



HAL
open science

Exploration de secrétomes d'*Aspergillus* spp. en vue de la complémentation du cocktail cellulolytique de

Trichoderma reesei

Camille Filiatrault Chastel

► To cite this version:

Camille Filiatrault Chastel. Exploration de secrétomes d'*Aspergillus* spp. en vue de la complémentation du cocktail cellulolytique de *Trichoderma reesei*. Biotechnologie. Institut agronomique, vétérinaire et forestier de France, 2019. Français. NNT : 2019IAVF0021 . tel-02790951

HAL Id: tel-02790951

<https://hal.inrae.fr/tel-02790951>

Submitted on 5 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2019IAVF0021

THESE DE DOCTORAT

préparée à l'Institut des sciences et industries du vivant et de l'environnement (AgroParisTech)

pour obtenir le grade de

Docteur de l'Institut agronomique vétérinaire et forestier de France

Spécialité : Biotechnologies

École doctorale n° 581

Agriculture, alimentation, biologie, environnement et santé (ABIES)

par

Camille FILIATRAULT CHASTEL

**Exploration de secrétomes d'*Aspergillus* spp. en vue de la
complémentation du cocktail cellulolytique de *Trichoderma reesei***

Directeurs de thèse : Antoine MARGEOT et Jean-Guy BERRIN

Co-encadrement de la thèse : Senta BLANQUET

Thèse présentée et soutenue à Rueil-Malmaison, le 4 novembre 2019

Composition du jury :

M. Philippe SILAR, Professeur, Université Paris-Diderot

Mme Isabelle BENOIT, Assistant Professor, Université Concordia

Mme Claire DUMON, Chargée de recherche, INRA

Mme Solange MORERA, Directrice de Recherche, CNRS

M. Gabriel PAES, Directeur de Recherche, INRA

Mme Senta BLANQUET, Chercheuse, IFP Energies Nouvelles

M. Antoine MARGEOT, Chercheur, IFP Energies Nouvelles

M. Jean-Guy BERRIN, Directeur de recherche, INRA

Président du jury

Rapporteuse

Rapporteuse

Examinatrice

Examineur

Co-encadrante de thèse

Directeur de thèse

Co-directeur de thèse

IFP Energies Nouvelles

1-4 avenue de Bois Préau, 92852 Rueil Malmaison

UMR Biodiversité et biotechnologies fongiques (INRA/AMU)

163 Avenue de Luminy, CP925, 13288 Marseille, Cedex 9

Remerciements

Cette thèse n'a bien évidemment pas été une aventure solitaire, et je tiens à remercier dans ces premières pages toutes les personnes qui m'ont aidée, accompagnée et soutenue au cours de ces trois années.

En premier lieu, bien sûr, mes deux directeurs de thèse Jean-Guy Berrin et Antoine Margeot, ainsi que mon encadrante Senta Blanquet. J'ai eu peur au début de ma thèse que le fait d'avoir trois encadrants dans deux institutions différentes soit difficile à gérer, mais ça a été tout le contraire : nos réunions et nos discussions se sont toujours révélées fertiles et ces différents échanges m'ont permis d'avancer en sachant que nous étions sur la même longueur d'onde. Je vous remercie tous les trois pour votre gentillesse et votre disponibilité, ainsi que pour vos conseils et vos encouragements qui m'ont aidé à améliorer ma confiance en moi et en mes résultats.

J'ai aussi eu la chance de travailler au cours de ma thèse dans deux laboratoires et avec deux équipes aux fonctionnements distincts, mais où je me suis sentie intégrée et soutenue du début à la fin de ma thèse.

Je remercie de tout cœur Mireille, David et Sacha qui depuis le tout début m'ont aidé de nombreuses fois dans mes manip et qui ont toujours été là pour répondre à mes questions (et mettre de l'ambiance dans le bureau !).

Je remercie tous mes collègues du laboratoire BBF pour m'avoir si bien accueillie. Si j'ai autant apprécié mes premières années de thèse à Marseille ce n'est pas uniquement grâce à la mer et au soleil, mais aussi grâce à l'ambiance que j'y ai trouvée : merci à tous pour les moments que nous avons partagés au labo, les pauses mots fléchés, les repas au soleil, les discussions autour d'un café et les pique-niques d'équipe. Et comme il n'y a pas que le labo dans la vie, merci aussi à Alexandra, Aurore, Kristian, David, Hayat, Bassem et les autres pour nos soirées marseillaises (au Black Stone ou ailleurs) et nos excursions dans cette si belle région.

Changer de laboratoire après deux ans de thèse n'a pas été chose facile, mais je ne regrette pas un seul instant d'être « remontée à Paris » et d'avoir fait la connaissance de mes collègues de l'IFPEN. Merci à tous ceux qui m'ont aidé à m'y retrouver dans ce nouvel environnement, et en particulier à Céline pour son aide précieuse avec les fermenteurs, et à Véronique pour son expertise qui m'a aidé maintes fois à dompter l'ICS. Je remercie également Paul pour l'aide qu'il m'a apportée pendant son stage de master, ainsi que Simon pour les nombreux échanges que nous avons eu autour des LPMO et pour ces fameuses journées RPE que nous avons partagées. Merci aussi à Francois, Vincent, Rémi, Sarah, Teddy, Tamiris, Emilie, Loïck et les autres, pour m'avoir motivé à aller à la salle de sport le midi, avoir supporté mes plaintes pendant notre fameuse pause de 16h, et m'avoir aidé à me changer les idées lors de nos nombreux afterworks, sessions de yoga, sorties au resto et autres soirées dégustation de vin !

Outre mes collègues de l'INRA et de l'IFPEN, j'ai eu la chance de collaborer avec de nombreuses personnes qui m'ont permis de faire appel à des techniques très différentes et d'élargir mes connaissances dans des domaines variés : merci à Didier Chevret de la plateforme PAPPSO (INRA Jouy-en-Josas) pour l'analyse protéomique des sécrétomes, à Pascal Mansuelle de la plateforme IMM (CNRS Marseille) pour le séquençage N-terminal et l'analyse de masse totale des protéines, à Mathieu Fanuel de la plateforme BIBS (INRA Nantes) pour l'analyse des produits oxydés en spectrométrie de masse, à Bernard Henrissat et toute l'équipe de la base de données CAZy pour leur aide avec l'analyse bioinformatique de la famille AA16, à Leila Lo Leggio et Kristian Frandsen de l'Université de Copenhague pour leur travail sur la caractérisation structurale des LPMO, et enfin à Nadia Touati et Laurent Binet de l'Institut de Recherche de Chimie Paris pour les expériences de RPE.

Enfin, je voudrais remercier ma famille et mes amis, qui ont toujours été là pour moi et qui ont été d'un soutien particulièrement important pendant ces 3 ans.

Merci à Pierre pour le soutien moral en toutes circonstances malgré la distance. Merci à Natalia, Martin, Kieran, Thomas, Charlie, Luc, Bruno, Florian et Victor pour ces deux voyages inoubliables en Canada et en Colombie, et pour tous nos autres séjours un peu partout en France et Europe. Merci à Adèle, Louise, Bastien et Fanny pour les soirées parisiennes, et à Amandine, Laure, Mathilde et Bonnard pour les weekends dans vos contrées respectives. Merci à mes copines de prépa pour nos week-ends de retrouvailles et pour ce super séjour en Roumanie.

Je remercie mes parents d'avoir cru en moi et de m'avoir toujours soutenu dans tout ce que j'ai fait, même si j'ai parfois dû leur faire peur avec certains de mes choix. Je remercie également mes grands-parents d'avoir orchestré mes nombreux déménagements et de m'avoir accueilli chez eux dès que j'en avais besoin. Merci à ma mère et à ma sœur pour leur visite en France qui m'a permis de prendre des vacances bien méritées avant ma soutenance, et merci à toute ma famille de m'avoir aidé à organiser mon pot de thèse et d'être venu me soutenir le jour de ma soutenance. Pour finir, un énorme merci à Alexandre, qui m'a soutenu sans failles pendant 3 ans et qui m'a aidé à surmonter les moments difficiles des derniers mois.

Le rire n'est pas un ennemi de l'apprentissage
- Walt Disney

Sommaire

INTRODUCTION GÉNÉRALE	1
CHAPITRE I : ETUDE BIBLIOGRAPHIQUE	3
1 De la biomasse lignocellulosique au bioéthanol	4
1.1 Structure et composition de la lignocellulose	4
1.1.1 Organisation de la paroi végétale	4
1.1.2 La cellulose.....	5
1.1.3 Les hémicelluloses.....	7
1.1.4 Les pectines	8
1.1.5 Les lignines.....	9
1.2 Utilisation de biomasse lignocellulosique pour la bioraffinerie.....	10
1.2.1 Sources de biomasse lignocellulosique	10
1.2.2 Concept de la bioraffinerie lignocellulosique.....	11
1.3 Procédés de production de bioéthanol.....	12
2 Enzymes impliquées dans la dégradation des polysaccharides	15
2.1 Principes de la classification CAZy	15
2.2 Les cellulases.....	16
2.3 Les hémicellulases.....	19
2.4 Les LPMO	20
2.4.1 Découverte de leur activité	20
2.4.2 Classification actuelle.....	21
2.4.3 Structure tridimensionnelle des LPMO	24
2.4.4 Mécanismes proposés	25
2.4.5 Partenaires redox	27
2.4.6 Utilisation des LPMO dans les cocktails cellulolytiques	28
3 Utilisation de secrétomes fongiques pour l'hydrolyse de biomasse lignocellulosique	30
3.1 Classifications et modes de vie des champignons.....	30
3.2 Utilisation du secrétome de <i>Trichoderma reesei</i>	32
3.2.1 Système cellulolytique de <i>Trichoderma reesei</i>	32
3.2.2 Etudes transcriptomiques et protéomiques	34
3.2.3 Ingénierie génétique des souches	36
3.2.4 Amélioration des enzymes.....	36
3.2.5 Mise en œuvre de l'hydrolyse enzymatique	38
3.3 Stratégies de supplémentation.....	40
3.3.1 Limitations du secrétome de <i>T. reesei</i>	40
3.3.2 La biodiversité fongique comme source d'inspiration	40
3.3.3 Supplémentation par des secrétomes fongiques	41
4 Enjeux et stratégie du projet de recherche.....	42

CHAPITRE II : MATÉRIEL ET MÉTHODES	44
1 Préparation des secrétomes.....	45
1.1 Souches fongiques.....	45
1.2 Milieux de culture	45
1.3 Culture et récolte des secrétomes	45
2 Tests de saccharification en supplémentation.....	46
2.1 Substrats utilisés.....	46
2.2 Cocktail enzymatique de référence	46
2.3 Tests de saccharification en tubes	46
2.4 Tests de saccharification en microplaques	46
2.5 Quantification des produits et calcul des rendements	47
3 Analyse du contenu protéique des secrétomes	48
3.1 Identification des protéines par LC-MS/MS	48
3.2 Analyse des données	48
4 Analyse bioinformatique	49
5 Production de protéines recombinantes	50
5.1 Construction et transformation des vecteurs d'expression.....	50
5.2 Production des protéines	50
5.2.1 Milieux de culture.....	50
5.2.2 Production en plaques Deepwell	51
5.2.3 Production en fioles	51
5.2.4 Production en bioréacteurs	51
5.3 Purification des protéines	52
5.3.1 Purification par chromatographie d'affinité	52
5.3.2 Purification par chromatographie d'échange d'anions.....	52
5.4 Caractérisation biochimique des protéines produites.....	53
5.4.1 Analyse ICP-MS.....	53
5.4.2 Séquençage N-terminal.....	53
5.4.3 Analyse de masse globale.....	53
5.4.4 Tests Amplex Red	53
6 Caractérisation de l'activité des LPMO	54
6.1 Substrats utilisés.....	54
6.2 Tests de dégradation de substrats	54
6.3 Tests de synergies sur cellulose	54
6.4 Analyse des produits solubles de dégradation	55
6.4.1 HPAEC-PAD.....	55
6.4.2 Spectrométrie de masse	55
6.5 Spectroscopie du site actif.....	56
6.6 Modélisation des structures 3D par homologie.....	56

CHAPITRE III : IDENTIFICATION D'UNE NOUVELLE FAMILLE DE LPMO DANS DES SÉCRÉTOMES FONGIQUES	57
1 Résumé	58
2 Background	59
3 Results	61
3.1 Exploration of fungal secretomes to improve biomass saccharification	61
3.2 Comparative proteomic analysis of fungal secretomes	64
3.3 Bioinformatic analysis of a new LPMO family	65
3.4 Heterologous expression and purification	68
3.5 Substrate specificity and regioselectivity of cleavage	69
3.6 Synergy assays	73
4 Discussion	73
5 Conclusions	75
CHAPITRE IV : CARACTÉRISATION DE TROIS NOUVELLES LPMO DE LA FAMILLE AA16	76
1 Résumé	77
2 Background	78
3 Results and discussion	79
3.1 Selection and production of three novel AA16 enzymes	79
3.2 Activity of AA16 enzymes on cellulose	84
3.3 Redox partners	86
3.4 Synergy assays	87
3.5 Spectroscopic analysis of the active site	87
3.6 Structural data	89
4 Conclusions	91
CHAPITRE V : DISCUSSION GÉNÉRALE ET CONCLUSIONS	92
1 Discussion générale	93
1.1 Exploration des sécrétomes et choix des cibles enzymatiques	93
1.2 Production et purification des LPMO AA16	102
1.3 Particularités de l'activité des AA16	104
2 Conclusions et perspectives	108
ANNEXES	110
RÉFÉRENCES	114

Index des figures

Figure 1 : Illustration des procédés d'obtention de biocarburants de 1ère et 2ème génération à partir de biomasse végétale	1
Figure 2 : Structure de la paroi végétale	4
Figure 3 : Structure chimique et organisation de la cellulose.....	6
Figure 4 : Représentation de différents allomorphes de la cellulose.....	6
Figure 5 : Représentation des différentes structures des hémicelluloses.....	8
Figure 6 : Représentation schématique et simplifiée des différents domaines de la pectine.....	8
Figure 7 : Composants structurels de la lignine.....	9
Figure 8 : Exemples de produits issus d'une bioraffinerie lignocellulosique.....	12
Figure 9 : Schéma des procédés de production de bioéthanol à partir de biomasse lignocellulosique	13
Figure 10 : Représentation schématique de l'action des cellulases sur la cellulose.....	17
Figure 11 : Exemples de structure 3D de cellulases, montrant trois types de sites actifs.....	17
Figure 12 : Comparaison des deux principaux mécanismes d'hydrolyse des liaisons glycosidiques	18
Figure 13 : Représentation schématique de différents types d'hémicellulases.....	19
Figure 14 : Régiosélectivité de l'oxydation des liaisons glycosidiques par les LPMO	23
Figure 15 : Structure générale des LPMO et interactions avec leurs substrats.....	24
Figure 16 : Structure tridimensionnelle de LPMO de différentes familles.....	25
Figure 17 : Mécanismes proposés pour l'oxydation de la liaison glycosidique par les LPMO	26
Figure 18 : Interactions des LPMO avec différents systèmes de transfert d'électrons.....	27
Figure 19 : Représentation de la synergie entre LPMO, hydrolases et cellobiose déshydrogénase.....	29
Figure 20 : Phylogénie du règne <i>Fungi</i> (ou <i>Eumycota</i>).....	31
Figure 21 : Illustration de la généalogie de quelques souches mutantes de <i>Trichoderma reesei</i>	33
Figure 22 : Expression des gènes codant pour différentes familles de CAZymes par <i>T. reesei</i> Rut C30 cultivée sur glucose, sophorose et glucose	35
Figure 23 : Différentes approches d'ingénierie pouvant être appliquées aux enzymes de <i>T. reesei</i>	37
Figure 24 : Déroulement typique de l'hydrolyse enzymatique de la cellulose.....	38
Figure 25 : Schéma d'une plaque utilisée pour les tests de saccharification miniaturisés.	47
Figure 26 : Hydrolysis of three pretreated lignocellulosic substrates by a cellulolytic cocktail of <i>Trichoderma reesei</i> strain CL847 produced on lactose	62
Figure 27 : Biomass saccharification performances of 15 <i>Aspergillus</i> spp. secretomes	63
Figure 28 : Protein content of the <i>Aspergillus</i> spp. secretomes	64
Figure 29 : Phylogeny and modularity of the new LPMO family	66
Figure 30 : Graphical representation of AA16 module consensus amino acids, based on the alignment of 213 sequences, generated using the WebLogo application.	67
Figure 31 : Time profiles of dissolved oxygen, temperature and optical density at 600 nm during the production of AaAA16 by <i>Pichia pastoris</i> in a 1.3-L bioreactor	68
Figure 32 : HPAEC-PAD chromatograms of PASC soluble degradation products, after incubation with <i>P. pastoris</i> bioreactor supernatant or with AaAA16 protein purified by affinity chromatography using a nickel HisTrap column	69
Figure 33 : SDS-PAGE analysis of AaAA16 before and after purification.	70
Figure 34 : Production of H ₂ O ₂ by AaAA16 in the presence of ascorbate or L-cysteine.....	70
Figure 35 : HPAEC-PAD chromatograms showing soluble products generated from cellohexaose using AaAA16 and 1 mM L-cysteine.....	71

