity determinants [60]. Therefore, arabinofuranosidases can act in

162   synergy with xylanases, improving the release of xylose when compared to the action of

163   xylanases alone [61,62], although this is not a general rule for all arabinofuranosidases,

164   because some appear to be more active on soluble arabinose-substituted xylo-

165   oligosaccharides, which are the logical end products of xylanase activity [63].

166   Arabinoxylans are extremely complex polymers because of the intrinsic chemical variability

167   of side-chain groups, whose overall distribution along the xylan main-chain varies according

168   to botanical origin and location within plant tissues [12,64]. In arabinoxylans displaying a

169 high L-arabinose to D-xylose ratio the frequency of di-substituted D-xylosyl moieties bearing

170 L-arabinosyl groups at both the *O*-2 and *O*-3 positions is increased [65]. These disubtituted D-

171 xylosyl residues constitute a formidable barrier to biomass deconstruction, because most

172 arabinofuranosidases are unable to release L-arabinose from them. Arabinofuranosidases that

173 can release an arabinose from disubtituted D-xylosyl residues (designated AXH-d2 or -d3)

174 have been identified, though these are quite rare [66-69]. Currently, no simple screening

175 strategies are available to easily detect arabinofuranosidases that act on disubtituted D-xylosyl

176 residues and no structural data is available. Importantly, some studies have revealed synergy

177 between arabinofuranosidases from families GH43 and 51, especially between those that act

178 on disubtituted D-xylosyl residues and those that act on mono-substituted D-xylosyl moieties

179 [67,70]. When AXH-d2 or AXH-d3 act on arabinoxylan, they increase the amount of

180 monosubstituted D-xylosyl moieties, which are then substrates for the arabinofuranosidases

181 that act on monosubstituted D-xylosyl residues.

182 Bifunctional enzymes displaying arabinofuranosidase and xylosidase activities and possessing

183 a single catalytic domain are not uncommon in the literature. However, bifunctionality is

184 often defined using labile synthetic aryl glycosides as substrates and thus certain results

185 obtained are somewhat debateable [71]. Nevertheless, some true bifunctional enzymes have

186 been reported, which release arabinose from arabinoxylans and hydrolyse xylo-

187 oligosaccharides [72]. Such dual behaviour has been rationalized by Hövel et *al*. through the

188 structural investigation of a GH51 arabinofuranosidase from *Geobacillus stearothermophilus*.

189 Basically, D-xylopyranose and L-arabinofuranose share similar spatial conformation allowing

190 them to bind into the more promiscuous subsites of certain enzymes [73]. Multi-modular

191 enzymes possessing multiple catalytic domains also occur in Nature and engineering of such

192 chimeras has been experimented. Notably, Fan et *al.* have constructed a xylanase-

193 arabinofuranosidase that releases 30% more xylose from wheat arabinoxylan than a mixture

194 of the parental xylanase and arabinofuranosidase activities [74]. This result underlines the

195 importance of synergy between xylanase and arabinofuranosidases.

196

197 *Esterases*

198 Specialized variants of carboxylic acid esterases (EC 3.1.1.1) are frequently identified

199 components in microbial lignocellulolytic arsenals [75,76] and well-studied examples are

200 those produced by *Aspergillus* sp. [77-80]. Presumably, these enzymes act in concert with

201 glycoside hydrolases to deconstruct lignocellulosic structures. Like glycoside hydrolases,

202 carbohydrate esterases (EC 3.1.1.72) have been grouped into 8 of the 16 carbohydrate esterase

203 families within the CAZy classification system [81]. However, so far all feruloyl esterases

204 (EC 3.1.1.73) are grouped into family CE1 (Table 1), although a sub-classification of these,

205 based on substrate specificity and phylogenetic relationships, has been proposed by Crepin et

206 *al*. [82]. According to this scheme, which defines four subclasses A to D, feruloyl esterases of

207 the subclass A are quite specific for cereal-derived substrates, cleaving the ester bond that

208 links a feruloyl group to an L-arabinosyl moiety at its *O*-5 position. Enzymes from subclass B

209 are more generally associated with the deconstruction of arabinose-containing structures

210 within the hairy regions of pectin. Therein, they act on ester bonds that link feruloyl groups to

211 either the *C*-2 of feruloylated arabinosyl or the *C*-6 galactosyl moieties [78]. Subclass C

212 enzymes have been observed to act either on pectin or on cereal-derived xylans [83] and are

213 generally described as broad specificity enzymes, being active on all four methyl esters of

214 hydroxycinnamic acids (methyl ferulate, methyl sinapate, methyl *p*-coumarate and methyl

215 caffeate). Finally subclass D esterases, which are xylanase-like enzymes, also display broad

216 specificity on the four methyl esters of hydroxycinnamic acids and, like enzyme from

217 subclass A, have been reported to release 5,5'-diferulic acid dimers [82,84].

218 Many carbohydrate esterases are multi-modular enzymes, often being associated with one or

219 more carbohydrate binding modules and/or catalytic domains [85-87]. In the latter case, the

220 physical proximity of esterases with other catalytic modules favors synergy of action, which

221 is clearly beneficial with regard to the deconstruction of complex biomass-derived substrates

222 [88]. The usefulness of esterases in the deconstruction of heteroxylans has been demonstrated

223 [89]. Interestingly, it has been shown that GH11 xylanases appear to be better partners for

224 feruloyl esterases than GH10 xylanases, although these latter are better partners for the release

225 of 5,5'-ferulate dimers [90]. Regarding acetyl xylan esterases, it has also shown that these

226 enzymes can act in concert with cellulases and/or xylanases to achieve biomass

227 deconstruction [91,92], especially in substrates such as corn stover, where acetylation of

228 heteroxylans has been pinpointed as a source of resistance towards hydrolysis by xylanases

229 alone [93]. Agger et *al.* have investigated the necessity for both feruloyl and acetyl xylan

230 esterases in designed or minimal cocktails for the deconstruction of corn bran heteroxylans.

231 According to these authors, the use of acetyl xylan esterase on insoluble corn bran leads to the

232 release of acetyl groups, whereas feruloyl esterase appears to be more suited to the release of

233 ferulic acid from extracted, soluble arabinoxylans.

234

235 **Biocatalysts for the conversion of pentose sugars into fuels and chemicals**

236 *Enzyme recruitment for the development of ethanol producing microorganisms*

237 Over the last few decades, the development of microbial strains that can produce ethanol

238 using D-xylose as the carbon source has been a high priority R&D target [94,95]. The reasons

239 for this are two-fold. Firstly, although D-xylose-utilizing microorganisms are quite

240 widespread, commonly used strains of *Saccharomyces cerevisiae*, the preferred industrial

241 workhorse for bioconversion of glucose into ethanol, have been described as unable to utilize

242 D-xylose despite the fact that they can uptake D-xylose using hexose transporters [96].

243 Secondly, in the context of lignocellulosic biorefining, the ability to convert D-xylose into

244 ethanol, or other useful chemicals, is of considerable economic importance, given the high

245 abundance of this sugar in biomass.

246 To develop D-xylose-utilizing strains of *S. cerevisiae*, showing a potential ability to use

247 lignocellulosic hydrolysates, two approaches have been investigated and extensively reviewed

248 [94,97]. Both strategies aim to drive D-xylose into the pentose phosphate pathway (PPP) via

249 the conversion of D-xylose into its ketose derivative, D-xylulose. Once phosphorylated, D-

250 xylulose-5-phosphate can enter the central metabolism of *S. cerevisae* via the pentose

251 phosphate pathway [98]. The first approach relies on the recruitment of xylose isomerase

252 activity, which is absent in *S. cerevisiae*, whereas the second approach focuses on the

253 improvement of the two consecutive reduction and oxidation reactions, mediated by xylose

254 reductase (XR) and xylitol dehydrogenase (XDH) respectively (Fig. 3). For many years both

255 approaches have disappointed expectations, either because exogeneous xylose isomerases

256 have proved to be difficult to express in *S. cerevisiae*, or because attempts to up-regulate the

257 XR and XDH genes have either failed or have caused redox imbalances [95]. Nevertheless, in

258 recent years significant progress has been made, though redox imbalances are still an issue.

259 Regarding rare eukaryotic xylose isomerases, a breakthrough was achieved by Kuyper and

260 coworkers [99] who succeeded in expressing at a high level a xylose isomerase from

261 *Piromyces* sp. in *S. cerevisiae*. This success in itself was insufficient to improve the growth

262 rate of the engineered strain on D-xylose. However, subsequent metabolic engineering has led

263 to some remarkable progress [100-102]. Interestingly, as progress in these well-studied areas

264 is being made, it is becoming increasingly apparent that a wider system approach will be

265 needed to obtain highly efficient industrial strains [103]. This is illustrated by the recent

266 observation that while the transport of D-xylose is not an issue in strains of *S. cerevisiae* that

267   poorly utilize this sugar, it becomes a limiting factor in improved strains [104]. Even so, a

268   recent study has shown that this particular problem is surmountable, using engineered strains

269   that have been specifically improved with regard to their ability to uptake xylose [105].

270   Logically, because L-arabinose is less abundant than D-xylose, its conversion to ethanol has

271   received less attention [95]. Nevertheless, it's almost inevitable presence in lignocellulosic

272   hydrolysates has driven research aimed at recruiting L-arabinose-acting enzymes into *S.*

273   *cerevisiae*. Like D-xylose, L-arabinose can be introduced into the central metabolism of *S.*

274   *cerevisiae* via D-xylulose-5-phosphate. To convert L-arabinose into this intermediate, two

275   routes are possible. The first route, present in bacteria, involves the conversion of L-arabinose

276   into L-ribulose, then L-ribulose-5-phosphate and finally D-xylose-5-phosphate, using L-

277   arabinose isomerase, L-ribulokinase and then L-ribulose-5-phosphate 4-epimerase

278   respectively. The second route, characteristic of L-arabinose-utilizing fungi, consists of four

279   alternating reduction–oxidation reactions that sequentially employ L-arabinose reductase, L-

280   arabinitol dehydrogenase, L-xylulose reductase and xylitol dehydrogenase to provide D-

281   xylulose, which is then phosphorylated before entering the pentose phosphate pathway [106].

282   The introduction of the enzymes of either pathway into *S. cerevisiae* has met with limited

283   success. Sedlak and Ho reported that the presence of the *E. coli* AraBAD operon in *S.*

284   *cerevisiae* led to the accumulation of L-arabinitol, with no trace of ethanol being detected

285   [107]. Similarly, the assembly of a fungal-type pathway in *S. cerevisiae* provided a strain that

286   produced only low amounts of ethanol [106]. Nevertheless, a recent report has described the

287   successful cloning of two genes of eukaryotic origin (L-arabinitol dehydrogenase from

288   *Trichoderma reesei* and L-xylulose reductase from *Ambrosiozyma monospora*) into a

289   previously engineered *S. cerevisiae* strain that already produces L-xylulose reductase and

290   xylitol dehydrogenase. When the new strain was grown on L-arabinose as the sole carbon

291   source, a volumetric ethanol yield of approximately 41% was achieved [108].

292

Reducing carbon catabolic repression in butanol producing microorganisms

294 Butanol is thought to be a better liquid fuel than ethanol. Therefore, considerable efforts are

295 focusing on the bioproduction of this alcohol. Several microbial systems are being developed,

296 among these clostridia such *Clostridium acetobutylicum* and *Clostridium beijerinckii* [109].

297 Significantly, clostridia have already been used for the industrial production of solvents (ABE

298 process producing acetone, butanol and ethanol) using starch or molasses as substrate,

299 although the ABE process is currently uncompetitive [110]. Nevertheless, the fact that solvent

300 producing clostridia are able to catabolize pentose sugars could be the key to renewed

301 economic viability in the future. However, to achieve this goal, it will first be necessary to

302 alleviate the inhibition of pentose catabolism, a phenomenon that is observed when D-glucose

303 is present in the growth medium [111].

304 Like *Bacillus subtilis*, clostridial xylose utilization (xyl) operons appear to be controlled by

305 the transcriptional regulator, Xyl R and the CcpA-Hpr complex, which are both activated by

306 the presence of D-glucose [112]. Therefore, these elements are targets for engineering. In a

307 recent study, the CcpA-encoding gene in *C. acetobutylicum* was identified and disrupted

308 [109]. This led to a reduction in D-glucose catabolism and a modest increase in D-xylose

309 utilization. However, the consequent absence of carbon catabolic repression also produced

310 collateral effects that were detrimental to the mutant strain's performance. Likewise, results

311 obtained by Hu *et al*. indicate that the disruption of a putative *xylR* gene in *C. acetobutylicum*

312 increased the utilization of D-xylose compared to the wild type ATCC 824 strain. This

313 increase was accompanied by an increase in the expression of *xylB*, which encodes a xylulose

314 kinase [114].

315

316 *Biocatalysts for sugar alcohol production from pentoses*

317 Xylitol, a hydrogenated derivative of D-xylose, is one of a number of polyols that are used by

318 the food industry as sugar-free, low calorie sweeteners. In particular xylitol is used in chewing

319 gums and sweets, basically because it procures a fresh oral sensation, thanks to its negative

320 heat of solution and good solubility [115]. Other potential applications for xylitol include the

321 synthesis of polymers [116-118]. Currently, the world market for xylitol is approximately

322 160,000 T, though this is likely to rise in the future thanks to strong growth in the Asian

323 market and the development of new applications [119].

324 The production of xylitol has so far been dominated by a chemical dehydrogenation process

325 that uses a Raney Nickel catalyst [120,121]. However, several biotechnological routes exist

326 and certain strains, notably *Candida* sp., can achieve high yields [120]. In pentose-growing

327 yeasts or fungi, D-xylose is reduced to xylitol by a NADPH or NADH-dependant xylose

328 reductase (XR). Unfortunately, XR often displays quite promiscuous action, converting both

329 D-xylose to xylitol and L-arabinose to L-arabinitol [122]. This lack of specificity, which is also

330 an issue for chemical catalysis, is an obstacle to economic viability, because to achieve high

331 purity (particularly necessary for food applications) one must either use pure D-xylose as

332 substrate or implement costly downstream purification schemes [123]. Nevertheless, protein

333 engineering using a combinatorial approach has produced a variant of the *Neurospora crassa*

334 XR that displays diminished affinity for L-arabinose and an almost 7 fold increase in (k_{cat}/K_M)

335 D-xylose relative to that of L-arabinose [124]. More recently, Nair *et al*. have shown that the

336 combination of smart strain engineering of *Escherichia coli*, including the introduction of the

337 mutated XR, provides an efficient route towards the production of nearly 100% pure xylitol,

338 starting with a mixed carbon source containing D-xylose, D-glucose and L-arabinose [122].

339

340   *Opportunities for other chemicals*

341   In addition to the two well known catabolic pathways for pentoses, which both rely on the

342   conversion of D-xylose and L-arabinose into D-xylulose-5-phosphate, there is also a third non-

343   phosphorylative route that produces ketoglutarate or pyruvate and glycolaldehyde via L-/D-2-

344   keto-3-deoxypentanoate intermediates (Fig. 4). This pathway, which appears to be specific to

345   the bacterial and archaeal domains, employs a series of enzyme activities including pentose

346   dehydrogenase, pentolactonase, pentonic acid dehydratase and 2-keto-3-deoxypentanoate

347   dehydratase, which together catalyze reactions that give access to useful chemicals and

348   intermediates [125,126]. One such chemical is xylonic acid, for which a large volume,

349   industrial application has been proposed [127]. In a recent report, Toivari *et al*. have described

350   the production of xylonic acid by an engineered *S. cerevisiae* strain [128]. Alternatively,

351   previous studies have indicated that xylonic acid could also be produced *in vitro* using

352   commercially available enzymes such as glucose oxidase [129]. Both D-xylonic acid and L-

353   arabonic acid are also precursors of L-/D-1,2,4-butanetriol. The production of this energetic

354   compound has been achieved using engineered *E. coli* cells bearing enzymatic components of

355   the non-phosphorylative pathway described above. Here the conversion of D-xylonic acid and

356   L-arabonic acid into D-/L-2-keto-3-deoxypentanoic acids is followed by the synthesis of L-/D-

357   3,4-hydroxybutanal (catalyzed by benzoylformate decarboxylase), which is converted into L-

358   /D-1,2,4-butanetriol via the action of a dehydrogenase [130]. Finally, it is important to

359   underline the fact that the non-phosphorylative pathway provides a convenient route to

360   tricarboxylic acid (TCA) cycle-derived compounds, such as fumaric acid [131].

361   The quest for modified nucleosides bearing the unnatural L-configuration, is motivated by the

362   fact that these compounds are potent antiviral agents [132]. The chemical synthesis of L-

363   ribose and its derivatives has been demonstrated using L-arabinose as the starting material

364   [133]. However, the biotechnological production is perfectly feasible either by converting L-

365   ribitol into L-ribose via L-ribulose, or by converting L-arabinose into L-ribose via the same

366   intermediate. Regarding the first option, the bioconversion of L-ribitol into L-ribulose can be

367   achieved using acetic acid bacteria, such as *Gluconobacter oxydans* [134]. However, L-ribitol

368   is itself a rare sugar that is quite expensive. In contrast, it is to be expected that L-arabinose,

369   which is naturally abundant, will become more readily and cheaply available with the

370   development of the biorefining industry. Therefore, recent work focused on the bioconversion

371   of L-arabinose into L-ribose using engineered, ribulokinase-deficient *E. coli* and *Lactobacillus*

372   *plantarum* strains is noteworthy. Through the introduction of a ribulose isomerase-encoding

373   gene into these hosts, it was shown that L-ribulose produced by the action of the endogeneous

374   L-arabinose isomerase (encoded by *araA*) was efficiently converted into L-ribose [135-137].

375

376   *Hemicellulase-mediated glycosynthesis*

377   Over the last decades enzyme technology has increasingly penetrated organic chemistry,

378   especially for the synthesis of sugar-based compounds. This is because enzymes present

379   several advantages when compared to traditional catalysts: they display stereospecificity (i.e.

380   anomeric specificity), are often regioselective, perform reactions in mild conditions and can

381   eliminate laborious protection-deprotection steps that are usually required when dealing with

382   polyhydroxylated carbohydrates. As well as the widely used lipases, glycosyl transferases and

383   glycoside hydrolases are increasingly being used in glycosynthetic strategies. The latter are

384   particularly interesting because they can operate either via reverse hydrolysis (thermodynamic

385   control) or through transglycosylation (kinetic control) using activated substrates, or in less

386   frequent cases via their innate transglycosidase properties. Likewise, for the enzyme-

387   catalyzed synthesis of pentose-based compounds, α-L-arabinofuranosidases, β-xylosidases

388   and xylanases are all potentially useful. However, because of the prevalence of hexose in

389   Nature and the wide availability of hexose-specific enzymes, the development of other

390   enzymes for the synthesis of pentose-based compounds is lagging behind.

391   The possible applications for pentose-specific enzymes are quite diverse. These include the

392   synthesis or tailoring of xylo-oligosaccharides, especially L-arabinose-substituted xylo-

393   oligosaccharides, which have good potential as prebiotic molecules [138,139] and the

394   preparation of alkyl polypentosides, which are known to display interesting foaming and

395   degreasing properties [140,141].

396   Several xylanases exhibit the ability to transfer xylotriosyl donor groups (or higher DP xylo-

397   oligosides) to different acceptor molecules [142-149]. In certain cases, high concentration of

398   substrate leads to reverse hydrolysis that produces oligosaccharides of higher degree of

399   polymerization compared to the starting substrates. Likewise, β-xylosidases from *Bacillus*

400   *halodurans* [42,150] and from *Thermoanaerobacterium* sp. [151] have been shown to possess

401   similar activity when xylobiose concentration is elevated. Other xylanases have also proved to

402   be good tools for the preparation of hetero-xylosyl derivatives, such as alkyl xylosides, using

403   birchwood or oat spelt xylan and alcohols as starting materials [152-155]. Similarly, the direct

404   synthesis of alkyl chain xylosides from heteroxylans has also been achieved using a

405   combination of xylanase and xylosidase [44,156]. Nevertheless, these reactions are often

406   limited by the fact that product yield decreases as a function of (i) increasing chain length of

407   the alcohol acceptor molecule or (ii) increasing number of substituting groups linked to the

408   alcohol-bearing carbon, with primary alcohols usually being preferred acceptors

409   [152,153,155]. However, one can expect that protein engineering could provide the means to

410   surmount these limits. In some cases transxylosylation of polyphenols, such as catechols, or

411   rare sugars, such as D-psicose, have also been reported [146-148]. The use of activated aryl-

412   xyloside donors is a widely employed approach to achieve transxylosylation by xylanases and

413   β-xylosidases [42,160,161]. Using this strategy, Eneyskaya et *al*. have been able to devise an

414   efficient strategy for the synthesis of a range of *p*-nitrophenyl or 4-methylumbelliferyl β-

415   (1,4)-D-xylo-oligosaccharides [162,163] that are useful substrates for the analysis of xylanase

416   activity.

417   Although most developments in the area of pentose-based, enzyme-mediated glycosynthesis

418   have focused on D-xylopyranose-acting enzymes some reports have described the

419   glycosynthetic activity of L-arabinose-acting enzymes. The first of these employed the α-L-

420   arabinofuranosidase produced by *Thermobacillus xylanilyticus*. This enzyme is able to

421   catalyse the transarabinofuranosylation of alcohols using aryl L-arabinofuranosides as donors

422   and can synthesize both L-arabinofurano- and D-galactofurano-oligosaccharides [164-166].

423   The analysis of other α-L-arabinofuranosidases has revealed similar activities using various

424   acceptors [167-169].

425   One major drawback when using glycoside hydrolases for glycosynthetic purposes is that the

426   enzyme tends to hydrolyze the glycosynthetic products, thus leading to poor yields of the

427   synthetic coupling products. One solution to this problem is the creation of a glycosynthase,

428   which is hydrolytically incompetent derivative of the parental enzyme created using site-

429   directed mutagenesis [170]. In the presence of activated sugars (e.g. glycosyl fluorides) that

430   mimic the glycosyl-enzyme covalent intermediate, such enzymes can promote the formation

431   of glycosidic bonds while remaining inert with regard to the synthetic product [171,172]. This

432   approach has been applied to xylanases from *Cellulomonas fimi* [173,174] and to a β-

433   xylosidase from *Geobacillus stearothermophilus* [175], all of which were engineered via the

434   substitution of the active nucleophile residue by an unreactive amino acid. However, so far

435   the glycosynthase approach has not been applied to arabinofuranose-acting enzymes.