Figure 36 : Oxidative cleavage of cellulose by AaAA16	71
Figure 37 : Synergy between AaAA16 and CBHI for the degradation of cellulosic substrates.....	73
Figure 38 : HPAEC-PAD chromatograms of PASC soluble degradation products, after incubation with AaAA16 alone or using PaCDHB or L-cysteine as electron donors	74
Figure 39 : Modularity and phylogeny of the produced AA16 LPMOs.....	80
Figure 40 : Time profiles of dissolved CO ₂ and protein production during the production of AjAA16 in a <i>Pichia pastoris</i> bioreactor	81
Figure 41 : Purification of AA16 enzymes	82
Figure 42 : HPAEC-PAD chromatograms showing soluble products generated from PASC using AA16 enzymes and L-cysteine.....	84
Figure 43 : Time course of cellulose degradation using AjAA16	85
Figure 44 : Influence of different electron donor on the products released by AaAA16.	86
Figure 45 : Improved cellulose degradation using TrCBHI after a pretreatment with AA16 enzymes.	87
Figure 46 : Continuous wave X-band EPR spectra of ApAA16 in water, recorded before and 20 h after addition of 100 equivalents of ascorbate.....	88
Figure 47 : Examples of crystals obtained with AaAA16 and AjAA16.....	89
Figure 48 : Plot of the predicted disorder in the AaAA16 sequence	89
Figure 49 : 3D structures of AA16 enzymes modeled using Phyre2 server.....	90
Figure 50 : Diagrammes de Venn représentant le nombre de protéines identifiées dans chaque type de sécrétome, classés selon l'inducteur sur lequel ils ont été produits ou les souches les ayant produits.....	94
Figure 51 : Double classification hiérarchique des familles de CAZymes identifiées dans les sécrétomes étudiés, en fonction de leur abondance	96
Figure 52 : Performances de saccharification de <i>Miscanthus</i> prétraité par des sécrétomes seuls et en supplémentation du cocktail de référence de <i>T. reesei</i>	101
Figure 53 : Effet de différents tampons et pH sur l'activité de la LPMO AaAA16, en présence ou non de cystéine.	105

Index des tableaux

Tableau 1 : Exemples de matières premières lignocellulosiques.....	10
Tableau 2 : Composition chimique de la paroi cellulaire de différents végétaux.....	11
Tableau 3 : Récapitulatif des principales caractéristiques des familles de LPMO existantes	22
Tableau 4 : Enzymes caractérisées de <i>T. reesei</i> ayant une activité sur la cellulose.....	34
Table 5 : Protein concentration (determined by BCA assay) of the secretomes of five <i>Aspergillus</i> spp. strains grown on three inducers, after seven days of culture.....	61
Table 6 : Main features of the produced LPMOs	83
Tableau 7 : Conditions de prétraitement et composition des trois biomasses modèles utilisées.....	97
Tableau 8 : Protéines plus abondantes dans les sécrétomes améliorant la saccharification de la paille à 24h que dans les 13 autres.	100
Tableau 9 : Protéines plus abondantes dans les sécrétomes améliorant la saccharification du <i>Miscanthus</i> à 24h que dans ceux n'apportant aucune amélioration.....	100

Liste des abréviations

3-HAA: acide 3-hydroxyanthranilic
AA : activité auxiliaire
AAQO : aryl-alcool quinone oxidoréductases
ADN : acide désoxyribonucléique
AGU: unité anhydroglucose
ARN : acide ribonucléique
ASM : autoclavé de son de maïs (ou MB, *Maize Bran*)
Avi : Avicel
BCA : acide bicinchoninic
BG : β -glucosidase
BSA : albumine sérique bovine
CAZyme : enzyme active sur les carbohydrates (*Carbohydrate Active enZyme*)
CBH : cellobiohydrolase
CBM : module de liaison aux carbohydrates (*Carbohydrate Binding Module*)
CBP : bioprocédé consolidé (*Consolidated Bioprocess*)
CDH : cellobiose déshydrogénase
CE : carbohydrate estérase
CIRM-CF : Centre International de Ressources Microbiennes - Champignons Filamenteux
Dha : acide docosaénoïque
DP : degré de polymérisation
EDS : systèmes de transfert d'électrons (*Electron Donating System*)
EG : endoglucanase
EMS : sulfonate d'éthyle méthane
ESI-MS : spectrométrie de masse par ionisation par nébulisation électrostatique (*Electrospray Ionization Mass Spectrometry*)
GDH : glucose déshydrogénase
GH : glycosyl-hydrolase
GPI : glycosylphosphatidylinositol
GT : glycosyl-tranférase
HMM : modèle de Markov caché (*Hidden Markov Model*)
HPAEC-PAD : chromatographie haute-performance par échange d'anions avec détection ampérométrique (*High Performance Anion Exchange Chromatography coupled with Pulsed Amperometric Detection*)

ICP-MS: spectrométrie de masse à plasma à couplage inductif (*Inductively Coupled Plasma Mass Spectrometry*)

IMAC : chromatographie d'affinité sur ions métalliques immobilisés (*Immobilized Metal ion Affinity Chromatography*)

Kdo : acide 3-désoxy-D-manno-oct-2-ulosonique

LC-MS/MS : chromatographie liquide couplée à de la spectrométrie de masse tandem (*Liquid Chromatography coupled to tandem Mass Spectrometry*)

LPMO : *Lytic Polysaccharide Monoxygenase*

MALDI-ToF : analyse de temps de vol par désorption-ionisation laser assistée par matrice (*Matrix-Assisted Laser Desorption Ionization Time-of-Flight*)

Man₅GlcNAc : (mannose)₅-N-acétylglucosamine

MS : masse sèche

MS/MS : spectrométrie de masse tandem

NFC : nanofibrilles de cellulose (*Nanofibrillated Cellulose*)

NTA : acide nitrilotriacétique

NTG : N-nitroguanidine

OD600 : densité optique à 600 nm

PAI : indice d'abondance des protéines (*Protein Abundance Index*)

PASC : *Phosphoric Acid Swollen Cellulose*

PCR : amplification en chaîne par polymérase (*Polymerase Chain Reaction*)

PL : polysaccharide lyase

RNA-seq : séquençage de l'ARN

RPE : résonance paramagnétique électronique (ou EPR, *Electronic Paramagnetic Resonance*)

SBP : pulpe de betterave (*Sugarbeet Pulp*)

SDS-PAGE: électrophorèse en gel de polyacrylamide contenant du laurylsulfate de sodium (*Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis*)

SHF : hydrolyse et fermentation séparées (*Separated Hydrolysis and Fermentation*)

SSCF : saccharification et co-fermentation simultanées (*Simultaneous Saccharification and Co-Fermentation*)

TFA : acide trifluoroacétique

UV : ultraviolet

Introduction générale

Dans un contexte d'accroissement permanent de la population et des besoins énergétiques mondiaux, les ressources fossiles traditionnellement utilisées tendent à s'épuiser rapidement, et sont remises en cause du fait de leur fort impact écologique. La réduction de la dépendance à ces ressources fossiles peut notamment passer par le recours à la biomasse, qui constitue une matière première renouvelable disponible en abondance. Celle-ci peut être utilisée pour la génération d'énergie ou pour la fabrication de divers produits de consommation dans les domaines de la santé, de l'alimentation humaine et animale, des matériaux, de la chimie etc., remplaçant ainsi la plupart des produits issus actuellement du raffinage du pétrole.

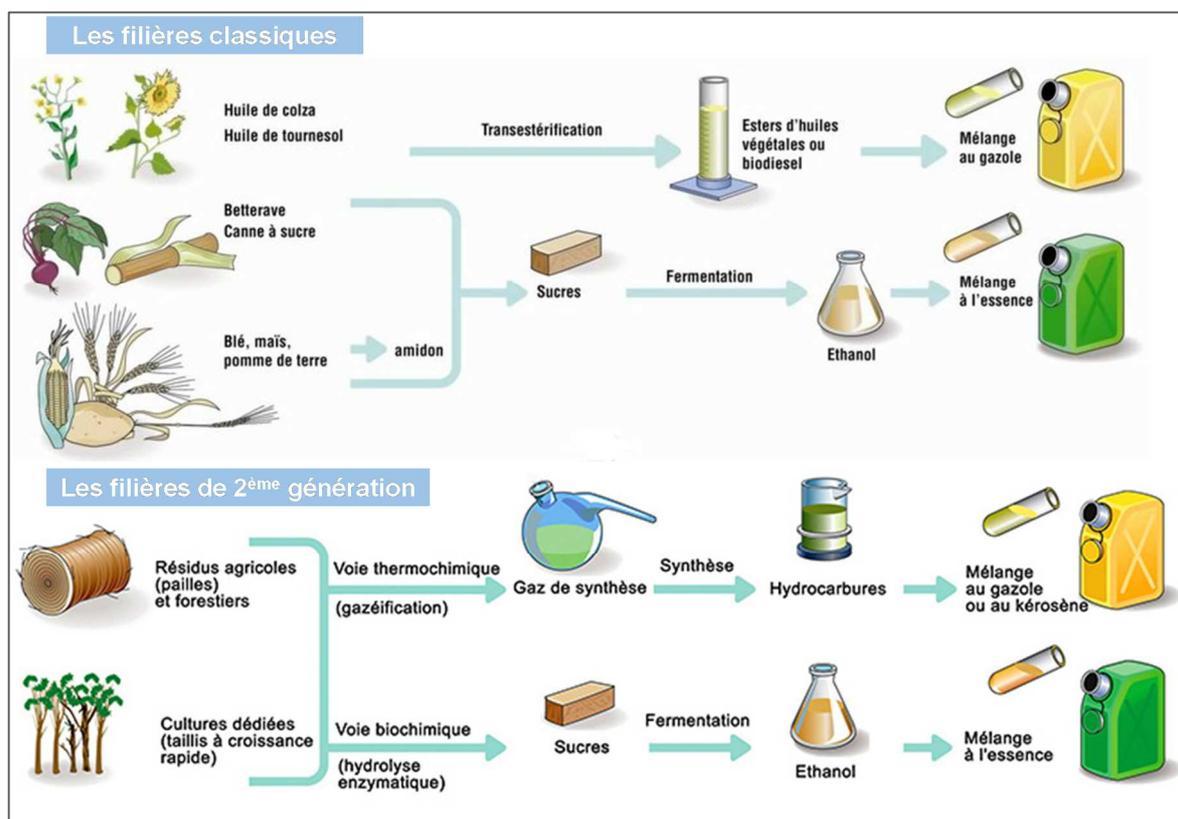


Figure 1 : Illustration des procédés d'obtention de biocarburants de 1ère et 2ème génération à partir de biomasse végétale (source : IFP Energies Nouvelles)

Parmi les secteurs concernés par cette transformation, celui des transports représente une part importante de l'énergie consommée et du CO₂ émis au niveau mondial. Une solution envisagée pour réduire les émissions polluantes est l'usage de biocarburants, utilisables en mélange avec les carburants traditionnels (voir Figure 1). Les biocarburants utilisés actuellement sont produits principalement à partir des organes de réserve de plantes oléagineuses (tournesol, colza, soja, palme), sucrières (betteraves, canne à sucre) et amylacées (maïs, blé). Ces biocarburants de 1ère génération sont remis en question sur plusieurs points, d'une part car ils entrent en compétition avec l'alimentation humaine et animale, et d'autre part car leur production à grande échelle

nécessite une consommation d'eau importante et l'utilisation de larges surfaces agricoles ne pouvant être gagnées dans certains pays qu'au prix d'une déforestation accrue.

De nombreux efforts tendent donc au développement de biocarburants, dits de 2^{ème} génération, obtenus à partir de plantes entières ou de leurs fractions non comestibles. Cette biomasse lignocellulosique, extrêmement abondante peu valorisée actuellement, est cependant plus récalcitrante à la dégradation que les matières premières précédemment citées. Elle peut être transformée soit par voie physico-chimique en hydrocarbures de synthèse, nécessitant une gazéification dans des conditions de température très élevées (>1000°C), soit par voie biologique en dégradant les polysaccharides de la paroi végétale en sucres qui peuvent ensuite être fermentés. Cette hydrolyse de la lignocellulose est possible grâce à l'utilisation d'enzymes produites par des micro-organismes dégradant naturellement les polysaccharides pariétaux.

Les champignons filamenteux, capables de se développer dans la nature sur une large gamme de biomasses végétales, sont des organismes de choix pour la production de telles enzymes. Grâce à ses souches hyper-sécrétrices, *Trichoderma reesei* est le producteur de cellulases et d'hémicellulases le plus utilisé dans l'industrie. Cependant, afin de rendre le bioéthanol de 2^{ème} génération compétitif, il est nécessaire de réduire encore les coûts de production, notamment en améliorant les rendements de l'étape d'hydrolyse enzymatique. Le séquençage du génome de *T. reesei* ayant révélé une faible diversité d'enzymes actives sur la lignocellulose, par rapport à d'autres espèces fongiques, des démarches de supplémentation ont été amorcées : il s'agit d'ajouter aux enzymes sécrétées par *T. reesei* des enzymes d'autres espèces, qui apportent des activités complémentaires ou synergiques. La recherche de tels enzymes peut notamment se faire par des approches de sécrétomique, c'est-à-dire l'étude de l'ensemble des protéines sécrétées par une souche dans des conditions données.

Les travaux réalisés dans le cadre de ce projet de thèse s'inscrivent dans cette démarche, en se proposant d'étudier les sécrétomes de différentes souches fongiques, dans le but d'identifier des enzymes capables d'améliorer la dégradation de la biomasse lignocellulosique. Le premier chapitre de ce rapport présente une étude bibliographique, portant d'une part sur la biomasse lignocellulosique et son utilisation dans les procédés biotechnologiques, et d'autre part sur les enzymes utilisées pour sa dégradation et les organismes fongiques qui les produisent, en citant les stratégies envisagées pour améliorer l'efficacité du procédé. Le deuxième chapitre présente le matériel et les méthodes utilisées pendant ce projet. Les chapitres 3 et 4 présentent les résultats obtenus, sous forme d'articles en anglais ; le premier décrit l'étude du contenu et des performances de plusieurs sécrétomes, ainsi que l'identification d'une nouvelle famille d'enzymes, tandis que le deuxième est consacré à la caractérisation de plusieurs enzymes issues de cette famille. Enfin, le cinquième chapitre est constitué d'une discussion portant sur l'ensemble des résultats obtenus, et présente les perspectives envisagées.