436

437

438 **Discovery of new pentose-specific enzymes – developments and perspectives**

439 In recent years the term metagenomics, which describes the culture-independent genomic

440 analysis of microbial populations, has become familiar to the wider scientific community.

441 This is because metagenomics is rapidly opening up access to the extensive chemistry present

442 in microbial diversity [176,177]. The ever wider application of this technology to the study of

443 biodiversity in many different ecosystems, including the sea [178], soil [179], the human gut

444 [180-183], the bovine rumen [184] and the termite hindgut [185], is revealing a vast and

445 hitherto hidden diversity of biocatalysts.

446 To discover new enzymes or even novel metabolic pathways, two fundamental types of

447 strategies are being developed, although many elegant variations on these approaches now

448 exist. Sequenced-based discovery of enzymes requires a minimum amount of knowledge

449 concerning the targeted enzymes and quite intensive sequencing resources, although "next

450 generation" technologies such as 454 pyrosequencing and, even more significantly, HiSeq

451 have considerably decreased the time and cost of sequencing [186] and makes possible the

452 production of huge amounts of sequence data, which nevertheless needs to be assembled,

453 annotated and appropriately analyzed [178,180]. Function-based discovery relies on good

454 knowledge of the chemical reaction under study, which is necessary to elaborate a selection or

455 screening strategy [187,188]. Both sequence-based and function-based strategies can benefit

456 from enrichment steps in order to home-in on functional microorganisms in a microbiome

457 [189-191].

458 Hydrolases are widespread in Nature and are key enzymes in catabolic processes breaking

459 down molecules such as polysaccharides to release energy. For this reason they are prime

460 targets for enzyme discovery, particularly now that biorefining of plant biomass has become a

461 major R&D pursuit. In this context, metagenomics has already proved to be extremely

462 powerful, especially to tap into microbial diversity in ecosystems that are expected to be rich

463 in biomass hydrolyzing activities. A milestone in this area is the work performed by

464 Warnecke *et al.*, who performed a first large scale metagenomic analysis on the microbiome

465 of the hindgut of a wood-feeding higher termite [185]. The study led to the identification of

466 more than 700 glycoside hydrolase-encoding genes, which represent 45 different CAZy

467 families. More than 100 genes were related to lignocellulose hydrolysis, with (putative)

468 pentose-acting enzymes from families GH10, GH11, GH26, GH43 and GH51 being

469 prominent among these. Likewise, genes encoding CE4 and CE34 and xylan esterases were

470 also identified. Similarly, more recent studies that have employed an initial enrichment step

471 aimed at specializing the microbial community prior to the metagenomic study, have shown

472 that compost microbial communities exposed to lignocellulosic biomass share common

473 features with microbiomes of wood-feeding termites and bovine rumen [192,193]. In

474 particular, cellulose degradation functions represented over 0.5 % of the annotated genes and

475 a high proportion of hemicellulases (GH51, 62, 67 and 43) were identified. Finally, a very

476 recent large-scale study (268 Gb sequence generated), also focusing on a bovine rumen

477 microbial community modified by enrichment on switchgrass, revealed 27,755 putative

478 glycoside hydrolase-encoding genes, of which only 5% were more than 75% identical to

479 sequences currently present in the CAZy database, thus underlining the power of the

480 metagenomic approach for novel enzyme discovery [194].

481 Functional screening of metagenomic librairies is an attractive approach because it holds the

482 potential to reveal hitherto unknown enzymes and guarantees that selected clones will harbor

483 proteins displaying the targeted activity, thus reducing the overall volume of high-throughput

484 sequencing [187]. Nevertheless, to implement function-based screening it is essential to use

485 appropriate screening strategies, typically using chromogenic substrates that can be

486 incorporated into solid agar medium and be visually detected upon hydrolysis. Hydrolases

487 such as lipases, amylases, cellulases and chitinases have often been the target of activity-

488 based screening [195], though examples of functional screening for hemicellulases are less

489 common [196-198]. The reasons for this are probably twofold: (i) hemicellulases have been

490 given less priority by researchers seeking to identify efficient cellulolytic systems and (ii)

491 chromogenic substrates for hemicellulases, such as L-arabinofuranosidases, are not readily

492 available even though the synthesis of 5-bromo-3-indolyl-alpha-L-arabinofuranoside was

493 described in 1996 [199]. Noteworthily, Marmuse *et al.* [200,201] recently described the

494 development of chromogenic substrates for feruloyl esterases and Fauré et *al*. has developed

495 novel chromogenic substrates for the detection of xylanase activity (unpublished results).

496 When coupled to fosmid library construction, which allows the insertion of large DNA

497 fragments, function-based approaches also provide access to multigenic clusters whose

498 components are often complementary to activity that is primarily targeted. This is illustrated

499 by a recent study by Tasse et *al*. who used a functional-based approach to mine the human gut

500 microbiome for dietary fiber catabolic enzymes. Hemicellulases, cellulases, amylases and

501 pectinases activity were detected in primary screening and 0.84 Mb of non redundant

502 metagenomic DNA was generated. Sequence analysis revealed 73 glycoside hydrolases,

503 belonging to 35 GH families, with many of these enzymes grouped into 18 multigenic

504 clusters, whose apparent functions are related to the degradation of plant cell walls [196].

505 Likewise, in very recent work, we have shown that the mining of the digestive microbiome of

506 the termite *Pseudacanthotermes militaris*, using hemicellulase detection for primary

507 screening, is a powerful way to discover components of microbial pathways for

508 hemicelluloses utilization (unpublished results).

509 Over the next years, undoubtedly metagenomics research will be a major source of new

510 enzymes for white biotechnology. However, to fully exploit this technology, it will be

511 necessary to overcome some of the outstanding difficulties, among which figures protein

512 expression bias. To address this problem, one solution will be to develop alternative hosts for

513 metagenomic libraries, expanding beyond *Escherichia coli*, which is today's workhorse.

514 Moreover it will be necessary to increase the throughput of cloning and expression of DNA,

515 which is a major bottleneck today, although solutions are now under development [202].

516 Beyond these critical technical obstacles, another key to further advancement is the clearer

517 understanding of how newly discovered enzymes function, especially within the context of

518 the complex multienzyme systems of natural microbiota. It is only with such precise

519 understanding of function that it will become possible to better conceive artificial enzyme

520 systems and/or microorganisms for complex tasks such as the degradation of lignocellulosic

521 biomass.

522

523 **Conclusions**

524 Clearly, the identification and engineering of new pentose-acting enzymes and

525 microorganisms are prerequisites for a better exploitation of pentose sugars in biorefining

526 strategies. The development of a varied toolbox of enzymes will provide considerable impetus

527 for innovation and will move the question of pentose valorization beyond the rather restrictive

528 D-xylose to ethanol framework. Among the key drivers of this transition, it can be expected

529 that metagenomics will provide many new enzymes, although major efforts regarding

530 screening will be required in order to go accelerate the discovery of other pentose-modifying

531 enzymes, such as isomerases, dehydratases and epimerases. These will be useful for novel

532 microbial strain engineering strategies, particularly using more ambitious approaches such as

533 synthetic biology.

534 Regarding biomass hydrolysis, it is increasingly recognized that hemicellulases are important

535 components of cellulolytic systems. Therefore, an intensification of research in this area can

536   be anticipated. Likewise, it is well recognized that consolidated biorefining provides a highly

537   optimized solution for biorefining lignocellulosic biomass, with considerable economic gains

538   expected [203,204]. However, to develop new efficient hydrolytic systems that rely, for

539   example, on designer cellulosomes, it will be vital to advance of our understanding of the

540   complex synergies that prevail in natural enzyme systems. Again, metagenomics will be

541   helpful in this area, but only if it is coupled to high-throughput, in-depth characterization of

542   the enzymes that are revealed.

543   Finally, glycosynthetic enzymes could play a future role in expanding the prospects for the

544   valorization of pentose sugars. In particular, we believe that the development of efficient

545   glycosynthetic tools could provide the next generation of tailored oligosaccharides and

546   pentose-based non-ionic detergents, especially if smart strategies involving recyclable,

547   immobilized enzymes are employed.

548

549   **References**

550   [1]   A.L. Demain, The business of biotechnology, Ind. Biotechnol. 3 (2007) 269-283.

551   [2]   M. Otero, José, G. Panagiotou, L. Olsson, Fueling industrial biotechnology growth
552   with bioethanol, Adv. Biochem. Eng/Biotechnol. 108 (2007) 1-40.

553   [3]   H. Ohara, Biorefinery, Appl. Microbiol. Biotechnol. 62 (2003) 474-477.

554   [4]   B. Kamm, M. Kamm, Principles of biorefineries, Appl. Microbiol. Biotechnol. 64
555   (2004) 137-145.

556   [5]   L. Mojović, D. Pejin, O. Grujić, S. Markov, M. Rakin, M. Vukašinović, et al., Progress
557   in the production of bioethnol from starch-based feedstocks, Chem. Ind. Chem. Eng.
558   Quart. 15 (2009) 211-226.

559   [6]   M.J.E.C. van der Maarel, B. van der Veen, J.C.M. Uitdehaag, H. Leemhuis, L.
560   Dijkhuizen, Properties and applications of starch-converting enzymes of the α-amylase
561   family, J. Biotechnol. 94 (2002) 137-155.

562 [7] C. Lequart, J.-M. Nuzillard, B. Kurek, P. Debeire, Hydrolysis of wheat bran and straw
563 by an endoxylanase: production and structural characterization of cinnamoyl-
564 oligosaccharides, Carbohydr. Res. 319 (1999) 102.

565 [8] S. Jin, H. Chen, Structural properties and enzymatic hydrolysis of rice straw, Process
566 Biochem. 41 (2006) 1261-1264.

567 [9] K.L. Kadam, L.H. Forrest, W.A. Jacobson, Rice straw as a lignocellulosic resource:
568 collection, processing, transportation, and environmental aspects, Biomass &
569 Bioenergy. 18 (2000) 369-389.

570 [10] H.V. Scheller, P. Ulvskov, Hemicelluloses, Ann. Rev. Plant Biol. 61 (2010) 263-289.

571 [11] B. Godin, F. Ghysel, R. Agneessens, T. Schmit, S. Gofflot, S. Lamaudière, et al.,
572 Détermination de la cellulose, des hémicelluloses, de la lignine et des cendres dans
573 diverses cultures lignocellulosiques dédiées à la production de bioéthanol de deuxième
574 génération, Biotechnol. Agron. Soc. Environ. 14 (2010) 549-560.

575 [12] M.S. Izydorczyk, C.G. Biliaderis, Cereal arabinoxylans: Advances in structure and
576 physicochemical properties, Carbohydr. Pol. 28 (1995) 33-48.

577 [13] L. Saulnier, Ferulic acid and diferulic acids as components of sugar-beet pectins and
578 maize bran, J. Agric. Food Chem. 79 (1999) 396-402.

579 [14] J. Ralph, Lignin-ferulate cross-links in grasses: active incorporation of ferulate
580 polysaccharide esters into ryegrass lignins, Carbohydr. Res. 275 (1995) 167-178.

581 [15] J. Grabber, Ferulate cross-linking in cell walls isolated from maize cell suspensions,
582 Phytochemistry. 40 (1995) 1077-1082.

583 [16] K. Iiyama, T.B.T. Lam, B.A. Stone, Covalent cross-links in the cell wall, Plant Physiol.
584 104 (1994) 315-320.