Chapitre I : Etude bibliographique

1 De la biomasse lignocellulosique au bioéthanol

1.1 Structure et composition de la lignocellulose

1.1.1 Organisation de la paroi végétale

Les cellules végétales possèdent à l'extérieur de leur membrane plasmique une enveloppe appelée paroi végétale, qui est impliquée dans le maintien de la plante et dans sa résistance à la déformation. Cette paroi est formée au cours du développement de la plante par le dépôt successif de plusieurs structures distinctes : la lamelle moyenne, la paroi primaire, et la paroi secondaire, constituée de trois couches (Figure 2A).

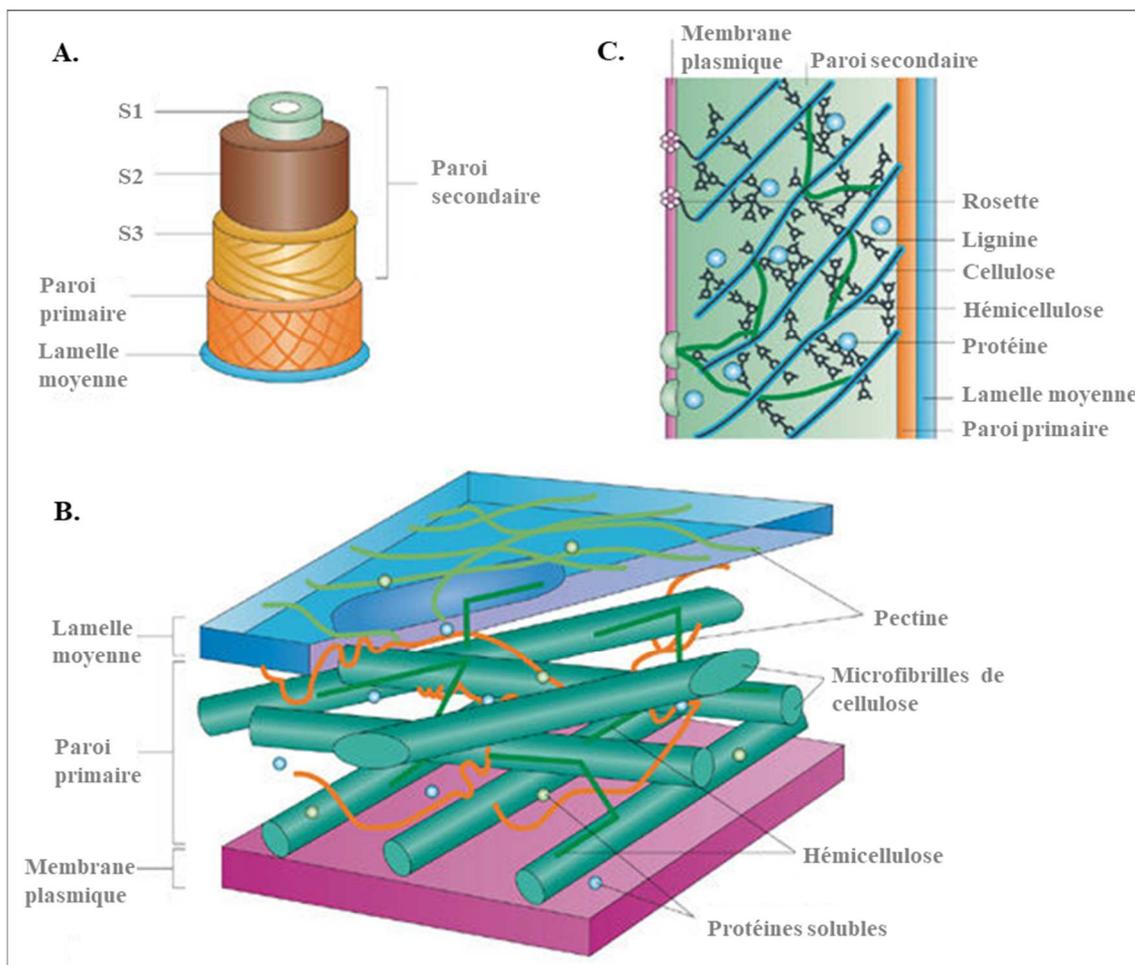


Figure 2 : Structure de la paroi végétale (Sticklen, 2008)

A. Organisation des différentes couches de la paroi végétale. **B.** Composition de la paroi primaire.

C. Composition de la paroi secondaire.

La lamelle moyenne, constituée de pectines et d'hémicelluloses, est située à l'interface entre deux cellules adjacentes. La paroi primaire, qui se forme durant la croissance de toutes les cellules végétales, contient des micro-fibrilles de cellulose emprisonnées dans une matrice de pectines et d'hémicelluloses, accompagnées de protéines (Figure 2B). Lorsqu'elles ont fini de croître, certaines cellules spécialisées développent une paroi secondaire (Figure 2C), où les micro-fibrilles de celluloses sont liées à des chaînes d'hémicelluloses et à des

lignines par des interactions covalentes et non-covalentes (liaisons hydrogènes et forces de van der Waals) ; les couches S1, S2 et S3 qui composent la paroi secondaire diffèrent par l'orientation de leurs fibres de cellulose (Davin et al., 2008). La dernière étape de maturation de certains végétaux est celle de la lignification, au cours de laquelle la lignine remplace peu à peu l'eau qui se trouvait dans les tissus, en commençant au niveau de la lamelle moyenne, jusqu'à la paroi primaire puis la paroi secondaire, formant ainsi une matrice hydrophobe dans laquelle sont emprisonnés les composants cellulosiques et non-cellulosiques ainsi que les protéines (Iiyama et al., 1994). La lignocellulose qui compose la paroi végétale est donc un enchevêtrement de différents composants, dont des polysaccharides (cellulose, hémicelluloses et pectines) et des polymères phénoliques (lignines), qui sont décrits plus précisément dans les paragraphes suivants.

1.1.2 La cellulose

La cellulose est le biopolymère le plus abondant sur Terre, et le composé principal de la paroi végétale. Elle est produite non seulement par les plantes, mais aussi par certains animaux marins, les tuniciers, ainsi que par des algues, des champignons, des bactéries et des invertébrés (Lavoine et al., 2012). Il s'agit d'un homopolymère linéaire de résidus anhydro-D-glucose, reliés entre eux par des liaisons glycosidiques de type β -(1 \rightarrow 4), ce qui le différencie de l'amidon où les résidus D-glucose sont liés en α -(1 \rightarrow 4). Chaque monomère effectuant une rotation de 180° par rapport à ses voisins, l'unité de répétition de ce polymère est un dimère appelé cellobiose. Chaque chaîne de cellulose possède une extrémité réductrice, porteuse d'un hémiacétal sur le carbone anomérique (C1) du résidu de glucose, et une extrémité appelée non-réductrice, porteuse d'une fonction hydroxyle (voir Figure 3A). Le degré de polymérisation d'une chaîne de cellulose peut aller d'environ 8000 dans la paroi primaire jusqu'à plus de 20000 dans la paroi secondaire (Brown, 2004; Habibi et al., 2010).

Dans la nature, les chaînes de cellulose sont assemblées entre elles pour former un ensemble hautement ordonné : approximativement 36 chaînes de cellulose sont assemblées en fibrilles élémentaires, qui elles-mêmes se regroupent en micro-fibrilles, qui s'associent aux autres composants de la paroi pour former des fibres cellulosiques (Figure 3B). Les fibrilles élémentaires sont composées de régions cristallines, où les chaînes sont étroitement liées entre elles par des forces de van der Waals et un réseau complexe de liaisons hydrogène inter- et intramoléculaires ; ces régions hautement ordonnées sont interrompues par des régions amorphes, probablement introduites par des tensions internes dans les fibrilles qui provoquent leur distorsion (Habibi et al., 2010).

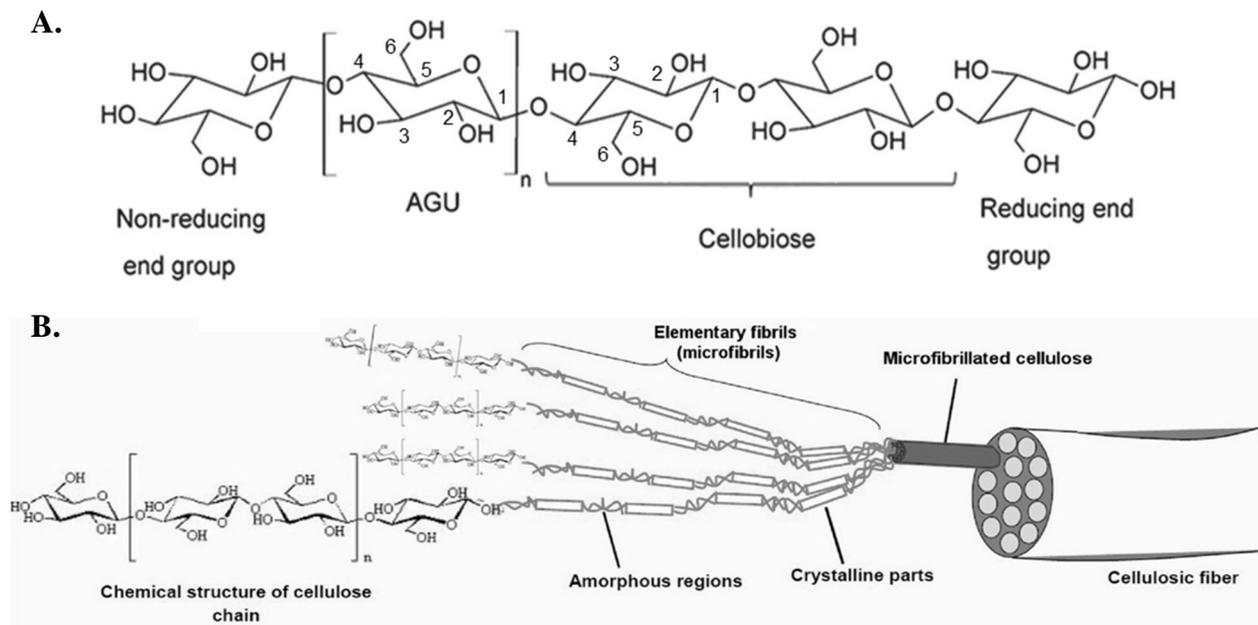


Figure 3 : Structure chimique et organisation de la cellulose

- A.** Structure chimique d'une chaîne de cellulose, montrant l'unité de répétition cellobiose, les extrémités réductrice et non-réductrice, ainsi que la numérotation des atomes de carbone au sein de l'unité anhydroglucose (AGU), montrant le carbone anomérique en position 1.
- B.** Représentation schématique des différents niveaux de structure de la cellulose (Lavoine et al., 2012).

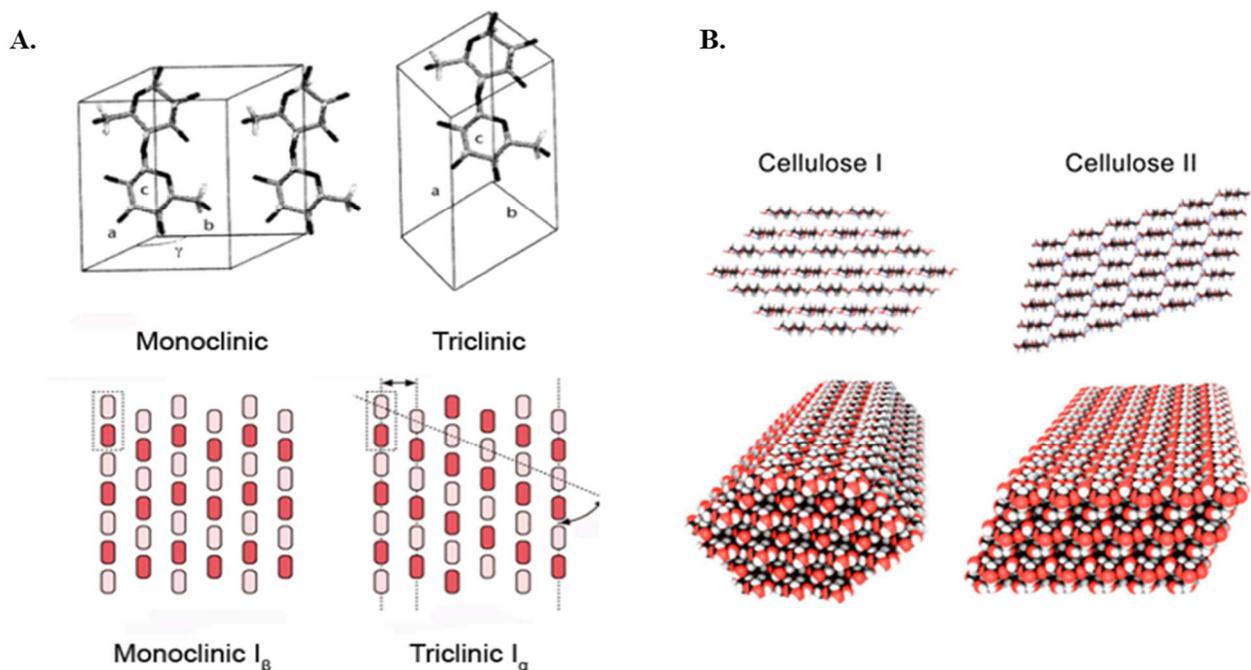


Figure 4 : Représentation de différents allomorphes de la cellulose

- A.** Différences entre les formes monoclinique et triclinique de la cellulose I, en terme de maille cristallographique et d'organisation intra-couche des chaînes de cellulose (Quiroz-Castañeda and Folch-Mallol, 2014).
- B.** Différences d'organisation des couches entre la cellulose I et la cellulose II (Mukarakate et al., 2016).

Les chaînes individuelles peuvent s'organiser entre elles de plusieurs manières pour former les régions cristallines, et il existe en tout six allomorphes différents de cellulose (O'Sullivan, 1997; Quiroz-Castañeda and Folch-Mallol, 2014) :

- dans la cellulose I, la plus abondante dans la nature, les chaînes cellulosiques s'organisent en couches grâce à un réseau bidimensionnel de liaisons hydrogène, et ces couches sont liées les unes aux autres grâce à des interactions plus faibles ; cette cellulose native est en fait un mélange entre deux phases cristallines I_α et I_β , qui diffèrent par l'organisation de leurs liaisons hydrogène intra-couches, résultant soit en une maille triclinique (I_α) soit en une maille monoclinique (I_β) (Figure 4A et 4B) ;
- dans la cellulose II, obtenue à partir de cellulose I soit par un traitement alcalin appelé mercerisation, soit par solubilisation et recristallisation, les chaînes cellulosiques sont réorganisées grâce à un réseau tridimensionnel de liaisons hydrogène intra- et inter-couches (Figure 4B) ;
- les celluloses III_I et III_{II} peuvent être formées respectivement à partir de cellulose I et II, par traitement à l'ammoniac réversible ;
- les celluloses IV_I et IV_{II} sont obtenues respectivement à partir des celluloses III_I et III_{II} par traitement au glycérol à 260 °C.

1.1.3 Les hémicelluloses

Les hémicelluloses constituent un groupe hétérogène de polysaccharides complexes, souvent ramifiés, dont la chaîne principale est composée de xylose, de glucose et/ou de mannose, et qui peuvent porter des substitutions très diverses : hexoses (glucose, mannose, galactose, fucose), pentoses (xylose, arabinose), acides uroniques (acide glucuronique). La composition des hémicelluloses varie énormément selon les groupes de plantes et les types de cellules.

On peut les classer en différents groupes qui sont représentés sur la Figure 5 :

- les xyloglucanes, les plus répandus, sont constitués d'une chaîne de glucose portant des unités xylose, elles-mêmes pouvant être substituées par plusieurs unités galactose, arabinose ou fucose ;
- les glucanes mixtes sont des polymères linéaires liés en β -(1→4), avec quelques liaisons β -(1→3) ;
- les xylanes possèdent une chaîne principale composée de résidus xylose, et peuvent porter différentes substitutions notamment composées d'acide glucuronique (qui peuvent être liés à la lignine lorsqu'ils sont méthylés) et d'arabinose (parfois estérifié par de l'acide férulique) ;
- les mannanes et les glucomannanes comportent des unités mannose liées en β -(1→4) dans leur chaîne principale, parfois alternés avec des unités glucose, et sont substitués ou non par des résidus galactose (galactomannane et galactoglucomannane).