585 [17] F. Okazaki, Y. Tamaru, S. Hashikawa, Y.-T. Li, T. Araki, Novel carbohydrate-binding
586 module of β-1,3-xylanase from a marine bacterium, Alcaligenes sp. strain XY-234, J.
587 Bacteriol. 184 (2002) 2399-2403.

588 [18] E. Frei, R.D. Preston, Non-cellulosic structural polysaccharides in algal cell walls. I.
589 Xylan in siphoneous green algae, Proc. R. Soc. London, Ser. B. 160 (1964) 293-313.

590 [19] J.R. Turvey, E.L. Williams, The structures of some xylans from red algae,
591 Phytochemistry. 9 (1970) 2383-2388.

592 [20] B. Henrissat, A. Bairoch, Updating the sequence-based classification of glycosyl
593 hydrolases, Biochem. J. 316 (1996) 695-696.

594 [21] B. Henrissat, G. Davies, Structural and sequence-based classification of glycoside
595 hydrolases, Curr. Opin. Struc. Biol. 7 (1997) 637-644.

596 [22] B. Henrissat, A classification of glycosyl hydrolases based on amino acid sequence
597 similarities, Biochem. J. 280 (1991) 309-316.

598 [23] B. Henrissat, A. Bairoch, New families in the classification of glycosyl hydrolases
599 based on amino acid sequence similarities, Biochem. J. 293 (1993) 781-788.

600 [24] T. Collins, C. Gerday, G. Feller, Xylanases, xylanase families and extremophilic
601 xylanases, FEMS Microbiol. Rev. 29 (2005) 3-23.

602 [25] N. Kulkarni, A. Shendye, M. Rao, Molecular and biotechnological aspects of
603 xylanases, FEMS Microbiol. Rev. 23 (1999) 411-456.

604 [26] J. Gebler, N.R. Gilkes, M. Claeyssens, D.B. Wilson, P. Béguin, W.W. Wakarchuk, et
605 al., Stereoselective hydrolysis catalyzed by related β-1,4-glucanases and β-1,4-
606 xylanases, J. Biol. Chem. 267 (1992) 12559-12561.

607 [27] S.G. Withers, D. Dombroski, L.A. Berven, D.G. Kilburn, R.C.J. Miller, R.A. Warren,
608 et al., Direct 1H NMR determination of the stereochemical course of hydrolyses
609 catalysed by glucanase components of the cellulase complex, Biochem. Biophys. Res.
610 Commun. 139 (1986) 487-494.

611 [28] L. Lo Leggio, S. Kalogiannis, M.K. Bhat, R.W. Pickersgill, High resolution structure
612 and sequence of T . aurantiacus xylanase I : implications for the evolution of
613 thermostability in family 10 xylanases and enzymes with $\beta\alpha$-barrel architecture,
614 Proteins: Struct., Funct., Genet. 36 (1999) 295-306.

615 [29] N. Hakulinen, O. Turunen, J. Janis, M. Leisola, J. Rouvinen, Three-dimensional
616 structures of thermophilic β-1,4-xylanases from Chaetomium thermophilum and
617 Nonomuraea flexuosa. Comparison of twelve xylanases in relation to their thermal
618 stability, Eur. J. Biochem. 270 (2003) 1399-1412.

619 [30] M. Vardakou, C. Dumon, J.W. Murray, P. Christakopoulos, D.P. Weiner, N. Juge, et
620 al., Understanding the structural basis for substrate and inhibitor recognition in
621 eukaryotic GH11 xylanases, J. Mol. Biol. 375 (2008) 1293-1305.

622 [31] A. Torronen, J. Rouvinen, Structural comparison of two major endo-1,4-xylanases
623 from Trichoderma reesei, Biochemistry. 34 (1995) 847-856.

624 [32] A. Pollet, J.A. Delcour, C.M. Courtin, Structural determinants of the substrate
625 specificities of xylanases from different glycoside hydrolase families, Crit. Rev.
626 Biotechnol. 30 (2010) 1-16.

627 [33] G. Pell, E.J. Taylor, T.M. Gloster, J.P. Turkenburg, C. Fontes, L.M.A. Ferreira, et al.,
628 The mechanisms by which family 10 glycoside hydrolases bind decorated substrates, J.
629 Biol. Chem. 279 (2004) 9597-9605.

630 [34] R. Fauré, C.M. Courtin, J.A. Delcour, C. Dumon, C.B. Faulds, G.B. Fincher, et al., A
631 brief and informationally rich naming system for oligosaccharide motifs of
632 heteroxylans found in plant cell walls, Aust. J. Chem. 62 (2009) 533-537.

633 [35] J. Beaugrand, G. Chambat, V.W.K. Wong, F. Goubet, C. Rémond, G. Paës, et al.,
634   Impact and efficiency of GH10 and GH11 thermostable endoxylanases on wheat bran
635   and alkali-extractable arabinoxylans., Carbohydr. Res. 339 (2004) 2529-40.

636 [36] P. Biely, M. Vrsanská, M. Tenkanen, D. Kluepfel, Endo-β-1,4-xylanase families:
637   differences in catalytic properties, J. Biotechnol. 57 (1997) 151-166.

638 [37] S.J. Charnock, J.H. Lakey, R. Virden, N. Hughes, M.L. Sinnott, G.P. Hazlewood, et al.,
639   Key residues in subsite F play a critical role in the activity of Pseudomonas fluorescens
640   subspecies cellulosa xylanase A against xylooligosaccharides but not against highly
641   polymeric substrates such as xylan, J Biol Chem. 272 (1997) 2942-2951.

642 [38] J.K. Yang, H.J. Yoon, H.J. Ahn, B.I. Lee, J.D. Pedelacq, E.C. Liong, et al., Crystal
643   structure of β-D-xylosidase from Thermoanaerobacterium saccharolyticum, a family 39
644   glycoside hydrolase, J. Mol. Biol. 335 (2004) 155-165.

645 [39] Y.-G. Kim, J.-H. Kim, K.-J. Kim, Crystal structure of the Salmonella enterica serovar
646   typhimurium virulence factor SrfJ, a glycoside hydrolase family enzyme, J. Bacteriol.
647   191 (2009) 6550-6554.

648 [40] C. Brüx, A. Ben-David, D. Shallom-Shezifi, M. Leon, K. Niefind, G. Shoham, et al.,
649   The structure of an inverting GH43 β-xylosidase from Geobacillus stearothermophilus
650   with its substrate reveals the role of the three catalytic residues., J. Mol. Biol. 359
651   (2006) 97-109.

652 [41] A. Miyanaga, T. Koseki, H. Matsuzawa, T. Wakagi, H. Shoun, S. Fushinobu, Crystal
653   structure of a family 54 α-L-arabinofuranosidase reveals a novel carbohydrate-binding
654   module that can bind arabinose, J. Biol. Chem. 279 (2004) 44907-44914.

655 [42] I. Smaali, C. Rémond, M.J. O'Donohue, Expression in Escherichia coli and
656   characterization of β-xylosidases GH39 and GH-43 from Bacillus halodurans C-125,
657   Appl. Microbiol. Biotechnol. 73 (2006) 582-590.

658 [43] I. Smaali, C. Rémond, Y. Skhiri, M.J. O'Donohue, Biocatalytic conversion of wheat
659   bran hydrolysate using an immobilized GH43 β-xylosidase, Biores. Technol. 100
660   (2009) 338-344.

661 [44] M. Muzard, N. Aubry, R. Plantier-Royon, M.J. O'Donohue, C. Rémond, Evaluation of
662   the transglycosylation activities of a GH 39 β-D-xylosidase for the synthesis of xylose-
663   based glycosides, J. Mol. Catal. B: Enzym. 58 (2009) 1-5.

664 [45] Q.K. Beg, M. Kapoor, L. Mahajan, G.S. Hoondal, Microbial xylanases and their
665   industrial applications: a review, Appl. Microbiol. Biotechnol. 56 (2001) 326-338.

666 [46] D. Dodd, I.K.O. Cann, Enzymatic deconstruction of xylan for biofuel production.,
667   GCB Bioenergy. 1 (2009) 2-17.

668 [47] B.C. Saha, Hemicellulose bioconversion, J. Ind. Microbiol. Biotechnol. 30 (2003) 279-
669   291.

670 [48] H.R. Sørensen, S. Pedersen, A. Viksø-nielsen, A.S. Meyer, Efficiencies of designed
671   enzyme combinations in releasing arabinose and xylose from wheat arabinoxylan in an
672   industrial ethanol fermentation residue, Enzy. Microbiol. Technol. 36 (2005) 773-784.

673 [49] H.R. Sørensen, S. Pedersen, C.T. Jørgensen, A.S. Meyer, Enzymatic hydrolysis of
674   wheat arabinoxylan by a recombinant " minimal " enzyme cocktail containing β-
675   xylosidase and novel endo-1,4 β-xylanase and α-L-arabinofuranosidase activities,
676   Biotechnol. Prog. 23 (2007) 100-107.

677 [50] R. Kumar, C.E. Wyman, Effects of cellulase and xylanase enzymes on the
678   deconstruction of solids from pretreatment of poplar by leading technologies,
679   Biotechnol. Prog. 25 (2009) 302-314.

680 [51] B. Yang, C.E. Wyman, Pretreatment: the key to unlocking low-cost cellulosic ethanol,
681   Biofuels, Bioprod. Bioref. 2 (2008) 26-40.

682 [52] R. Kumar, C.E. Wyman, Effect of enzyme supplementation at moderate cellulase
683   loadings on initial glucose and xylose release from corn stover solids pretreated by
684   leading technologies, Biotechnol. Bioeng. 102 (2009) 457-467.

685 [53] R. Kumar, C.E. Wyman, Effect of xylanase supplementation of cellulase on digestion
686   of corn stover solids prepared by leading pretreatment technologies, Biores. Technol.
687   100 (2009) 4203-4213.

688 [54] S. Mcintosh, T. Vancov, Enhanced enzyme saccharification of Sorghum bicolor straw
689   using dilute alkali pretreatment, Biores. Technol. 101 (2010) 6718-6727.

690 [55] R.B. Hespell, P.J. Obryan, M. Moniruzzaman, R.J. Bothast, Hydrolysis by commercial
691   enzyme mixtures of AFEX-treated corn fiber and isolated xylans, Appl. Biochem.
692   Biotechnol. 62 (1997) 87-97.

693 [56] D. Shallom, Y. Shoham, Microbial hemicellulases, Curr. Opin. Microbiol. 6 (2003)
694   219-228.

695 [57] S.M. Pitson, G. Beldman, Stereochemical course of hydrolysis catalyzed by
696   arabinofuranosyl hydrolases, FEBS Lett. 398 (1996) 7-11.

697 [58] B.C. Saha, α-L-Arabinofuranosidases: biochemistry, molecular biology and application
698   in biotechnology, Biotechnol. Adv. 18 (2000) 403-423.

699 [59] T. Debeche, S. Cummings, I. Connerton, P. Debeire, M.J. O'Donohue, Genetic and
700   biochemical characterization of a highly thermostable α-L-arabinofuranosidase from
701   Thermobacillus xylanilyticus, Appl. Environ. Microbiol. 66 (2000) 1734-1736.

702 [60] M. Vardakou, J. Flint, P. Christakopoulos, R.J. Lewis, H.J. Gilbert, J.W. Murray, A
703   family 10 Thermoascus aurantiacus xylanase utilizes arabinose decorations of xylan as
704   significant substrate specificity determinants, J. Mol. Biol. 352 (2005) 1060-1067.

705 [61] M. Numan, N. Bhosle, α-L-Arabinofuranosidases: the potential applications in
706   biotechnology, J. Ind. Microbiol. Biotechnol. 33 (2006) 247-260.

707 [62] R.P. de Vries, H.C. Kester, C.H. Poulsen, J.A.E. Benen, J. Visser, Synergy between
708 enzymes from Aspergillus involved in the degradation of plant cell wall
709 polysaccharides., Carbohydr. Res. 327 (2000) 401-410.

710 [63] C. Rémond, I. Boukari, G. Chambat, M.J. O'Donohue, Action of a GH 51α- L-
711 arabinofuranosidase on wheat-derived arabinoxylans and arabino-xylooligosaccharides,
712 Carbohydr. Pol. 72 (2008) 424-430.

713 [64] J. Beaugrand, G. Paës, D. Reis, M. Takahashi, P. Debeire, M.J. O'Donohue, et al.,
714 Probing the cell wall heterogeneity of micro-dissected wheat caryopsis using both
715 active and inactive forms of a GH11 xylanase., Planta. 222 (2005) 246-57.

716 [65] C.J.A. Vinkx, J.A. Delcour, M.A. Verbruggen, H. Gruppen, Rye water-soluble
717 arabinoxylans also vary in their 2-monosubstituted xylose content, Cereal Chem. 72
718 (1995) 1993-1994.

719 [66] K.M.J. Van Laere, G. Beldman, A.G.J. Voragen, A new arabinofuranohydrolase from
720 Bifidobacterium adolescentis able to remove arabinosyl residues from double-
721 substituted xylose units in arabinoxylan, Appl. Microbiol. Biotechnol. 47 (1997) 231-
722 235.

723 [67] H.R. Sørensen, C.T. Jørgensen, C.H. Hansen, C.I. Jørgensen, S. Pedersen, A.S. Meyer,
724 A novel GH43 α-L-arabinofuranosidase from Humicola insolens: mode of action and
725 synergy with GH51 α-L-arabinofuranosidases on wheat arabinoxylan, Appl Microbiol
726 Biotechnol. 73 (2006) 850-861.

727 [68] S. Lagaert, A. Pollet, J.A. Delcour, R. Lavigne, C.M. Courtin, G. Volckaert, Substrate
728 specificity of three recombinant α-L-arabinofuranosidases from Bifidobacterium
729 adolescentis and their divergent action on arabinoxylan and arabinoxylan
730 oligosaccharides, Biochem. Biophys. Res. Commun. 402 (2010) 644-650.

731 [69] L.A.M. van den Broek, R.M. Lloyd, G. Beldman, J.C. Verdoes, B.V. McCleary, A.G.J.
732 Voragen, Cloning and characterization of arabinoxylan arabinofuranohydrolase-D3
733 (AXHd3) from Bifidobacterium adolescentis DSM20083, Appl. Microbiol. Biotechnol.
734 67 (2005) 641.

735 [70] A.S. Meyer, L. Rosgaard, H.R. Sørensen, The minimal enzyme cocktail concept for
736 biomass processing, J. Cereal Sci. 50 (2009) 337-344.

737 [71] N.N.M.E. Van Peij, J. Brinkmann, M. Vrsanská, J. Visser, L.H. De Graaff, Beta-
738 Xylosidase activity, encoded by xlnD, is essential for complete hydrolysis of xylan by
739 Aspergillus niger but not for induction of the xylanolytic enzyme spectrum, Eur. J.
740 Biochem. 245 (1997) 164-173.

741 [72] K. Wagschal, C. Heng, C. Lee, D. Wong, Biochemical characterization of a novel dual-
742 function arabinofuranosidase/xylosidase isolated from a compost starter mixture, Appl.
743 Microbiol. Biotechnol. 81 (2009) 855-863.

744 [73] K. Hövel, D. Shallom, K. Niefind, V. Belakhov, G. Shoham, T. Baasov, et al., Crystal
745   structure and snapshots along the reaction pathway of a family 51 α-L-
746   arabinofuranosidase, Embo J. 22 (2003) 4922-32.

747 [74] Z. Fan, K. Wagschal, C.C. Lee, Q. Kong, K. a Shen, I.B. Maiti, et al., The construction
748   and characterization of two xylan-degrading chimeric enzymes., Biotechnol. Bioeng.
749   102 (2009) 684-92.

750 [75] O. Guais, G. Borderies, C. Pichereaux, M. Maestracci, V. Neugnot, M. Rossignol, et
751   al., Proteomics analysis of "Rovabio (TM) Excel", a secreted protein cocktail from the
752   filamentous fungus Penicillium funiculosum grown under industrial process
753   fermentation, J. Ind. Microbiol. Biotechnol. 35 (2008) 1659-1668.

754 [76] S.S. Adav, C.S. Ng, M. Arulmani, S.K. Sze, Quantitative iTRAQ secretome analysis of
755   cellulolytic Thermobifida fusca, J. Proteome Res. 9 (2010) 3016-3024.

756 [77] C.B. Faulds, G. Williamson, Ferulic acid esterase from Aspergillus niger - purification
757   and partial characterization of two forms from a commercial source of pectinase,
758   Biotechnol. Biochem. 17 (1993) 349-359.

759 [78] P.A. Kroon, C.B. Faulds, G. Williamson, Purification and characterization of a novel
760   esterase induced by growth of Aspergillus niger on sugar beet pulp, Biotechnol. Appl.
761   Biochem. 23 (1996) 255-262.

762 [79] C.B. Faulds, G. Williamson, Purification and characterization of a ferulic acid esterase
763   (Fae-Iii) from Aspergillus niger - Specificity for the phenolic moiety and binding to
764   microcrystalline cellulose, Microbiology. 140 (1994) 779-787.

765 [80] R.P. de Vries, B. Michelsen, C.H. Poulsen, P.A. Kroon, R.H. van den Heuvel, C.B.
766   Faulds, et al., The faeA genes from Aspergillus niger and Aspergillus tubingensis
767   encode ferulic acid esterases involved in degradation of complex cell wall
768   polysaccharides, Appl. Environ. Microbiol. 63 (1997) 4638-4644.

769 [81] B.L. Cantarel, P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, The
770   carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics,
771   Nucl. Acids Res. 37 (2009) D233-238.

772 [82] V.F. Crepin, C.B. Faulds, I.F. Connerton, Functional classification of the microbial
773   feruloyl esterases, Appl. Microbiol. Biotechnol. 63 (2004) 647-652.

774 [83] M. Moukouli, E. Topakas, P. Christakopoulos, Cloning, characterization and functional
775   expression of an alkalitolerant type C feruloyl esterase from Fusarium oxysporum,
776   Appl. Microbiol. Biotechnol. 79 (2008) 245-254.

777 [84] B. Bartolome, C.B. Faulds, P.A. Kroon, K. Waldron, H.J. Gilbert, G. Hazlewood, et al.,
778   An Aspergillus niger esterase (ferulic acid esterase III) and a recombinant
779   Pseudomonas fluorescens subsp. cellulosa esterase (Xy1D) release a 5-5' ferulic
780   dehydrodimer (diferulic acid) from barley and wheat cell walls, Appl. Environ.
781   Microbiol. 63 (1997) 208-212.

782 [85] D.L. Blum, I.A. Kataeva, X.-L. Li, L.G. Ljungdahl, Feruloyl esterase activity of the
783   Clostridium thermocellum cellulosome can be attributed to previously unknown
784   domains of XynY and XynZ, J. Bacteriol. 182 (2000) 1346-1351.

785 [86] S. Yoshida, R.I. Mackie, I.K.O. Cann, Biochemical and domain analyses of
786   FSUAxe6B, a modular acetyl xylan esterase, identify a unique carbohydrate binding
787   module in Fibrobacter succinogenes S85, J. Bacteriol. 192 (2010) 483-493.

788 [87] C.-K. Pai, Z.-Y. Wu, M.-J. Chen, Y.-F. Zeng, J.-W. Chen, C.-H. Duan, Molecular
789   cloning and characterization of a bifunctional xylanolytic enzyme from Neocallimastix
790   patriciarum, Appl. Microbiol. Biotechnol. 85 (2010) 1451-1462.

791 [88] A. Levasseur, D. Navarro, P.J. Punt, J.P. Belaich, M. Asther, E. Record, Construction
792   of engineered bifunctional enzymes and their overproduction in Aspergillus niger for
793   improved enzymatic tools to degrade agricultural by-products, Appl. Environ.
794   Microbiol. 71 (2005) 8132-8140.

795 [89] C.B. Faulds, What can feruloyl esterases do for us?, Phytochem. Rev. 9 (2010) 121-
796   132.

797 [90] C.B. Faulds, D. Zanichelli, V.F. Crepin, I.F. Connerton, N. Juge, M.K. Bhat, et al.,
798   Specificity of feruloyl esterases for water-extractable and water-unextractable
799   feruloylated polysaccharides: influence of xylanase, J. Cereal Sci. 38 (2003) 281-288.

800 [91] M. Selig, W. Adney, M. Himmel, S. Decker, The impact of cell wall acetylation on
801   corn stover hydrolysis by cellulolytic and xylanolytic enzymes, Cellulose. 16 (2009)
802   711-722.

803 [92] M.J. Selig, E.P. Knoshaug, W.S. Adney, M.E. Himmel, S.R. Decker, Synergistic
804   enhancement of cellobiohydrolase performance on pretreated corn stover by addition of
805   xylanase and esterase activities, Biores. Technol. 99 (2008) 4997-5005.

806 [93] M.M. Appeldoorn, M. Kabel, D. Van Eylen, H. Gruppen, H. Schols, Characterization
807   of oligomeric xylan structures from corn fiber resistant to pretreatment and
808   simultaneous saccharification and fermentation, J. Agric. Food Chem. (2010) 11294-
809   11301.

810 [94] B. Hahn-Hägerdal, K. Karhumaa, C. Fonseca, I. Spencer-Martins, M. Gorwa-
811   Grauslund, Towards industrial pentose-fermenting yeast strains, Appl. Microbiol.
812   Biotechnol. 74 (2007) 937-953.

813 [95] B. Hahn-Hägerdal, K. Karhumaa, M. Jeppsson, M.F. Gorwa-Grauslund, Metabolic
814   engineering for pentose utilization in Saccharomyces cerevisiae, Adv. Biochem.
815   Eng/Biotechnol. 108 (2007) 147-177.

816 [96] T. Hamacher, J. Becker, M. Gardonyi, B. Hahn-Hägerdal, E. Boles, Characterization of
817   the xylose-transporting properties of yeast hexose transporters and their influence on
818   xylose utilization, Microbiology. 148 (2002) 2783-2788.

819   [97]   A.J.A. van Maris, D.A. Abbott, E. Bellissimi, J. van den Brink, M. Kuyper, M.A.H.
820   Luttik, et al., Alcoholic fermentation of carbon sources in biomass hydrolysates by
821   Saccharomyces cerevisiae : current status, Antonie Van Leeuwenhoek. 90 (2006) 391-
822   418.

823   [98]   L.-C. Chiang, C.-S. Gong, L.-F. Chen, G.T. Tsao, D-Xylulose fermentation to ethanol
824   by Saccharomyces cerevisiae, Appl. Environ. Microbiol. 42 (1981) 284-289.