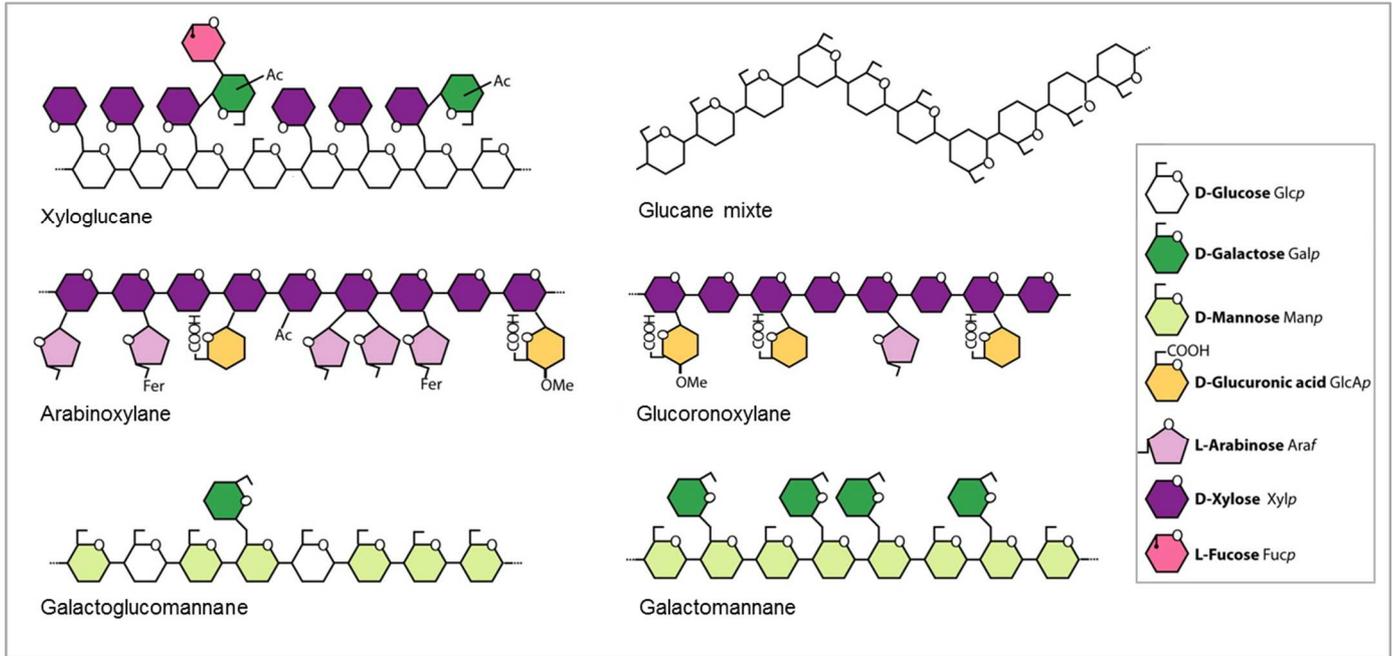


Figure 5 : Représentation des différentes structures des hémicelluloses (Scheller and Ulvskov, 2010)

1.1.4 Les pectines

Les pectines, présentes dans la matrice des parois primaires, sont des hétéropolymères complexes, contenant de l'acide galacturonique, qui forment un gel stable lorsqu'elles s'associent via l'intermédiaire d'ions calcium (Caffall and Mohnen, 2009).

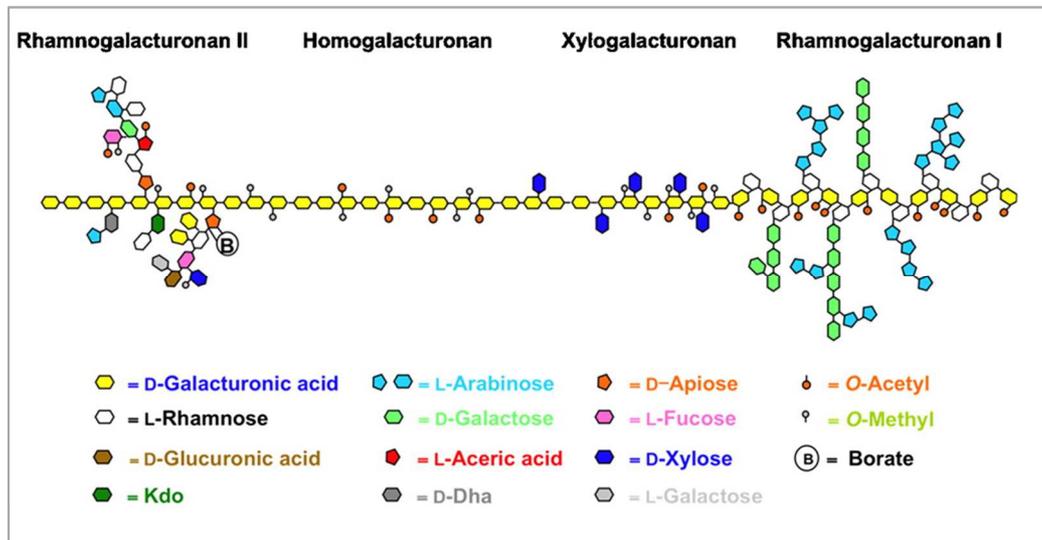


Figure 6 : Représentation schématique et simplifiée des différents domaines de la pectine (Harholt et al., 2010).

Kdo : acide 3-désoxy-D-manno-oct-2-ulosonique ; Dha : acide docosahexaénoïque

Elles contiennent quatre types de régions, représentées sur la Figure 6 :

- les régions homogalacturonanes, les plus abondantes, qui sont constituées d'une chaîne linéaire d'acide D-galacturonique liée en α -(1→4) ;

- les régions xylogalacturonanes, dans lesquelles la chaîne principale d'acide galacturonique est substituée par des résidus xyloses ;
- les régions rhamnogalacturonanes de type I, qui sont des polymères branchés dont la chaîne principale possède une unité de base (α -1,4-D-galacturonate- α -1,2-L-rhamnose), et dont les chaînes latérales sont composées de galactane et d'arabinane ;
- les régions rhamnogalacturonanes de type II, dont la chaîne principale d'acide galacturonique est substituée avec des chaînes latérales très complexes, contenant 12 types de groupements différents.

1.1.5 Les lignines

Les lignines sont, après la cellulose, les substances organiques les plus abondantes dans la nature. Composées d'hétéropolymères hydrophobes et résistants à la compression, elles confèrent aux végétaux leur solidité, et sont présentes dans les cellules des tissus dédiés au transport de l'eau et au soutien des parties aériennes de la plante (Iiyama et al., 1994). Elles sont le produit de la voie métabolique des phénylpropanoïdes, et dérivent principalement de trois monolignols : les alcools coumarylique, coniférylique et sinapylique (voir Figure 7). Ces précurseurs sont à l'origine respectivement des unités *p*-hydroxyphényle (notée H), guaiacyle (G) et syringyle (S), qui sont associées pour former des polymères amorphes et hautement complexes (Davin et al., 2008).

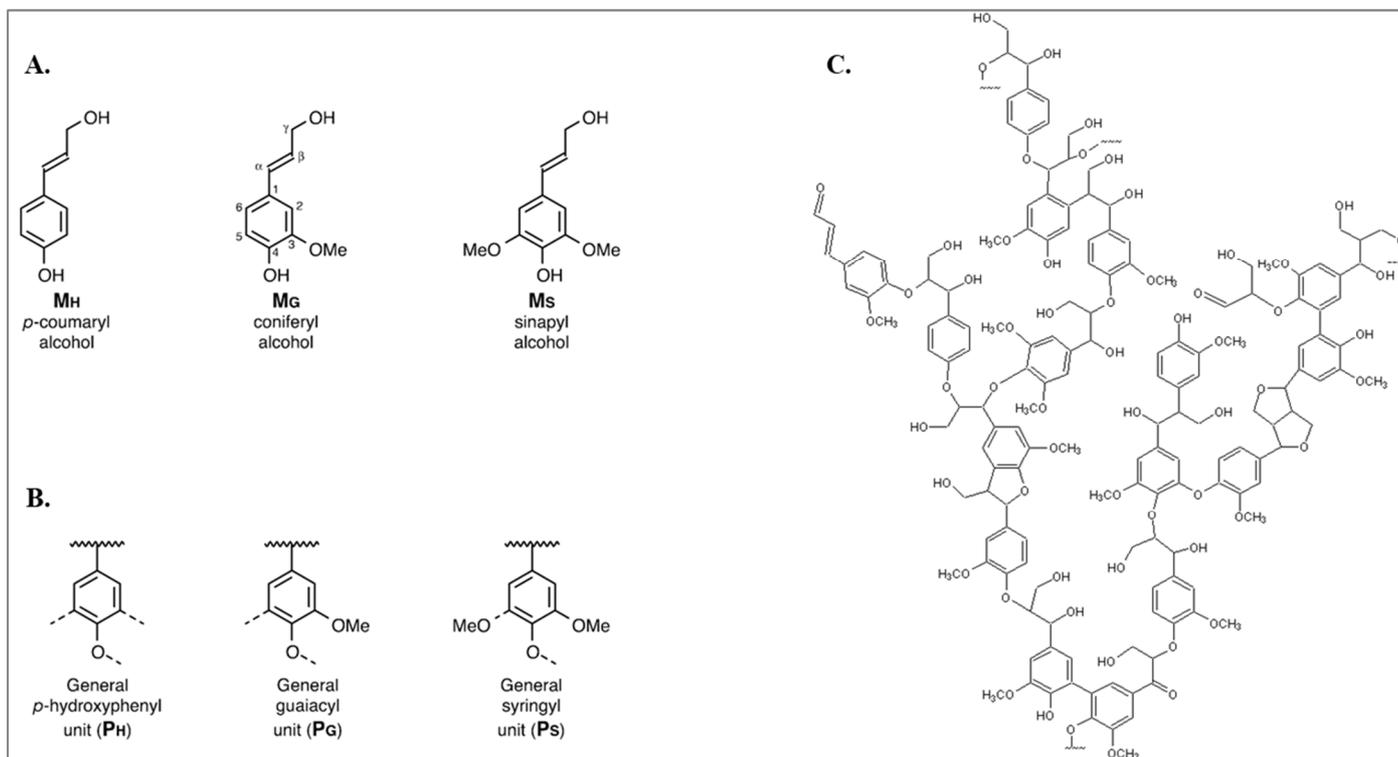


Figure 7 : Composants structurels de la lignine (Adler, 1977; Ralph et al., 2004)

A. Les trois monolignols **B.** Les trois unités composant la lignine **C.** Exemple de structure de lignine (bois tendre)

1.2 Utilisation de biomasse lignocellulosique pour la bioraffinerie

1.2.1 Sources de biomasse lignocellulosique

L'appellation de « biomasse lignocellulosique » est utilisée pour désigner toute matière végétale contenant principalement de la lignocellulose. Cette appellation très large recouvre donc un grand nombre de produits, qui peuvent avoir différentes origines :

- résidus agricoles produits lors de la culture de différentes plantes destinées à l'alimentation ;
- déchets forestiers de bois dur ou de bois tendre, issus de l'industrie du bois ou des pâtes et papiers ;
- plantes énergétiques issues de cultures dédiées (herbes vivaces et taillis à courte rotation) ;
- déchets solides municipaux et industriels.

Des exemples de chacune de ces sources de biomasse sont présentés dans le Tableau 1.

La plupart de ces matières sont considérées comme des déchets ou des sous-produits d'autres industries, et sont donc facilement disponibles, en très grandes quantités, sans accaparer de terres cultivables supplémentaires. Certains de ces sous-produits peuvent être utilisés comme source de chaleur et d'énergie, comme fourrage ou répandus dans les champs pour les fertiliser, mais ils sont aussi parfois simplement jetés ou brûlés ; il apparaît donc avantageux d'en transformer une partie en produits à plus haute valeur ajoutée (Gupta and Verma, 2015; Talebnia et al., 2010). Seules les plantes énergétiques font l'objet d'une culture dédiée, et bien qu'elles ne soient pas utilisées dans l'alimentation ou la production de fibres, elles pourraient être en compétition pour l'usage des terres. Cependant, leur capacité à pousser rapidement sur divers types de terrains, avec des besoins en eau limités ou une bonne tolérance aux basses températures, pourraient permettre leur exploitation sur des terres non-agricoles (García et al., 2014; Ho et al., 2014).

Tableau 1 : Exemples de matières premières lignocellulosiques

(de Bhowmick et al., 2018; García et al., 2014; Gupta and Verma, 2015; Ho et al., 2014)

	Herbes et biomasse non-boisée	Bois dur (feuillus et plantes à fleurs)	Bois tendre (conifères)
Cultures dédiées	Miscanthus Panic érigé (<i>Switchgrass</i>)	Saule Peuplier	
Résidus agricoles	Rafles de maïs Bagasse de canne à sucre Pailles (blé, riz, avoine, orge) Pulpe de betterave	Branches d'oliviers et d'arbres fruitiers Résidus de taille de vigne Tiges de coton	
Résidus forestiers		Ecorce d'eucalyptus	Sciure de pin Branches d'épicéa
Autres déchets	Papiers, cartons Pelures de fruits Boues de papeterie		

Tableau 2 : Composition chimique de la paroi cellulaire de différents végétaux (Rytioja et al., 2014)

La composition est indiquée en pourcentage du poids sec (M : quantités mineures).

	Cellulose	Hémicelluloses				Pectines	Lignine
		Mannane	Xylanes	β -Glucanes	Xyloglucane		
Bois tendre	33–42	10–15	5–11	—	—	—	27–32
Bois dur	38–47	2–5	15–30	—	—	—	21–31
Monocotylédones							
Paroi primaire	20–30	M	20–40	10–30	1–5	5	M
Paroi secondaire	35–45	M	40–50	M	M	M	20
Dicotylédones							
Paroi primaire	15–30	5–10	5	—	20–25	20–30	M
Paroi secondaire	45–50	3–5	20–30	—	M	M	7–10

Les différentes matières premières lignocellulosiques contiennent les polymères pariétaux présentés plus haut, mais leurs compositions et structures varient considérablement selon le type de végétaux dont elles sont issues (Tableau 2). La principale différence de composition chimique entre le bois dur et le bois tendre est liée aux hémicelluloses : les galactoglucomannanes sont majoritaires dans le bois tendre (pins, épicéa...) alors que le bois dur (bouleau, chêne, peuplier...) contient majoritairement des glucuronoxylanes. Les monocotylédones (herbes et graminées) se différencient des dicotylédones (plantes à fleurs) par une fraction hémicellulosique plus pauvre en mannanes et xyloglucanes, et plus riche en xylanes, ainsi que par une quantité de pectines largement plus faible. Dans les deux cas, la paroi secondaire est enrichie en cellulose et en lignines, mais celle des dicotylédones contient plus de cellulose que celle des monocotylédones, alors que ces dernières contiennent plus de lignines (Rytioja et al., 2014).

1.2.2 Concept de la bioraffinerie lignocellulosique

Les différents types de biomasse lignocellulosique précédemment cités peuvent être utilisés comme matière première et valorisés par différents procédés dans ce qu'on appelle des bioraffineries, où la cellulose, les hémicelluloses et la lignine sont transformées pour obtenir une large palette de molécules plateformes (alcools, phénols, acides organiques, hydrocarbures, produits oxydés...) destinées à remplacer celles issues du raffinage du pétrole (Figure 8). Ces molécules plateformes peuvent ensuite être transformées en produits de consommation dans les domaines de l'industrie, de la santé, de l'alimentation humaine et animale, des matériaux et de l'énergie (Cherubini, 2010; de Bhowmick et al., 2018). Tous les constituants de la biomasse peuvent ainsi être valorisés, à condition de réussir à les atteindre, et donc à surmonter la résistance de la structure complexe formée par la cellulose, les hémicelluloses et la lignine. Les biomasses utilisables ayant des compositions chimiques et structurales très variées, l'un des défis de cette approche consiste à adapter en

fonction des matières premières les prétraitements et procédés utilisés (de Bhowmick et al., 2018; Menon and Rao, 2012).