825   [99]   M. Kuyper, H.R. Harhangi, A. Kristin, A.A. Winkler, M.S.M. Jetten, W.T.A.M. de
826   Laat, et al., High-level functional expression of a fungal xylose isomerase : the key to
827   efficient ethanolic fermentation of xylose by Saccharomyces cerevisiae ?, FEMS Yeast
828   Res. 4 (2003) 69-78.

829   [100] M. Kuyper, M.J. Toirkens, J.A. Diderich, A.A. Winkler, J.P. van Dijken, J.T. Pronk,
830   Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting
831   Saccharomyces cerevisiae strain, FEMS Yeast Res. 5 (2005) 925-934.

832   [101] M. Kuyper, M.M.P. Hartog, M.J. Toirkens, M.J.H. Almering, A.A. Winkler, J.P. van
833   Dijken, et al., Metabolic engineering of a xylose isomerase expressing Saccharomyces
834   cerevisiae strain for rapid anaerobic xylose fermentation, FEMS Yeast Res. 5 (2005)
835   399-409.

836   [102] M. Kuyper, A.A. Winkler, J.P. van Dijken, J.T. Pronk, Minimal metabolic engineering
837   of Saccharomyces cerevisiae for efficient anaerobic xylose fermentation: a proof of
838   principle, FEMS Yeast Res. 4 (2004) 655-664.

839   [103] E. Young, S.M. Lee, H. Alper, Optimizing pentose utilization in yeast: the need for
840   novel tools and approaches, Biotechnol. Biofuels. 3 (2010).

841   [104] D. Runquist, B. Hahn-Hägerdal, P. Radstrom, Comparison of heterologous xylose
842   transporters in recombinant Saccharomyces cerevisiae, Biotechnol. Biofuels. 3 (2010).

843   [105] C. Fonseca, K. Olofsson, C. Ferreira, D. Runquist, L.L. Fonseca, B. Hahn-Hägerdal, et
844   al., The glucose/xylose facilitator Gxf1 from Candida intermedia expressed in a xylose-
845   fermenting industrial strain of Saccharomyces cerevisiae increases xylose uptake in
846   SSCF of wheat straw, Enz. Microb. Technol. 48 (2011) 518-525.

847   [106] P. Richard, R. Verho, M. Putkonen, J. Londesborough, M. Penttilä, Production of
848   ethanol from L-arabinose by Saccharomyces cerevisiae containing a fungal L-arabinose
849   pathway, FEMS Yeast Res. 3 (2003) 185–189.

850   [107] M. Sedlak, N.W.Y. Ho, Expression of E. coli araBAD operon encoding enzymes for
851   metabolizing L-arabinose in Saccharomyces cerevisiae, Enz. Microb. Technol. 28
852   (2001) 16-24.

853   [108] A.K. Bera, M. Sedlak, A. Khan, N.W.Y. Ho, Establishment of L-arabinose
854   fermentation in glucose/xylose co-fermenting recombinant Saccharomyces cerevisiae
855   424A(LNH-ST) by genetic engineering, Appl. Microbiol. Biotechnol. 87 (2010) 1803-
856   1811.

857 [109] C. Weber, A. Farwick, F. Benisch, D. Brat, H. Dietz, T. Subtil, et al., Trends and
858   challenges in the microbial production of lignocellulosic bioalcohol fuels, Applied
859   Microbiology And Biotechnology. 87 (2010) 1303.

860 [110] V.V. Zverlov, O. Berezina, G.A. Velikodvorskaya, W.H. Schwarz, Bacterial acetone
861   and butanol production by industrial fermentation in the Soviet Union: use of
862   hydrolyzed agricultural waste for biorefinery, Appl Microbiol Biotechnol. 71 (2006)
863   587-597.

864 [111] K. Ounine, H. Petitdemange, G. Raval, R. Gay, Regulation and butanol inhibition of D-
865   xylose and D-glucose uptake in Clostridium acetobutylicum, Appl Environ Microbiol.
866   49 (1985) 874-878.

867 [112] C. Grimmler, C. Held, W. Liebl, A. Ehrenreich, Transcriptional analysis of catabolite
868   repression in Clostridium acetobutylicum growing on mixtures of D-glucose and D-
869   xylose, J. Biotechnol. 150 (2010) 315-323.

870 [113] C. Ren, Y. Gu, S. Hu, Y. Wu, P. Wang, Y. Yang, et al., Identification and inactivation
871   of pleiotropic regulator CcpA to eliminate glucose repression of xylose utilization in
872   Clostridium acetobutylicum, Metabolic Engineering. 12 (2010) 446-454.

873 [114] S. Hu, H. Zheng, Y. Gu, J. Zhao, W. Zhang, Y. Yang, et al., Comparative genomic and
874   transcriptomic analysis revealed genetic characteristics related to solvent formation and
875   xylose utilization in Clostridium acetobutylicum EA 2018, BMC Genomics. 12 (2011).

876 [115] T.B. Granström, K. Izumori, M. Leisola, A rare sugar xylitol. Part I: the biochemistry
877   and biosynthesis of xylitol, Appl. Microbiol. Biotechnol. 74 (2007) 277-281.

878 [116] M.G. García-Martín, R.R. Pérez, E.B. Hernández, J.A. Galbis, Synthesis of L-
879   arabinitol and xylitol monomers for the preparation of polyamides. Preparation of an L-
880   arabinitol-based polyamide, Carbohydr. Res. 333 (2001) 95-103.

881 [117] A. Alla, K. Hakkou, F. Zamora, A. MartinezdeIlarduya, J.A. Galbis, S. Munoz-Guerra,
882   Poly(butylene terephthalate) copolyesters derived from L-arabinitol and xylitol,
883   Macromolecules. 39 (2006) 1410-1416.

884 [118] J.P. Bruggeman, C.J. Bettinger, C.L. Nijst, D.S. Kohane, R. Langer, Biodegradable
885   xylitol-based polymers, Adv. Mater. 20 (2008) 1922-1927.

886 [119] M. McCoy, A sweet result with a life-cycle assessment , Danisco seeks to enhance
887   status of its xylitol, Chem. Eng. News. 88 (2010) 14-15.

888 [120] T.B. Granström, K. Izumori, M. Leisola, A rare sugar xylitol. Part II: biotechnological
889   production and future applications of xylitol, Appl. Microbiol. 74 (2007) 273-276.

890 [121] A. Corma, S. Iborra, A. Velty, Chemical routes for the transformation of biomass into
891   chemicals, Chem. Rev. 107 (2007) 2411-2502.

892 [122] N.U. Nair, H. Zhao, Selective reduction of xylose to xylitol from a mixture of
893   hemicellulosic sugars, Metabol. Eng. 12 (2010) 462-468.

894 [123] J.J. Bozell, G.R. Petersen, Technology development for the production of biobased
895   products from biorefinery carbohydrates-the US Department of Energy's "Top 10"
896   revisited, Green Chem. 12 (2010) 539-554.

897 [124] N.U. Nair, H. Zhao, Evolution in reverse: engineering a D-xylose-specific xylose
898   reductase, ChemBioChem. 9 (2008) 1213-1215.

899 [125] U. Johnsen, M. Dambeck, H. Zaiss, T. Fuhrer, J. Soppa, U. Sauer, et al., D-Xylose
900   degradation pathway in the halophilic archaeon Haloferax volcanii, J. Appl. Microbiol.
901   284 (2009) 27290-27303.

902 [126] C.E.M. Nunn, U. Johnsen, P. Schönheit, T. Fuhrer, U. Sauer, D.W. Hough, et al.,
903   Metabolism of pentose sugars in the hyperthermophilic archaea Sulfolobus solfataricus
904   and Sulfolobus acidocaldarius, J. Biol. Chem. 285 (2010) 33701-33709.

905 [127] B.-W. Chun, B. Dair, P.J. Macuch, D. Wiebe, C. Porteneuve, A. Jeknavorian, The
906   development of cement and concrete additive: based on xylonic acid derived via
907   bioconversion of xylose, Appl. Biochem. Biotechnol. 129-132 (2006) 645-658.

908 [128] M.H. Toivari, L. Ruohonen, P. Richard, M. Penttilä, M.G. Wiebe, Saccharomyces
909   cerevisiae engineered to produce D-xylonate, Appl. Microbiol. Biotechnol. 88 (2010)
910   751-760.

911 [129] F. Pezzotti, M. Therisod, Enzymatic synthesis of aldonic acids, Carbohydr. Res. 341
912   (2006) 2290-2292.

913 [130] W. Niu, M.N. Molefe, J.W. Frost, Microbial synthesis of the energetic material
914   precursor 1,2,4-butanetriol, J. Am. Chem. Soc. 125 (2003) 12998-12999.

915 [131] J.P. Meijnen, J.H. de Winde, H.J. Ruijssenaars, Establishment of oxidative D-xylose
916   metabolism in Pseudomonas putida S12, Appl. Environ. Microbiol. 75 (2009) 2784-
917   2791.

918 [132] P. Wang, J.H. Hong, J.S. Cooperwood, C.K. Chu, Recent advances in L-nucleosides:
919   chemistry and biology, Antiviral Res. 40 (1998) 19-44.

920 [133] Y. Chong, C.K. Chu, Efficient synthesis of 2-deoxy-erythro-pentose (2-deoxy--ribose)
921   from L-arabinose, Carbohydr. Res. 337 (2002) 397-402.

922 [134] C. De Muynck, C. Pereira, W. Soetaert, E. Vandamme, Dehydrogenation of ribitol with
923   Gluconobacter oxydans: production and stability of L-ribulose, J. Biotechnol. 125
924   (2006) 408-415.

925 [135] Y.-W. Zhang, M. Jeya, J.-K. Lee, L-Ribulose production by an Escherichia coli
926   harboring L-arabinose isomerase from Bacillus licheniformis, Appl. Microbiol.
927   Biotechnol. 87 (2010) 1993-1999.

928 [136] M. Helanto, K. Kiviharju, T. Granström, M. Leisola, A. Nyyssölä, Biotechnological
929   production of L-ribose from L-arabinose, Appl. Microbiol. Biotechnol. 83 (2009) 77-
930   83.

931 [137] M. Helanto, K. Kiviharju, M. Leisola, A. Nyyssölä, Metabolic engineering of
932   Lactobacillus plantarum for production of L-ribulose, Appl; Environ. Microbiol. 73
933   (2007) 7083-7091.

934 [138] C.M. Courtin, W.F. Broekaert, K. Swennen, O. Lescroart, O. Onagbesan, J. Buyse, et
935   al., Dietary inclusion of wheat bran arabinoxylooligosaccharides induces beneficial
936   nutritional effects in chickens, Cereal Chem. 85 (2008) 607-613.

937 [139] C. Grootaert, J. Delcour, C. Courtin, W. Broekaert, W. Verstraete, T. Vandewiele,
938   Microbial metabolism and prebiotic potency of arabinoxylan oligosaccharides in the
939   human intestine, Trends Food Sci. Technol. 18 (2007) 64-71.

940 [140] F. Bouxin, S. Marinkovic, J. Le Bras, B. Estrine, Direct conversion of xylan into alkyl
941   pentosides., Carbohydr. Res. 345 (2010) 2469-2473.

942 [141] B. Estrine, C. Ernenwein, A. Bresin, A. Wintrebert, T. Benvegnu, D. Plusquellec,
943   Novel family of alkyl polyglycoside compositions and compounds and compounds
944   derived from glycine betain, use as surfactant, Patent.WO/2005/121294. (2005)
945   121294.

946 [142] P. Biely, M. Vrsanská, Xylosyl transfer to cellobiose by an endo-(1-4)-β-D-xylanase of
947   Crytococcus albus, Carbohydr. Res. 123 (1983) 97-107.

948 [143] M. Vrsanská, I.V. Gorbacheva, Z. Krátký, P. Biely, Reaction pathways of substrate
949   degradation by an acidic endo-1,4-β-xylanase of Aspergillus niger, Biochim. Biophys.
950   Acta. 704 (1982) 114-122.

951 [144] A. Moreau, F. Shareck, D. Kluepfel, R. Morosoli, Alteration of the cleavage mode and
952   of the transglycosylation reactions of the xylanase A of Streptomyces lividans 1326 by
953   site-directed mutagenesis of the Asn173 residue, Eur. J. Biochem. 219 (1994) 261-266.

954 [145] S. Armand, S.R. Andrews, S.J. Charnock, H.J. Gilbert, Influence of the aglycone
955   region of the substrate binding cleft of Pseudomonas xylanase 10A on catalysis,
956   Biochemistry. 40 (2001) 7404-7409.

957 [146] S. Watanabe, D.N. Viet, J. Kaneko, Y. Kamio, S. Yoshida, Cloning, expression, and
958   transglycosylation reaction of Paenibacillus sp. strain W-61 xylanase 1, Biosci.,
959   Biotechnol., Biochem. 72 (2008) 951-958.

960 [147] M.K. Kang, P.J. Maeng, Y.H. Rhee, Purification and characterization of two xylanases
961   from alkalophilic Cephalosporium sp. strain RYM-202., Appl. Environ. Microbiol. 62
962   (1996) 3480-3482.

963 [148] B.S. Chadha, B.K. Ajay, F. Mellon, M.K. Bhat, Two endoxylanases active and stable at
964   alkaline pH from the newly isolated thermophilic fungus, Myceliophthora sp. IMI
965   387099., J. Biotechnol. 109 (2004) 227-237.

966 [149] A.K. Badhan, B.S. Chadha, J. Kaur, K.G. Sonia, H.S. Saini, M.K. Bhat, Role of
967   transglycosylation products in the expression of multiple xylanases in Myceliophthora
968   sp. IMI 387099, Curr. Microbiol. 54 (2007) 405-409.

969   [150] K. Wagschal, D. Franqui-Espiet, C.C. Lee, G.H. Robertson, D.W.S. Wong, Cloning,
970   expression and characterization of a glycoside hydrolase family 39 xylosidase from
971   Bacillus halodurans C-125, Appl. Biochem. Biotechnol. 146 (2008) 69-78.

972   [151] K. Wagschal, D. Franqui-Espiet, C.C. Lee, G.H. Robertson, D.W.S. Wong, Enzyme-
973   coupled assay for beta-xylosidase hydrolysis of natural substrates, Appl. Environ.
974   Microbiol. 71 (2005) 5318-5323.

975   [152] S. Matsumura, K. Sakiyama, K. Toshima, Preparation of octyl β-D-xylobioside and
976   xyloside by xylanase-catalyzed direct transglycosylation reaction of xylan and octanol,
977   Biotechnol. Lett. 21 (1999) 17-22.

978   [153] Z.Q. Jiang, Y.P. Zhu, L. Li, X.H. Yu, I. Kusakabe, M. Kitaoka, et al.,
979   Transglycosylation reaction of xylanase B from the hyperthermophilic Thermotoga
980   maritima with the ability of synthesis of tertiary alkyl β-D-xylobiosides and xylosides,
981   J. Biotechnol. 114 (2004) 125-134.

982   [154] N. Kadi, J. Crouzet, Transglycosylation reaction of endoxylanase from Trichoderma
983   longibrachiatum, Food Chem. 106 (2008) 466-474.

984   [155] G. Mamo, S. Kasture, R. Faryar, S. Hashim, R. Hatti-Kaul, Surfactants from xylan:
985   Production of n-octyl xylosides using a highly thermostable xylanase from Thermotoga
986   neapolitana, Process Biochem. 45 (2010) 700-705.

987   [156] A. Tramice, E. Pagnotta, I. Romano, A. Gambacorta, A. Trincone, Transglycosylation
988   reactions using glycosyl hydrolases from Thermotoga neapolitana, a marine hydrogen-
989   producing bacterium, J. Mol. Catal. B: Enzym. 47 (2007) 21-27.

990   [157] K. Chiku, J. Uzawa, H. Seki, S. Amachi, T. Fujii, H. Shinoyama, Characterization of a
991   novel polyphenol-specific oligoxyloside transfer reaction by a family 11 xylanase from
992   Bacillus sp. KT12, Biosci., Biotechnol., Biochem. 72 (2008) 2285-2293.

993   [158] H. Oshima, I. Kimura, K. Izumori, Synthesis and structure analysis of novel
994   disaccharides containing D-psicose produced by endo-1,4-β-D-xylanase from
995   Aspergillus sojae, J. Biosci. Bioeng. 101 (2006) 280-283.

996   [159] N. Kadi, L. Belloy, P. Chalier, J. Crouzet, Enzymatic synthesis of aroma compound
997   xylosides using transfer reaction by Trichoderma longibrachiatum xylanase, J. Agric.
998   Food Chem. 50 (2002) 5552-5557.

999   [160] J. Zhengqiang, M. Ahsan, L.I. Lite, Characterization of a thermostable family 10 endo-
1000   xylanase (XynB) from Thermotoga maritima that cleaves para-nitrophenyl-β-D-
1001   xyloside, J. Biosci. Bioeng. 92 (2001) 423-428.

1002   [161] E.V. Eneyskaya, D.R. Ivanen, K.S. Bobrov, L.S. Isaeva-Ivanova, K. Shabalin, A.N.
1003   Savelev, et al., Biochemical and kinetic analysis of the GH3 family β-xylosidase from
1004   Aspergillus awamori X-100, Arch. Biochem. Biophys. 457 (2007) 225-234.

1005   [162] E.V. Eneyskaya, H. Brumer, L.V. Backinowsky, D.R. Ivanen, A. Kulminskaya, K.
1006   Shabalin, et al., Enzymatic synthesis of beta-xylanase substrates: transglycosylation

1007   reactions of the β-xylosidase from Aspergillus sp., Carbohydr. Res. 338 (2003) 313-
1008   325.

1009   [163] E.V. Eneyskaya, D.R. Ivanen, K.A. Shabalin, A.A. Kulminskaya, L.V. Backinowsky,
1010   H. Brumer, et al., Chemo-enzymatic synthesis of 4-methylumbelliferyl β-(1→4)-D-
1011   xylooligosides: new substrates for β-D-xylanase assays, Org Biomol Chem. 3 (2005)
1012   146-151.

1013   [164] C. Rémond, M. Ferchichi, N. Aubry, R. Plantier-Royon, C. Portella, M.J. O'Donohue,
1014   Enzymatic synthesis of alkyl arabinofuranosides using a thermostable alpha-L-
1015   arabinofuranosidase, Tetrahedron Lett. 43 (2002) 9653-9655.

1016   [165] C. Rémond, R. Plantier-Royon, N. Aubry, E. Maes, C. Bliard, M.J. O'Donohue,
1017   Synthesis of pentose-containing disaccharides using a thermostable α-L-
1018   arabinofuranosidase, Carbohydr. Res. 339 (2004) 2019-2025.

1019   [166] C. Rémond, R. Plantier-Royon, N. Aubry, M.J. O'Donohue, An original
1020   chemoenzymatic route for the synthesis of beta-D-galactofuranosides using an alpha-L-
1021   arabinofuranosidase, Carbohydr. Res. 340 (2005) 637-644.

1022   [167] I. Chlubnová, D. Filipp, V. Spiwok, H. Dvoráková, R. Daniellou, C. Nugier-Chauvin,
1023   et al., Enzymatic synthesis of oligo-D-galactofuranosides and L-arabinofuranosides:
1024   from molecular dynamics to immunological assays, Org. Biomol. Chem. 8 (2010)
1025   2092-2102.

1026   [168] T. Sakamoto, T. Fujita, H. Kawasaki, Transglycosylation catalyzed by a Penicillium
1027   chrysogenum exo-1,5-α-L-arabinanase, Biochim. Biophys. Acta. 1674 (2004) 85-90.

1028   [169] C.F. Wan, W.H. Chen, C.T. Chen, M.D.T. Chang, L.C. Lo, Y.K. Li, Mutagenesis and
1029   mechanistic study of a glycoside hydrolase family 54 α-L-arabinofuranosidase from
1030   Trichoderma koningii, Biochem. J. 401 (2007) 551-558.

1031   [170] L.F. Mackenzie, Q. Wang, R.A.J. Warren, S.G. Withers, Glycosynthases: mutant
1032   glycosidases for oligosaccharide synthesis, J. Am. Chem. Soc. 120 (1998) 5583-5584.

1033   [171] M. Kitaoka, Y. Honda, S. Fushinobu, H. Masafumi, K. Takane, Y. Kenji, Conversion
1034   of inverting glycoside hydrolases into catalysts for synthesizing glycosides employing
1035   a glycosynthase strategy, Trends Glycosci. Glycotechnol. 21 (2009) 23-39.

1036   [172] B. Rakic, S.G. Withers, Recent developments in glycoside synthesis with
1037   glycosynthases and thioglycoligases, Aust. J. Chem. 62 (2009) 510-520.

1038   [173] Y.-W. Kim, D.T. Fox, O. Hekmat, T. Kantner, L.P. McIntosh, R.A.J. Warren, et al.,
1039   Glycosynthase-based synthesis of xylo-oligosaccharides using an engineered retaining
1040   xylanase from Cellulomonas fimi, Org. Biomol. Chem. 4 (2006) 2025-2032.

1041   [174] M. Sugimura, M. Nishimoto, M. Kitaoka, Characterization of glycosynthase mutants
1042   derived from glycoside hydrolase family 10 xylanases, Biosci., Biotechnol., Biochem.
1043   70 (2006) 1210-1217.

1044 [175] A. Ben-David, T. Bravman, Y.S. Balazs, M. Czjzek, D. Schomburg, G. Shoham, et al.,
1045   Glycosynthase activity of Geobacillus stearothermophilus GH52 β-xylosidase: efficient
1046   synthesis of xylooligosaccharides from α-D-xylopyranosyl fluoride through a
1047   conjugated reaction, ChemBioChem. 8 (2007) 2145-2151.

1048 [176] J. Handelsman, M.R. Rondon, S.F. Brady, J. Clardy, R.M. Goodman, Molecular
1049   biological access to the chemistry of unknown soil microbes: a new frontier for natural
1050   products, Chem. Biol. 5 (1998) 245-249.

1051 [177] W.B. Whitman, D.C. Coleman, W.J. Wiebe, Prokaryotes: The unseen majority, Proc.
1052   Natl. Acad. Sci. U.S.A. 95 (1998) 6578-6583.

1053 [178] J.C. Venter, K. Remington, J.F. Heidelberg, A.L. Halpern, D. Rusch, J.A. Eisen, et al.,
1054   Environmental Genome Shotgun Sequencing of the Sargasso Sea, Science. 304 (2004)
1055   66-74.