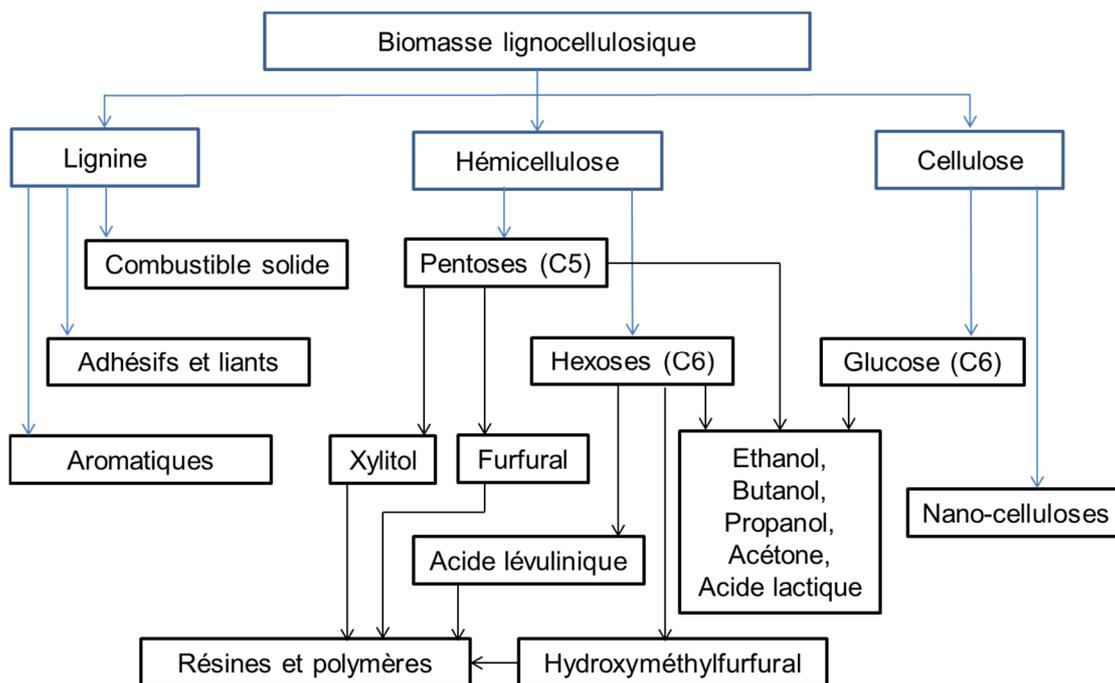


Figure 8 : Exemples de produits issus d'une bioraffinerie lignocellulosique (d'après Villegas and Gnansounou, 2008; Witzler et al., 2018)

Une fois séparées, les différentes fractions peuvent être converties ou fonctionnalisées grâce à un panel varié de procédés thermochimiques et biochimiques. La plupart de ces techniques sont encore en développement, mais devraient à terme permettre l'accès à de nombreux synthons, réduisant ainsi la dépendance aux produits issus du raffinage du pétrole, tout en minimisant les rejets de déchets qui sont transformés en produits à haute valeur ajoutée (Menon and Rao, 2012). Parmi les différents produits issus de la biomasse lignocellulosique, l'un des plus étudiés et des plus avancés industriellement est le bioéthanol, dont le procédé de production est détaillé ci-après.

1.3 Procédés de production de bioéthanol

Le procédé classique d'obtention d'éthanol de deuxième génération à partir de biomasse lignocellulosique se décompose en quatre étapes principales, représentées sur la Figure 9A.

1°) Le prétraitement de la biomasse est nécessaire pour casser la structure de la paroi lignocellulosique, en solubilisant ou en séparant ses différents composants. Il existe différents types de prétraitements :

- des prétraitements physiques (mécaniques ou hydrothermaux), qui permettent la réduction de la taille des particules, la diminution de la cristallinité de la cellulose et la solubilisation d'une partie des hémicelluloses ;

- des prétraitements chimiques qui utilisent des acides, bases ou agents oxydants pour solubiliser les hémicelluloses, retirer la lignine ou en modifier la structure, et parfois diminuer la cristallinité de la cellulose ;
- des prétraitements biologiques, qui utilisent des espèces fongiques capables de dégrader la lignine ou les hémicelluloses.

Certains prétraitements combinent des méthodes physiques et chimiques, comme par exemple l'explosion à la vapeur en présence d'acide dilué ou l'explosion à froid à l'ammoniaque (Gupta and Verma, 2015; Talebnia et al., 2010).

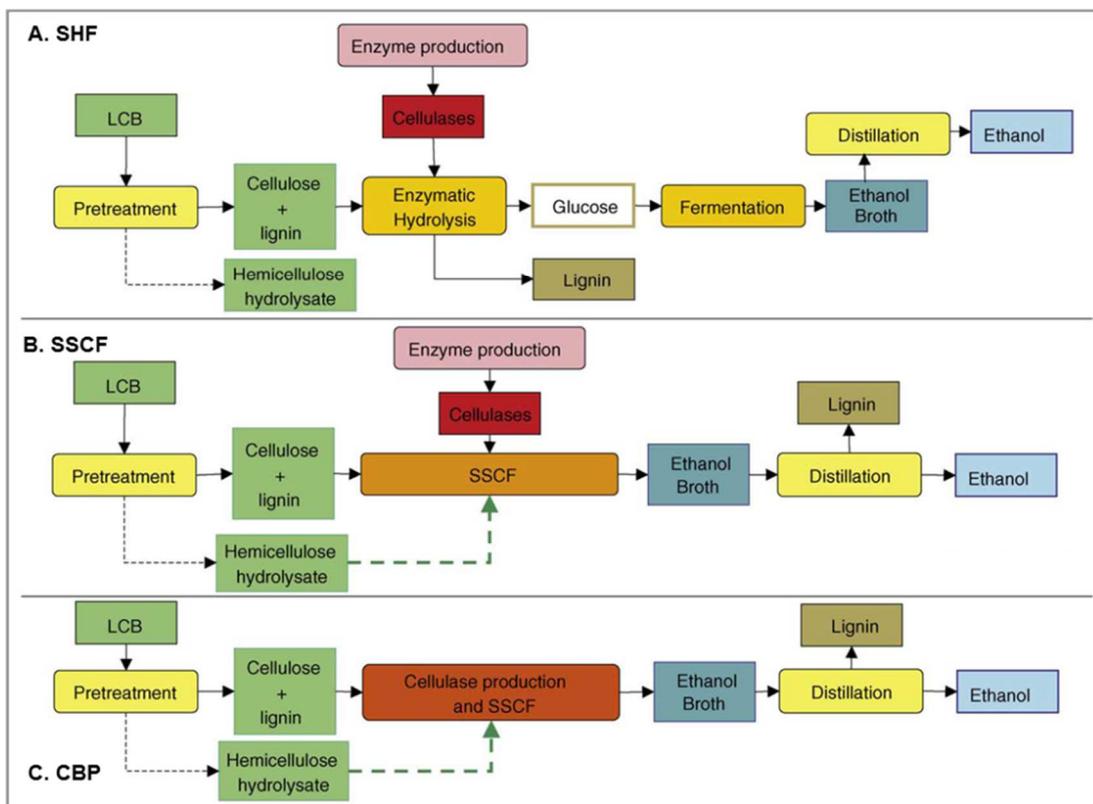


Figure 9 : Schéma des procédés de production de bioéthanol à partir de biomasse lignocellulosique (LCB) (Margeot et al., 2009)

A. Hydrolyse et fermentation séparées (SHF) **B.** Saccharification et co-fermentation simultanées (SSCF)

C. Bioprocédé consolidé (CBP)

2°) L'hydrolyse enzymatique, ou saccharification, consiste à dépolymériser la cellulose et les hémicelluloses restantes en monomères, à l'aide d'enzymes spécifiques. Cette étape d'hydrolyse est donc précédée par celle de la production des enzymes. On peut utiliser pour cela une variété d'organismes cellulolytiques, aérobies ou anaérobies, mésophiles ou thermophiles ; il peut s'agir de bactéries (notamment des genres *Clostridium*, *Thermomonospora*, *Cellulomonas*, *Bacteriodes* ou *Streptomyces*), et plus souvent de champignons filamenteux (*Trichoderma*, *Aspergillus*, *Penicillium* ou *Myceliophthora*,) (Sun and Cheng, 2002). Ce sont ces enzymes et organismes fongiques auxquels on s'intéressera plus particulièrement dans la suite de ce rapport (voir parties 2 et 3 de ce chapitre).

3°) La fermentation par des microorganismes permet d'obtenir de l'éthanol à partir des sucres (pentoses et hexoses) issus de l'étape d'hydrolyse. La levure *Saccharomyces cerevisiae* est le microorganisme le plus utilisé dans l'industrie, grâce à ses forts rendements et productivités, et sa sensibilité modérée aux inhibiteurs. Cependant, elle n'est naturellement pas capable d'utiliser les pentoses produits lors de l'hydrolyse enzymatique. Afin d'améliorer les rendements, des souches modifiées capables de fermenter des sucres tels que le xylose ou l'arabinose ont donc été développées (Dashtban et al., 2009).

4°) La distillation, la rectification et la déshydratation permettent enfin de purifier l'éthanol et de le rendre propre à une utilisation en tant que carburant, qui doit répondre à un certain nombre de spécifications.

Ces différentes étapes peuvent être réalisées séparément, comme c'est le cas dans le procédé conventionnel noté SHF (*Separate Hydrolysis and Fermentation*, Figure 9A) où les enzymes et les levures peuvent être utilisées dans leurs conditions optimales respectives. Il est aussi possible de réaliser simultanément les étapes d'hydrolyse et de fermentation, un procédé que l'on nomme SSCF (*Simultaneous Saccharification and Co-Fermentation*, Figure 9B), et qui nécessite de faire un compromis sur les conditions de réaction, mais permet aux sucres formés d'être immédiatement utilisés par les levures, et donc d'éviter l'accumulation d'inhibiteurs. Enfin, le concept de bioprocédé consolidé (*Consolidated Bioprocessing* ou CBP, Figure 9C) propose d'effectuer la production de cellulases pour l'hydrolyse enzymatique et la fermentation à l'aide d'un seul microorganisme, ce qui nécessite des approches d'ingénierie métabolique ou de biologie synthétique (Hasunuma et al., 2013; Lynd et al., 2002).

Afin de rendre le bioéthanol lignocellulosique rentable, il est nécessaire d'optimiser chacune des étapes de sa production. Il est notamment essentiel de bien choisir le prétraitement adapté au type de biomasse traité, afin de maximiser son effet, tout en limitant la formation d'inhibiteurs et en réduisant les coûts et la consommation d'énergie. L'hydrolyse enzymatique est également considérée comme une étape critique en termes de rendement et de coût. La production des enzymes est en effet un procédé très coûteux, et beaucoup d'efforts tendent aujourd'hui à améliorer les rendements et diminuer les charges d'enzymes utilisées (Dashtban et al., 2009; Klein-Marcuschamer et al., 2012).

La biomasse lignocellulosique apparaît comme une matière première de choix pour l'obtention de nombreux produits, dont le bioéthanol. Sa structure en fait un matériau particulièrement résistant, mais la dégradation de certains de ces composants (cellulose et hémicelluloses) en sucres est possible, grâce à des enzymes spécialisées.

2 Enzymes impliquées dans la dégradation des polysaccharides

2.1 Principes de la classification CAZy

Les polysaccharides pariétaux précédemment cités, mais aussi tous les autres polysaccharides, notamment l'amidon (un polymère de glucose lié en α -1,4 trouvé dans les organes de réserve de nombreuses plantes) et la chitine (un polymère de N-acétylglucosamine qui forme la carapace des insectes et crustacés et se trouve également dans la paroi fongique), ainsi que tous les glycanes et tous les sucres, sont regroupés sous le terme de carbohydrates. Les carbohydrates font partie des quatre principales classes de biomolécules, avec les protéines, les acides nucléiques et les lipides. Il s'agit d'une classe de substrats extrêmement divers, étant donné le nombre de monosaccharides existants, les différentes manières dont ils peuvent s'assembler, et la variété des autres molécules auxquelles ils peuvent se lier (Lombard et al., 2014). Les organismes vivants, pour s'adapter à leur environnement, ont donc développé un panel d'enzymes actives sur ces carbohydrates (*Carbohydrate-Active enZymes*, ou CAZymes), qui contrôlent leur assemblage, leur modification et leur dégradation, et possèdent une très grande diversité de structures et de spécificités.

Afin de faciliter la gestion d'une telle diversité, une classification des CAZymes, basée sur les différences structurelles plutôt que sur les seules spécificités de substrats, a été mise en place (Cantarel et al., 2009; Henrissat, 1991). Mise en ligne sous la forme d'une base de donnée depuis 1999 (www.cazy.org), et basée sur les similarités de séquences d'acides aminés et de structures secondaires, elle permet d'affecter une séquence de protéine de fonction inconnue à une famille comportant au moins une protéine caractérisée expérimentalement, permettant de faire dans certains cas une prédiction du type de substrats sur lesquels elle pourrait être active, ou sur son mécanisme. Elle prend également en compte le caractère modulaire de beaucoup de CAZymes, qui peuvent regrouper plusieurs modules catalytiques et non-catalytiques issus de familles différentes (Lombard et al., 2014). Cette base de donnée est complétée par une encyclopédie en ligne créée en 2007 appelée *CAZypedia* (www.cazypedia.org), qui rassemble des informations détaillées sur 170 familles de CAZymes et de modules associés, chaque famille faisant l'objet d'une page comportant un résumé des résultats les plus importants la concernant (Brumer et al., 2018).

Sans cesse alimentée par de nouvelles séquences, la base de donnée CAZy comporte 435 familles de modules (au 1^{er} janvier 2020), appartenant aux classes suivantes (Cantarel et al., 2009; Levasseur et al., 2013) :

- les glycoside hydrolases (GH), qui sont les plus représentées avec 165 familles, sont capables d'hydrolyser les liaisons glycosidiques et de réaliser des réactions de transglycosylation ;
- les glycosyl-transférases (GT) sont responsables de la formation de liaisons glycosidiques ;
- les polysaccharides lyases (PL) clivent elles aussi les liaisons glycosidiques (dans les polysaccharides contenant des acides uroniques) mais en utilisant un mécanisme de β -élimination ;

- les carbohydrate estérases (CE) permettent de cliver les O- et N-acylations des sucres estérifiés, facilitant ainsi l'accès des glycosyl-hydrolases aux polysaccharides ;
- les activités auxiliaires (AA) regroupent des enzymes oxydoréductrices qui agissent en synergie avec les autres CAZymes, avec d'une part des enzymes lignolytiques, qui améliorent l'accessibilité aux polysaccharides en dégradant la lignine environnante, et d'autre part des enzymes capables d'oxyder les sucres et les polysaccharides ;
- les modules de liaison aux carbohydrates (CBM) sont des modules non-catalytiques qui facilitent l'activité enzymatique des modules catalytiques associés en améliorant les interactions avec un substrat donné.

Parmi ces familles, certaines enzymes sont plus particulièrement liées à la dégradation de la cellulose et à l'amélioration de son accessibilité, et donc particulièrement intéressantes pour la production de bioéthanol.

2.2 Les cellulases

La dégradation complète de la cellulose en glucose nécessite le clivage de toutes les liaisons glycosidiques β -(1 \rightarrow 4), ce qui fait appel à trois types d'activités complémentaires, qui se répartissent chez les champignons dans 7 familles de glycosyl-hydrolases (Couturier and Berrin, 2013).

- Les exo- β -(1,4)-D-glucanases ou cellobiohydrolases (CBH), appartenant aux familles GH6 et GH7, attaquent les chaînes de cellulose par leur extrémité réductrice ou non-réductrice, et libèrent du cellobiose de manière processive.
- Les endo- β -(1,4)-glucanases ou endoglucanases (EG), appartenant aux familles GH5, 6, 7, 12 et 45, agissent de manière apparemment aléatoire sur la cellulose amorphe, libérant du cellobiose ou de courtes chaînes de cellulose appelées cellodextrines, créant ainsi de nouvelles extrémités que les cellobiohydrolases peuvent attaquer.
- Les β -D-glucosidases (BG), qui appartiennent aux familles GH1 et GH3, agissent sur le cellobiose et les cellodextrines produits par les autres cellulases et les dégradent en monomères de glucose ; cela permet d'éviter l'accumulation de cellobiose, qui est un inhibiteur des cellobiohydrolases et des endoglucanases.

Qu'elles agissent de manière endo ou exo, les cellulases sont fréquemment liées à des CBM qui leur permettent d'approcher le module catalytique de son substrat, grâce à des interactions avec les résidus hydrophobes qu'ils portent (Gilbert et al., 2013; Himmel et al., 2010). Grâce à leurs différences de spécificité, ces trois types d'enzymes agissent de manière synergique sur la cellulose, comme représenté sur la Figure 10.

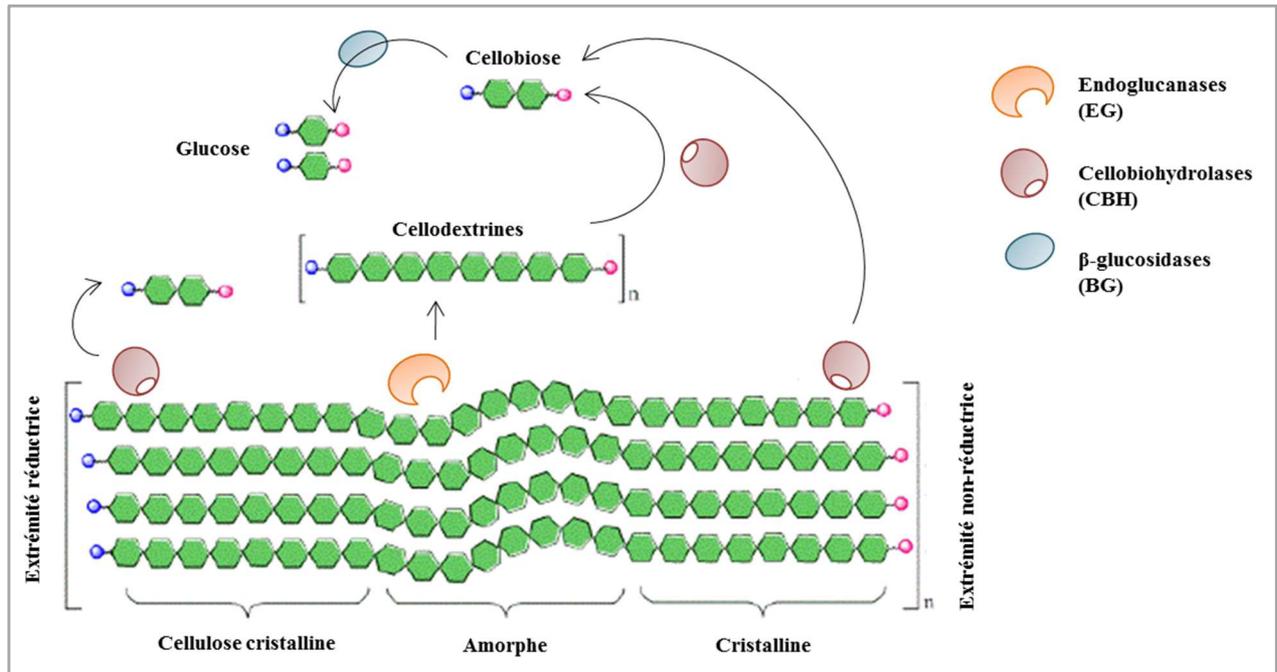


Figure 10 : Représentation schématique de l'action des cellulases sur la cellulose (Sreedharan et al., 2016).

Le mode d'action des différentes cellulases est lié à leur structure tridimensionnelle. Le site actif des CBH possède une structure en forme de tunnel (voir Figure 11) qui accueille l'extrémité d'une chaîne de cellulose d'un côté et permet à l'enzyme de glisser sur cette chaîne vers le prochain site de clivage tout en libérant le cellobiose produit par l'autre côté, lui conférant ainsi sa processivité. Les CBH de la famille GH7, qui attaquent l'extrémité réductrice des fragments de cellulose, ont un mécanisme catalytique de rétention (dans lequel le carbone anomérique du produit garde après le clivage la même configuration que celui du substrat, voir Figure 12A), tandis que celles de la famille GH6, qui attaquent l'extrémité non-réductrice, ont un mécanisme d'inversion (dans lequel le carbone anomérique change de stéréochimie au cours de la réaction, voir Figure 12B) (Couturier et al., 2016; Davies and Henrissat, 1995).

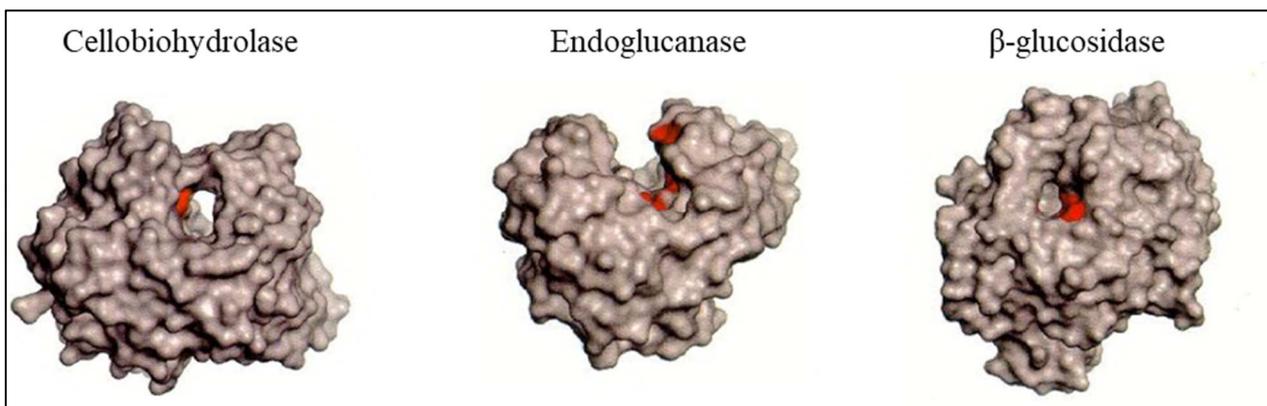


Figure 11 : Exemples de structure 3D de cellulases, montrant trois types de sites actifs (Davies and Henrissat, 1995)

Les structures représentées montrent un site en forme de tunnel pour la cellobiohydrolase, en forme de vallée pour l'endoglucanase, et en forme de poche pour la β -glucosidase. Les résidus catalytiques supposés sont colorés en rouge.

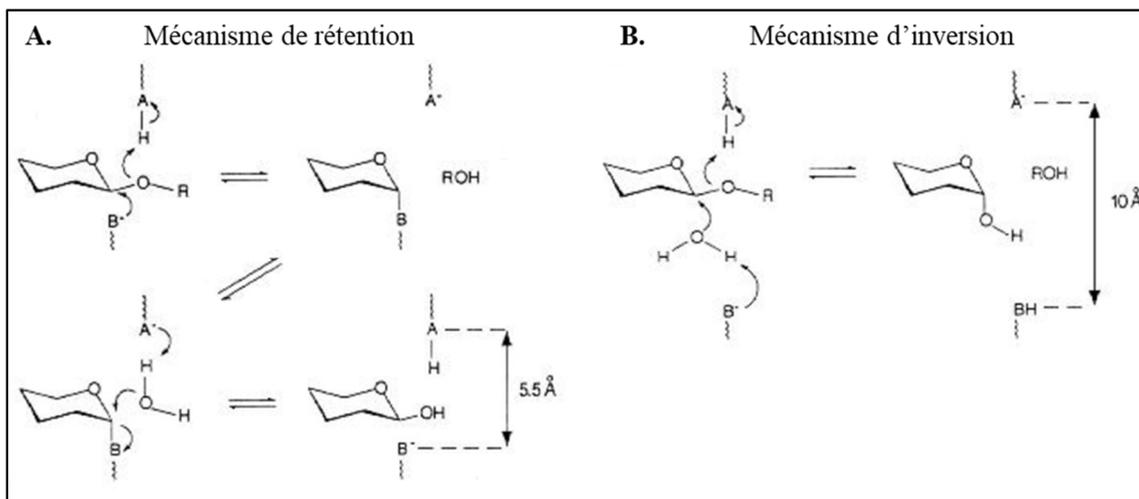


Figure 12 : Comparaison des deux principaux mécanismes d'hydrolyse des liaisons glycosidiques (Davies and Henrissat, 1995)

A. Lors du mécanisme de rétention, l'oxygène de la liaison glycosidique est protoné par un résidu catalytique acide (AH), et la liaison est rompue par l'attaque nucléophile d'un résidu jouant le rôle de base (B⁻) ; la liaison ainsi formée est ensuite hydrolysée par une molécule d'eau, et cette seconde substitution nucléophile génère un produit ayant la même stéréochimie que le substrat.

B. Dans le cas du mécanisme d'inversion : l'oxygène est également protoné par le résidu acide, mais une seule attaque nucléophile, par une molécule d'eau activée par la base, se produit, résultant en une inversion de configuration. La base est dans ce cas plus éloignée du substrat afin d'accommoder cette molécule d'eau, et les deux résidus catalytiques sont plus éloignés que dans les enzymes ayant un mécanisme de rétention.

Les endoglucanases, qui se retrouvent dans plusieurs familles CAZy, ont une grande diversité de structure, et peuvent utiliser soit un mécanisme d'inversion soit un mécanisme de rétention. Il existe néanmoins un point commun entre toutes les EG : elles possèdent un site actif en forme de vallée ouverte, qui leur permet d'accommoder la chaîne de cellulose sans forcément l'attaquer par l'une de ses extrémités. La différence entre cette topologie et celle du tunnel, notamment dans les familles GH6 et GH7 qui contiennent à la fois des CBH et des EG, est due à l'existence de longues boucles qui viennent couvrir une partie du site actif chez les CBH, et qui sont absentes chez les EG (Davies and Henrissat, 1995). Cependant, ces boucles peuvent être plus ou moins flexibles, et la différence entre les activités exo- et endo-lytiques n'est pas toujours bien marquée ; certaines CBH possèdent en effet une faible activité endo (Badino et al., 2017; Ståhlberg et al., 1993). Beaucoup d'organismes sécrètent plusieurs endoglucanases, ce qui pourrait s'expliquer par des différences entre les familles, notamment en termes de taille et de spécificité de substrat. En effet, les enzymes des familles GH12 et GH45 ont un poids moléculaire beaucoup plus faible que celles des autres familles, ce qui leur permettrait d'accéder aux plus petits pores du substrat (Payne et al., 2015), et la plupart des EG ont des activités complémentaires sur les hémicelluloses, dépendant de leur famille : il existe par exemple des GH5 actives sur les mannanes et galactomannanes, des GH6 sur l'arabinoxylane, des GH12 sur le xylane et le xyloglucane, et des GH7 actives sur ces trois derniers substrats (Vlasenko et al., 2010).

Les β -glucosidases possèdent un site actif en forme de poche qui leur permet d'accommoder l'extrémité non-réductrice des cellodextrines, et de cliver un seul monomère de sucre. Avant le clivage suivant, le substrat doit être relâché pour laisser le monomère produit sortir du site actif, ce qui explique le caractère non-processif de ces enzymes (Couturier et al., 2016).

2.3 Les hémicelluloses

Les hémicelluloses ont une structure complexe impliquant des liaisons osidiques très variées ; leur dépolymérisation nécessite donc l'action d'un grand nombre d'enzymes spécialisées, appartenant chez les champignons à une vingtaine de familles de GH et plusieurs familles de CE, qui agissent en synergie (Himmel et al., 2010; van den Brink and Vries, 2011).

Pour chacun des trois principaux types d'hémicelluloses, il existe des enzymes qui attaquent la chaîne principale, comme illustré sur la Figure 13 :

- les xyloglucane β -1,4-endoglucanases (XEG), appartenant aux familles GH12 et GH74 ;
- les β -1,4-endoxy lanases (XLN) des familles GH10, GH11 et GH30, certaines GH30 ayant également montré une activité exo-xylobiohydrolase (XBH) (Katsimpouras et al., 2019; Nakamichi et al., 2019) ;
- les β -1,4-endomannanases (MAN) des familles GH5 et GH26 et les β -1,4-mannosidases (MND) de la famille GH2.

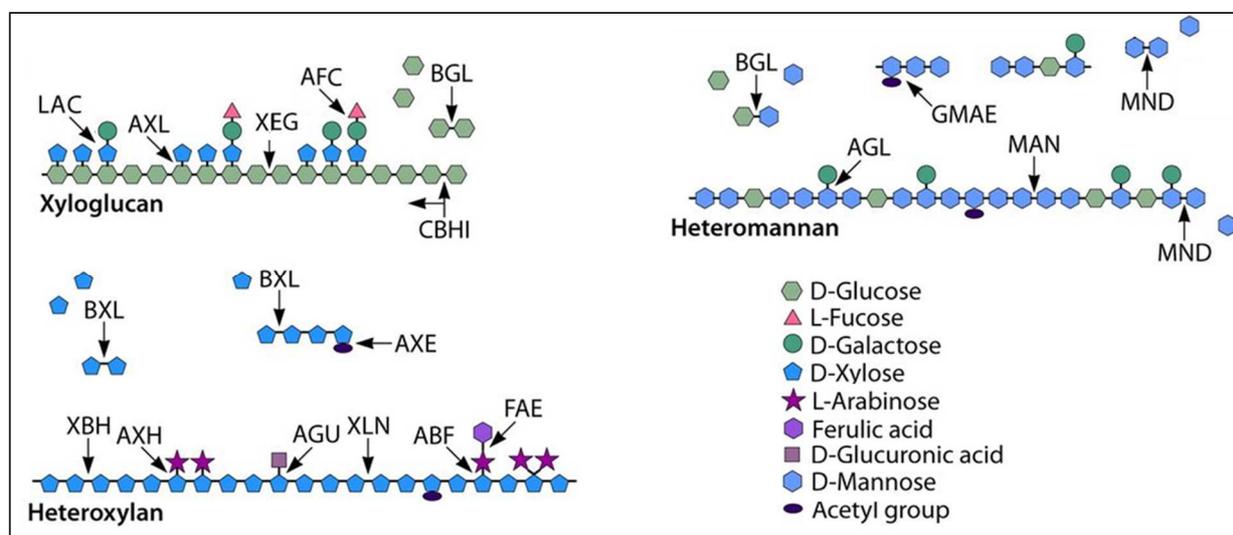


Figure 13 : Représentation schématique de différents types d'hémicelluloses (Rytioja et al., 2014; van den Brink and Vries, 2011).

AFC : α -fucosidase ; ABF : arabinofuranosidase ; AGU : α -glucuronidase ; AGL : α -1,4-galactosidase ; AXH : arabinoxylyane arabinofuranosidase ; AXE : acétyle xylane estérase ; AXL : α -xylosidase ; BGL : β -1,4-glucosidase ; BXL : β -1,4-xylosidase ; CBH : cellobiohydrolase ; FAE : féruloyl estérase ; GMAE : galactomannane acétyle estérase ; LAC : β -1,4-galactosidase ; MAN : β -1,4-endomannanase ; MND : β -1,4-mannosidase ; XBH : exo-xylobiohydrolase ; XEG : xyloglucane β -1,4-endoglucanases ; XLN : β -1,4-endoxy lanases.