1056 [179] S. Voget, C. Leggewie, A. Uesbeck, C. Raasch, K.E. Jaeger, W.R. Streit, Prospecting
1057   for novel biocatalysts in a soil metagenome, Appl. Environ. Microbiol. 69 (2003) 6235-
1058   6242.

1059 [180] J. Qin, R. Li, J. Raes, M. Arumugam, K.S. Burgdorf, C. Manichanh, et al., A human
1060   gut microbial gene catalogue established by metagenomic sequencing, Nature. 464
1061   (2010) 59-67.

1062 [181] S.R. Gill, M. Pop, R.T. Deboy, P.B. Eckburg, P.J. Turnbaugh, B.S. Samuel, et al.,
1063   Metagenomic analysis of the human distal gut microbiome., Science. 312 (2006) 1355-
1064   1359.

1065 [182] P.J. Turnbaugh, R.E. Ley, M.A. Mahowald, V. Magrini, E.R. Mardis, J.I. Gordon, An
1066   obesity-associated gut microbiome with increased capacity for energy harvest, Nature.
1067   444 (2006) 1027-1031.

1068 [183] P.J. Turnbaugh, M. Hamady, T. Yatsunenko, B.L. Cantarel, A. Duncan, R.E. Ley, et
1069   al., A core gut microbiome in obese and lean twins, Nature. 457 (2009) 480-484.

1070 [184] J.M. Brulc, D.A. Antonopoulos, M.E. Berg Miller, M.K. Wilson, A.C. Yannarell, E.A.
1071   Dinsdale, et al., Gene-centric metagenomics of the fiber-adherent bovine rumen
1072   microbiome reveals forage specific glycoside hydrolases, Proc. Natl. Acad. Sci. U.S.A.
1073   106 (2009) 1948-1953.

1074 [185] F. Warnecke, P. Luginbühl, N. Ivanova, M. Ghassemian, T.H. Richardson, J.T. Stege,
1075   et al., Metagenomic and functional analysis of hindgut microbiota of a wood-feeding
1076   higher termite, Nature. 450 (2007) 560-565.

1077 [186] D. MacLean, J.D.G. Jones, D.J. Studholme, Application of "next-generation"
1078   sequencing technologies to microbial genetics, Nature Rev. Microbiol. 7 (2009) 287-
1079   296.

1080 [187] L.L. Li, S.R. McCorkle, S. Monchy, S. Taghavi, D. van der Lelie, Bioprospecting
1081   metagenomes: glycosyl hydrolases for converting biomass, Biotechnol. Biofuels. 2
1082   (2009).

1083 [188] C. Simon, R. Daniel, Achievements and new knowledge unraveled by metagenomic
1084   approaches, Appl. Microbiol. Biotechnol. 85 (2009) 265-276.

1085 [189] D. Cowan, Q. Meyer, W. Stafford, S. Muyanga, R. Cameron, P. Wittwer, Metagenomic
1086   gene discovery: past, present and future, Trends Biotechnol. 23 (2005) 321-329.

1087 [190] M.G. Dumont, J.C. Murrell, Stable isotope probing linking microbial identity to
1088   function, Nature Rev. Microbiol. 3 (2005) 499-504.

1089 [191] S. Radajewski, P. Ineson, N.R. Parekh, J.C. Murrell, Stable-isotope probing as a tool in
1090   microbial ecology, Nature. 403 (2000) 646-649.

1091 [192] K. DeAngelis, J. Gladden, M. Allgaier, P. D'haeseleer, J. Fortney, A. Reddy, et al.,
1092   Strategies for enhancing the effectiveness of metagenomic-based enzyme discovery in
1093   lignocellulolytic microbial communities, Bioenergy Res. 3 (2010) 146-158.

1094 [193] M. Allgaier, A. Reddy, J.I. Park, N. Ivanova, P. D'Haeseleer, S. Lowry, et al., Targeted
1095   discovery of glycoside hydrolases from a switchgrass-adapted compost community,
1096   PLoS One. 5 (2010) e8812.

1097 [194] M. Hess, A. Sczyrba, R. Egan, T.-W. Kim, H. Chokhawala, G. Schroth, et al.,
1098   Metagenomic discovery of biomass-degrading genes and genomes from cow rumen,
1099   Science. 331 (2011) 463-467.

1100 [195] P. Lorenz, J. Eck, Metagenomics and industrial applications, Nature Rev. Microbiol. 3
1101   (2005) 510-516.

1102 [196] L. Tasse, J. Bercovici, S. Pizzut-Serin, P. Robe, J. Tap, C. Klopp, et al., Functional
1103   metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes,
1104   Genome Res. 20 (2010) 1605-1612.

1105 [197] Z.M. Fan, K. Wagschal, W. Chen, M.D. Montross, C.C. Lee, L. Yuan, Multimeric
1106   hemicellulases facilitate biomass conversion, Appl. Environ. Microbiol. 75 (2009)
1107   1754-1757.

1108 [198] C.C. Lee, R.E. Kibblewhite-Accinelli, K. Wagschal, G.H. Robertson, D.W.S. Wong,
1109   Cloning and characterization of a cold-active xylanase enzyme from an environmental
1110   DNA library, Extremophiles. 10 (2006) 295-300.

1111 [199] W. Berlin, B. Sauer, In situ color detection of α-L-arabinofuranosidase, a "no-
1112   background" reporter gene, with 5-bromo-3-indolyl-α-L-arabinofuranoside, Anal.
1113   Biochem. 243 (1996) 171-175.

1114 [200] L. Marmuse, M. Asther, D. Navarro, L. Lesage-Meessen, S. Fort, H. Driguez,
1115   Chromogenic substrates for feruloyl esterases, Carbohydr. Res. 342 (2007) 2316-2321.

1116 [201] L. Marmuse, M. Asther, E. Fabre, D. Navarro, L. Lesage-Meessen, M.J. O'Donohue, et
1117   al., New chromogenic substrates for feruloyl esterases, Org. Biomol. Chem. 6 (2008)
1118   1208-1214.

1119 [202] A. Groisillier, C. Hervé, A. Jeudy, E. Rebuffet, P.F. Pluchon, Y. Chevolot, et al.,
1120   MARINE-EXPRESS: taking advantage of high throughput cloning and expression
1121   strategies for the post-genomic analysis of marine organisms., Microb. Cell Fact. 9
1122   (2010).

1123 [203] L.R. Lynd, W.H. van Zyl, J.E. McBride, M. Laser, Consolidated bioprocessing of
1124   cellulosic biomass: an update, Curr. Opin. Biotechnol. 16 (2005) 577-583.

1125 [204] L.R. Lynd, P.J. Weimer, W.H. van Zyl, I.S. Pretorius, Microbial cellulose utilization:
1126   Fundamentals and biotechnology, Microbiol. Mol. Biol. Rev. 66 (2002) 506-577.

1127   **Table Legend**

1128

1129   **Table 1** – Summary of classification data related to the principal arabinoxylan-degrading
1130   enzymes

1131

1132   **Figure legends**

1133

1134   **Figure 1. -** Schematic representation of arabinoxylan structures found in graminaceous plants.

1135   **Figure 2** - Subsite tolerance and hydrolytic capability of GH10 and GH11 xylanases on

1136   ramified arabinoxylan substrates. Arrows indicate potential cleavage sites and circles

1137   represent β-1,4-linked D-xylosyl subunits.

1138   **Figure 3** - Overview of the initial steps of the phosphorylative pentose utilization pathways

1139   classically employed by bacteria and fungi.

1140   **Figure 4** - Overview of the steps in the alternative non phosphorylative pentose utilization

1141   pathway found in certain archaebacteria and bacterial species.

Table 1

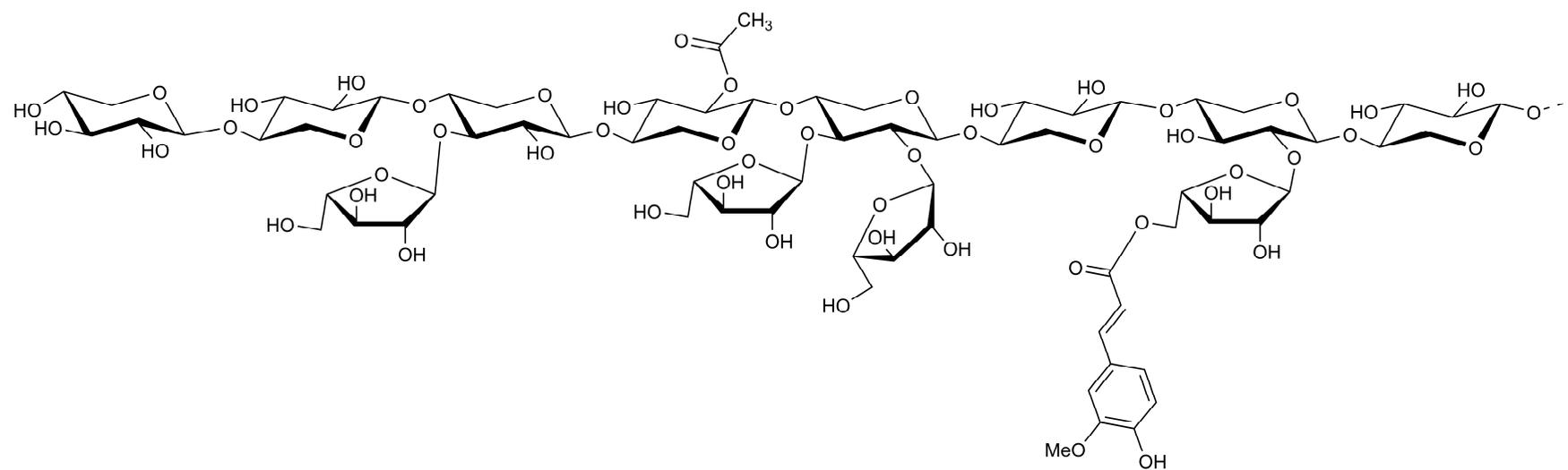| Enzyme	Abbreviation	EC Activities	Mechanism	CAZy Family	GH Clan	Fold
endo-1,4-β-xylanase (xylanase)	Xyn	3.2.1.8	Retaining	GH 5	A	(β/α)$_8$
				GH 10	A	(β/α)$_8$
				GH 11	C	β-jelly roll
			Inverting	GH 8	M	(β/α)$_8$
exo-1,4-β-xylosidase (β-xylosidase)	Xyl	3.2.1.37	Retaining	GH 3	Unk	(β/α)$_8$
				GH 30	A	(β/α)$_8$
				GH 39	A	(β/α)$_8$
				GH 52	Unk	Unk
				GH 54	Unk	β-jelly roll
				GH 116	Unk	Unk
			Inverting	GH 43	F	5-fold β-propeller
			Unk	GH 120	Unk	Unk
α-L-arabinofuranosidase (arabinoxylan arabinofuranohydrolases)	Abf (AXH)	3.2.1.55	Retaining	GH 3	Unk	(β/α)$_8$
				GH 51	A	(β/α)$_8$
				GH 54	Unk	β jelly roll
			Inverting	GH 43	F	5-fold β-propeller
			Unk	GH 62	Unk	Unk
Feruloyl esterases	Fae	3.1.1.73	na	CE 1	na	(α/β/α)-sandwich

Unk: Not Known

na: not applicable

Figure 1

Figure 2

Figure 3

Figure 4