Afin de compléter la dépolymérisation, il est nécessaire de débrancher les différentes substitutions, ce qui fait appel à au moins 10 activités différentes, elles aussi résumées sur la Figure 13 : β -1,4-galactosidase (LAC), α -xylosidase (AXL), α -fucosidase (AFC), arabinoxylane α -arabinofuranosidase (AXH), α -glucuronidase (AGU), α -arabinofuranosidase (ABF), féruloyle estérase (FAE), acétyl xylane estérase (AXE), α -1,4-galactosidase (AGL), galactomannane acétyl estérase (GMAE).

Les oligosaccharides issus de cette dégradation sont ensuite décomposés en monomères par les β -1,4-glucosidases (BGL) des familles GH1 et 3, les β -1,4-xylosidases (BXL) des familles GH3 et 43, et les β -1,4-mannosidases (MND) de la famille GH2.

2.4 Les LPMO

2.4.1 Découverte de leur activité

Dès 1974, les travaux de Eriksson et al. mentionnent l'existence d'enzymes oxydatives, sécrétées par des champignons cellulolytiques, qui interviennent dans la dégradation de la cellulose aux côtés des hydrolases. Dans les années 90 et 2000, il a été démontré que certains champignons dits de « pourriture brune » (voir paragraphe 3.1 de ce chapitre) oxydent la cellulose de manière non-enzymatique, à l'aide de radicaux libres produits de manière extracellulaire par des réactions de Fenton impliquant des peroxydes et des ions fer (par exemple : $\text{Fe}^{2+} + \text{H}_2\text{O}_2 \rightarrow \text{Fe}^{3+} + \text{OH}\cdot + \text{OH}^-$) (Arantes et al., 2012). Cependant, aucune oxydation enzymatique n'a été reportée pendant près de 40 ans, jusqu'à la fin des années 2000, lorsque deux protéines précédemment assignées aux familles GH61 et CBM33 ont révélé des activités jusqu'à présent ignorées.

La protéine CBP21 de la bactérie *Serratia marcescens* était classée comme un module de liaison à la chitine de la famille des CBM33, sans activité catalytique, mais il a ensuite été montré que sa présence permettait d'améliorer l'activité des chitinases (Vaaje-Kolstad et al., 2005), et ce grâce à un mécanisme de clivage oxydatif de la chitine (Vaaje-Kolstad et al., 2010). En parallèle, lors de la résolution de la structure de la protéine Cel61B de *Trichoderma reesei* appartenant à la famille des GH61, d'importantes similarités structurelles avec CBP21 sont apparues (Karkehabadi et al., 2008). La caractérisation de plusieurs GH61, dont certaines avaient montré une faible activité endoglucanase, a prouvé leur capacité à améliorer l'activité des cellulases (Harris et al., 2010). Il a enfin été montré que les GH61 et CBM33 étaient bien des métallo-enzymes contenant du cuivre, capables en présence d'oxygène et d'un donneur d'électrons d'oxyder respectivement la cellulose (Forsberg et al., 2011; Quinlan et al., 2011) et la chitine (Vaaje-Kolstad et al., 2012).

Les familles GH61 et CBM33 ont par la suite été regroupées sous le terme de "*Lytic Polysaccharide Monoxygenases*" (Horn et al., 2012), et respectivement renommées AA9 et AA10 dans la nouvelle classe des activités auxiliaires (AA) de la base de données CAZy (Levasseur et al., 2013). La présence de ces protéines chez un nombre très important de micro-organismes impliqués dans la dégradation de la matière organique mène à penser que les LPMO pourraient être un élément fondamental de leur mode de vie, au même titre que les hydrolases ; cette observation, ainsi que le potentiel de ces enzymes pour la complémentation des cocktails cellulolytiques classiques, ont logiquement mené à un très fort intérêt de la part de la communauté scientifique (Agostoni et al., 2017; Horn et al., 2012; Johansen, 2016; Morgenstern et al., 2014). Ces recherches ont mené à la découverte de plusieurs nouvelles familles de LPMO, et les familles CAZy AA11, AA13, AA14 et AA15 ont été créées.

2.4.2 Classification actuelle

Les LPMO sont donc regroupées dans six familles CAZy, qui se différencient par leurs séquences d'acides aminés ; ces familles sont présentes dans différents types d'organismes et présentent des spécificités de substrats variées (voir résumé dans le Tableau 3).

- La famille fongique des AA9 regroupe des enzymes actives sur la cellulose, mais aussi sur les cello-oligosaccharides et les hémicelluloses contenant des unités glucose liées en β -1,4 : xyloglucanes, glucomannanes, β -glucanes mixtes (Bennati-Granier et al., 2015; Fanuel et al., 2017; Frommhagen et al., 2016; Isaksen et al., 2014) ; une protéine active sur le xylane, mais seulement en présence de cellulose, a également été caractérisée chez *Myceliophthora thermophila* (Frommhagen et al., 2015), et l'une des AA9 de *Lentinus similis* est capable de cliver le xylane et les xylo-oligosaccharides (Simmons et al., 2017).
- Les LPMO de la famille AA10 se trouvent principalement chez les bactéries et les virus, et sont actives sur la chitine, la cellulose, et parfois les deux (Agostoni et al., 2017; Forsberg et al., 2014b) ; récemment une AA10 de l'actinobactérie *Kitasatospora papulosa* a montré une activité à la fois sur la cellulose, la chitine et le xylane (Corrêa et al., 2019). Quelques AA10 sont trouvées dans les génomes de champignons de la classe des Ustilaginomycètes, tels que le pathogène du maïs *Ustilago maydis*, mais aucune de ces enzymes n'a encore été caractérisée.
- Les familles AA11 et AA13 ont été découvertes chez les champignons grâce à des homologies de séquence avec des domaines portés par les familles de LPMO déjà connues ; elles ne contiennent pour l'instant que quelques protéines caractérisées, et sont actives respectivement sur la chitine et les composants de l'amidon (amylose et amylopectine) (Hemsworth et al., 2014; Lo Leggio et al., 2015; Vu et al., 2014b; Vu et al., 2019).

- Les AA14 ont été découvertes grâce au développement d'approches sécrétomiques chez les champignons (Berrin et al., 2017) ; la caractérisation de deux d'entre elles, issues de *Pycnoporus coccineus*, a révélé qu'elles sont actives sur le xylane recouvrant les fibres de cellulose dans le bois (Couturier et al., 2018).
- La famille AA15 récemment découverte regroupe des enzymes, actives sur la cellulose et la chitine, issues d'organismes très variés : oomycètes, algues unicellulaires et multicellulaires, virus, invertébrés terrestres et marins (crustacés, mollusques, insectes, millipèdes, araignées) (Sabbadin et al., 2018).

Les modules catalytiques de ces différentes familles sont fréquemment associés à d'autres modules, notamment des modules de liaison aux carbohydrates (CBM), dont les spécificités correspondent aux substrats identifiés pour chaque famille (voir Tableau 3) (Berrin et al., 2017; Couturier et al., 2018; Horn et al., 2012; Lenfant et al., 2017; Sabbadin et al., 2018; Vu et al., 2019).

Tableau 3 : Récapitulatif des principales caractéristiques des familles de LPMO existantes.

Les données sont issues de CAZy ainsi que des différentes références citées dans le corps du texte.

Famille CAZy	Type d'organismes	CBM associés	Nombre d'enzymes caractérisées	Spécificité de substrat	Régiosélectivités
AA9	Champignons	CBM1 (cellulose), CBM18 (chitine)	37	Cellulose, cello-oligosaccharides, glucanes mixtes, xyloglucane, glucomananne, xylane	C1, C4, et C1/C4
AA10	Bactéries, virus, archées, plantes, champignons (<i>Ustilaginaceae</i>)	CBM1 (cellulose), CBM2 (cellulose/chitine/xylane), CBM12, CBM5 (cellulose/chitine), CBM14 (chitine), CBM20 (amidon)	23	Cellulose, chitine, xylane	C1 et C1/C4 (cellulose) C1 (chitine)
AA11	Champignons	Aucun	2	Chitine	C1
AA13	Champignons	CBM20 (amidon)	4	Amidon	C1
AA14	Champignons	Aucun	2	Xylane associé à la cellulose	C1
AA15	Invertébrés, oomycètes, algues, virus	CBM1 (cellulose), CBM14 (chitine)	2	Cellulose, chitine	C1

Les champignons sont donc les organismes chez lesquels on trouve le plus grand nombre de familles différentes de LPMO. Ils possèdent généralement plusieurs gènes de chaque famille, et notamment de la famille AA9, avec souvent plus de 10 gènes pour une seule espèce, et parfois jusqu'à plus de 30. Les raisons d'une telle expansion sont encore mal connues, mais elle pourrait être liée à la diversification de la famille en termes de spécificités de substrat et d'autres propriétés biochimiques, et refléter des différences de régulation permettant de s'adapter à des substrats variés (Berrin et al., 2017). Les études sécrétomiques semblent aller dans ce sens, puisqu'elles montrent que seule une partie des LPMO encodées dans les génomes fongiques sont effectivement sécrétées, et qu'il ne s'agit pas des mêmes selon les substrats utilisés (Berrin et al., 2017). Parmi ces protéines sécrétées simultanément, on retrouve souvent des représentants de plusieurs types de régiosélectivités, qui pourraient avoir des activités complémentaires pour la dégradation efficace des polysaccharides.

Les différentes LPMO caractérisées varient en effet par leur régiosélectivité. L'oxydation des liaisons β -1,4 peut avoir lieu sur le carbone C1, générant une lactone qui est spontanément hydratée en un acide aldonique plus stable, ou sur le carbone C4, générant un kétoaldose en équilibre avec le gemdiol correspondant (voir Figure 14). Parmi les AA9, certaines oxydent sélectivement en C1, d'autres en C4, et d'autres encore n'ont pas de préférence et sont capables d'oxyder les deux carbones (Frommhagen et al., 2018). Les AA10 actives sur la cellulose sont capables d'oxyder en C4, mais pas celles actives sur la chitine, et jusqu'à présent les LPMO des autres familles n'ont montré qu'une régiosélectivité pour le carbone C1 (voir Tableau 3) (Couturier et al., 2018; Forsberg et al., 2018; Hemsworth et al., 2014; Sabbadin et al., 2018; Vu et al., 2019; Wang et al., 2018a).

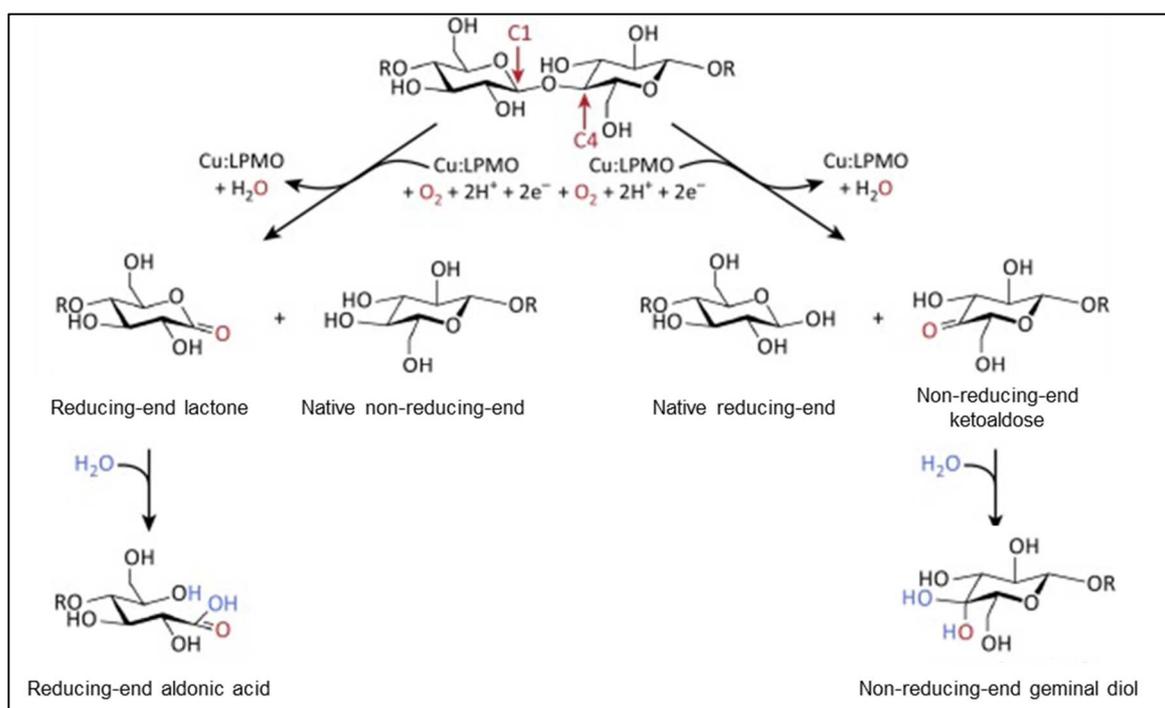


Figure 14 : Régiosélectivité de l'oxydation des liaisons glycosidiques par les LPMO (d'après Hemsworth et al., 2015)

2.4.3 Structure tridimensionnelle des LPMO

De manière générale, les LPMO se caractérisent par une structure de type immunoglobuline, constituée de deux feuillets β antiparallèles, de 4 à 5 brins chacun, qui forment ce qu'on appelle un β -sandwich. Elles ne possèdent pas de site actif en forme de tunnel ou de vallée, mais plusieurs boucles formant une surface plus ou moins plane, comportant des résidus aromatiques et hydrophobes qui servent aux interactions avec le substrat ; cette surface plane leur permet de se lier à la surface de substrats cristallins comme la cellulose ou la chitine, au lieu de reconnaître une chaîne isolée comme le font les cellulases (voir Figure 15).

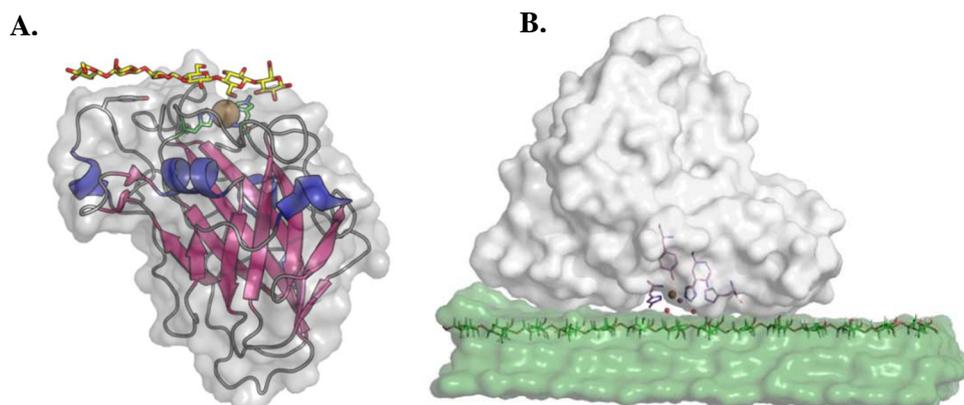


Figure 15 : Structure générale des LPMO et interactions avec leurs substrats

A. Exemple de l'AA9 de *Lentinus similis* en interaction avec du cellohexaose (en bâtonnets jaunes), montrant la structure centrale de type immunoglobuline (en rose), le centre actif formé par les deux histidines (batônnetts verts) et le cuivre (sphère orange).

Cette structure est déposée sur PDB sous le code d'accès 5ACI (Frandsen et al., 2016).

B. Exemple de l'AA9 de *Thermoascus aurantiacus* en interaction avec une surface de cellulose cristalline (en vert). Cette structure est déposée sur PDB sous le code d'accès 2YET (Kim et al., 2014).

Exposé sur cette surface se trouve le site actif, constitué d'un ion cuivre coordonné sur le plan équatorial par 3 atomes d'azote portés par deux résidus histidines, dont l'un est situé à l'extrémité N-terminale, formant ce que l'on appelle l'"*histidine brace*" (voir Figure 16). Egalement dans la sphère de coordination du cuivre, en position axiale, on trouve généralement une tyrosine, qui est remplacée dans les LPMO des familles AA10 et AA15 actives sur la chitine par un résidu phénylalanine (Frandsen and Lo Leggio, 2016; Tandrup et al., 2018). L'histidine N-terminale est systématiquement méthylée dans les LPMO fongiques mais pas dans les LPMO bactériennes, les champignons filamenteux étant les seuls à posséder la méthyltransférase nécessaire à cette modification post-traductionnelle. Le rôle fonctionnel de cette modification n'est pas encore totalement élucidé, mais il est clair que les LPMO restent actives en son absence, comme c'est le cas lorsqu'elles sont produites dans la levure *Pichia pastoris*. Une étude récente de Petrovic et al. (2018) impliquant deux variantes de la même protéine, avec et sans méthylation, a montré que bon nombre de leurs caractéristiques étaient identiques (spécificité de substrat, régiosélectivité, affinité pour le cuivre, potentiel redox...) mais que la version méthylée semblait plus résistante à la présence de peroxyde d'hydrogène, qui peut entraîner des

dommages oxydatifs (voir aussi paragraphe suivant). La méthylation de l’histidine N-terminale serait donc impliquée dans la protection des LPMO contre l’inactivation par l’auto-oxydation.

La présence d’un ion cuivre (II) dans l’*histidine brace* permet, outre les études structurales, d’en étudier la géométrie par résonance paramagnétique électronique (RPE), une technique qui a été très utilisée pour étudier l’environnement immédiat du cuivre dans les LPMO, seules ou en interaction avec leurs substrats (Ciano et al., 2018). Dans les protéines à cuivre, il existe plusieurs types de centres, classés en fonction de leurs caractéristiques spectrales UV, visible et RPE. Toutes les LPMO caractérisées jusqu’à présent ont un spectre RPE typique des centres de type 2, qui sont mononucléaires et coordonnés dans une géométrie plane ou tétragonale par plusieurs histidines (souvent 3 ou 4 ; le centre des LPMO est donc assez inhabituel puisqu’il n’en contient que deux, dont l’une est doublement impliquée puisqu’elle interagit avec le cuivre via ses deux atomes d’azote) (Frandsen and Lo Leggio, 2016; Hemsworth et al., 2018).

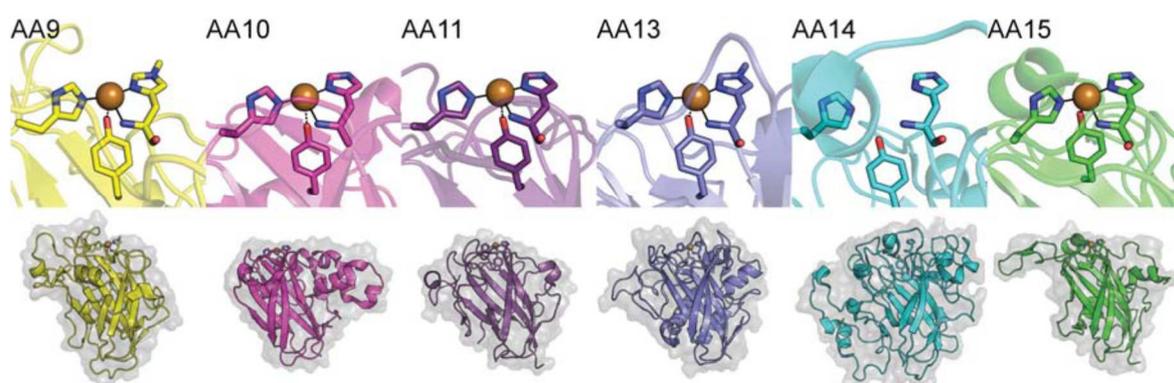


Figure 16 : Structure tridimensionnelle de LPMO de différentes familles

Structures représentatives des différentes familles de LPMO, avec en haut un agrandissement des sites actifs montrant les résidus de l’*histidine brace* (histidines, méthylées ou non, et tyrosines, représentés en bâtonnets) coordonnant l’ion cuivre (sphère orange).

Les codes PDB des protéines sont les suivants : 5ACH (AA9, en jaune), 5OPF (AA10, en rose), 4MAI (AA11, en violet), 4OPB (AA13, en bleu foncé), 5NO7 (AA14, en bleu clair) et 5MSZ (AA15, en vert) (Tandrup et al., 2018).

2.4.4 Mécanismes proposés

Concernant le mécanisme exact de l’oxydation, plusieurs possibilités existent, et ce sujet fait débat dans la communauté. L’appellation de “monooxygénase” qui a été donnée aux LPMO repose sur le fait que, comme d’autres enzymes, elles peuvent former un groupe hydroxyle à partir d’un atome d’oxygène issu d’une molécule d’O₂, ce qui a été prouvé par des expériences utilisant l’isotope radioactif ¹⁸O₂ (Vaaje-Kolstad et al., 2010). Plusieurs mécanismes basés sur cette observation ont été proposés, tous commençant par la réduction, grâce à un donneur d’électrons, du complexe LPMO-Cu(II) en LPMO-Cu(I), qui active ensuite l’O₂ en superoxyde pour donner un intermédiaire LPMO-Cu(II)-O₂^{•-}. Celui-ci réagit avec le substrat selon un

mécanisme radicalaire faisant intervenir un électron et deux ions H^+ , pour finalement aboutir au produit oxydé et au complexe LPMO-Cu(II) initial (Beeson et al., 2015); un exemple est donné sur la Figure 17A.

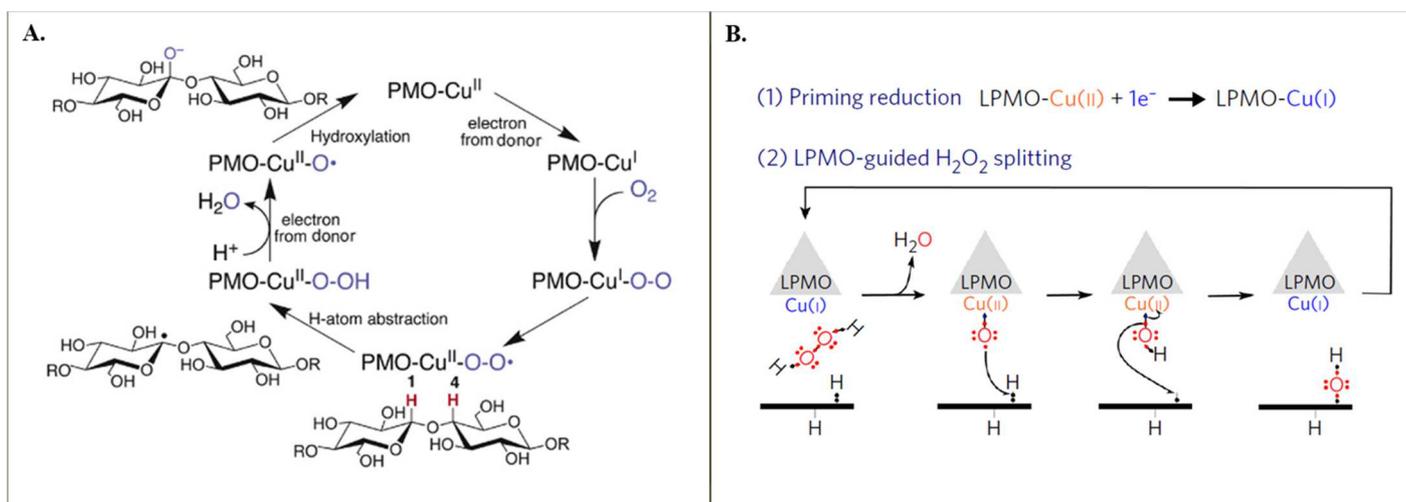


Figure 17 : Mécanismes proposés pour l'oxydation de la liaison glycosidique par les LPMO

A. Exemple de mécanisme initialement proposé faisant intervenir le dioxygène (Phillips et al., 2011)

B. Mécanisme proposé plus récemment faisant intervenir le peroxyde d'hydrogène (Bissaro et al., 2017)

En l'absence de substrat, les LPMO sont également capables de produire du peroxyde d'hydrogène (H_2O_2), propriété qui a été utilisée pour mesurer leur activité de réduction de l'oxygène (Kittl et al., 2012). Dans un article paru en 2017, Bissaro et al. suggèrent que le peroxyde d'hydrogène serait en fait le véritable cofacteur des LPMO, et que le mécanisme faisant intervenir le dioxygène n'aurait lieu qu'en son absence. Ils montrent l'incorporation dans la chaîne de cellulose d'un atome d'oxygène marqué provenant du substrat $H_2^{18}O_2$, et prouvent la présence d'une activité en l'absence d'oxygène lorsque du peroxyde d'hydrogène est ajouté. Ils proposent un mécanisme où le donneur d'électrons, présent en quantités catalytiques, n'interviendrait que pour la réduction initiale du complexe LPMO-Cu(II) en LPMO-Cu(I), qui réagirait ensuite avec H_2O_2 et le substrat selon un mécanisme radicalaire pour donner le produit oxydé et régénérer le centre Cu(I), qui peut commencer un nouveau cycle catalytique (voir Figure 17B).

Cependant, le peroxyde d'hydrogène peut aussi inactiver les LPMO lorsqu'il est présent en trop grande quantité, en générant des espèces réactives de l'oxygène spécifiquement dans le centre catalytique, qui mènent à des dommages oxydatifs sur les résidus les plus proches du cuivre, et notamment les deux histidines qui voient leur cycle imidazole modifié. La liaison des LPMO à leur substrat leur permet de se protéger de ces dommages oxydatifs menant à l'inactivation (Bissaro et al., 2017; Loose et al., 2018).

2.4.5 Partenaires redox

Quel que soit le mécanisme, l'action des LPMO requiert l'apport d'électrons, et donc la présence d'un réducteur. Dans les tests d'activité *in vitro*, les réducteurs utilisés sont souvent de petites molécules (acide ascorbique, L-cystéine, glutathion, acide gallique et autres phénols) ou divers composants de la lignine, qui ont des efficacités variables selon les protéines (Frommhagen et al., 2016; Westereng et al., 2015). L'identité des donneurs d'électrons *in vivo* est encore incertaine, mais les candidats sont nombreux : en plus des composés produits par les champignons ou les plantes (c'est le cas de l'acide ascorbique, du glutathion et de nombreux phénols) ou présents dans le bois (tels que l'acide gallique ou la lignine), plusieurs enzymes peuvent également jouer ce rôle.

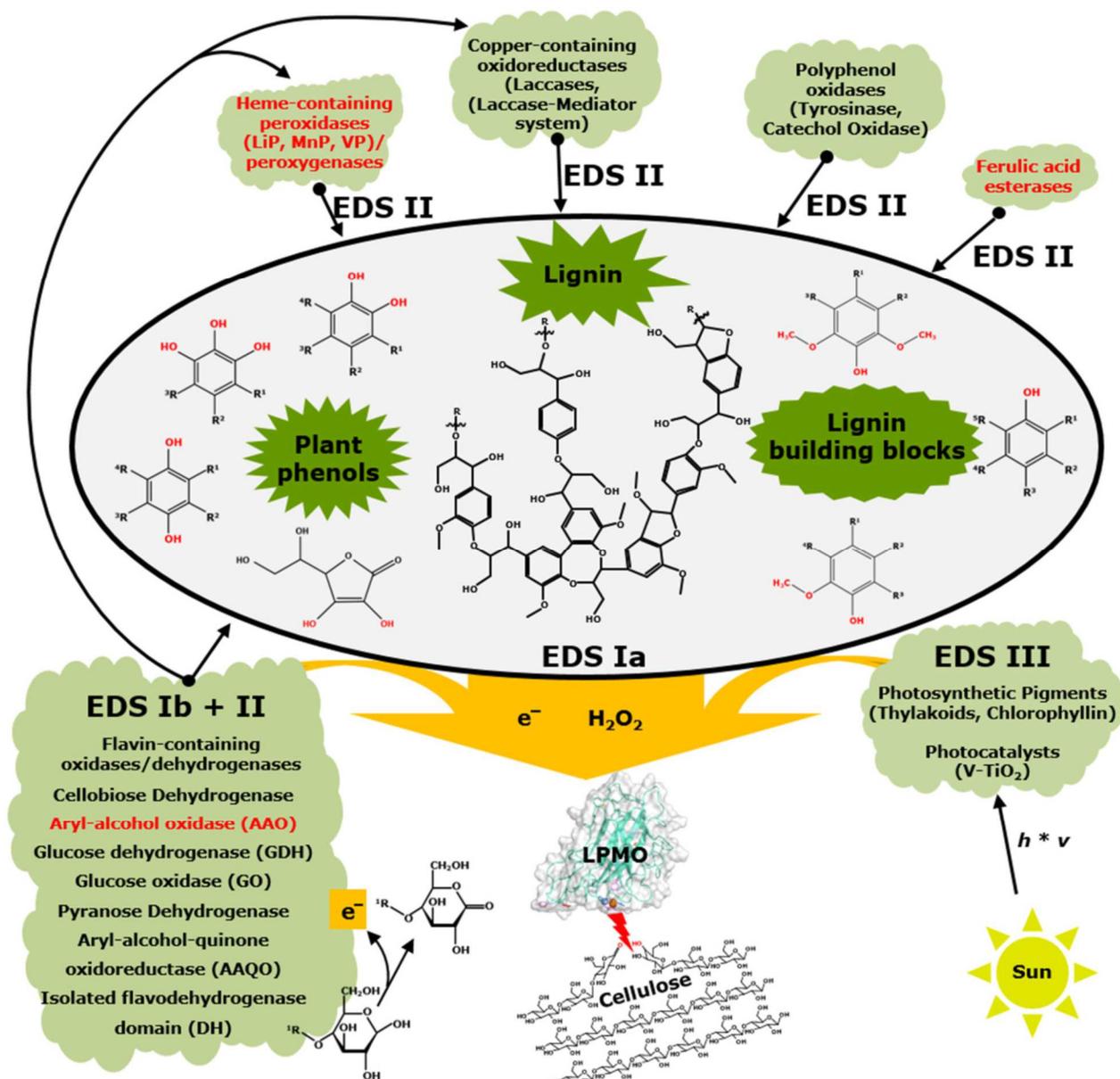


Figure 18 : Interactions des LPMO avec différents systèmes de transfert d'électrons (EDS) (Frommhagen et al., 2018)

Système I : donneurs directs, non-enzymatiques (Ia) et enzymatiques (Ib). Système II : donneurs indirects. Système III : systèmes photo-activés.

La cellobiose déshydrogénase (CDH, de la famille AA_1), une enzyme fongique responsable de l'oxydation du cellobiose et d'autres oligosaccharides, a été identifiée très tôt comme un partenaire redox potentiel pour les LPMO ; ce rôle est aujourd'hui bien établi, et supporté par le fait que les CDH sont souvent co-exprimées avec les AA9, ou encore que la délétion du gène *cdh* entraîne une baisse d'efficacité lignocellulolytique (Kracher et al., 2016; Phillips et al., 2011; Tangthirasunun et al., 2017). Cependant cette enzyme est absente chez les bactéries, et chez certains champignons. D'autres enzymes peuvent alors jouer le rôle de réducteur, comme les glucose déshydrogénases (GDH) et les aryl-alcool quinone oxidoréductases (AAQO) de la famille AA3_2 (Garajova et al., 2016) ; les autres enzymes de la famille AA3_2 (voir Figure 18) peuvent quant à elles intervenir dans le transfert d'électrons de manière indirecte, en régénérant les phénols capables de réduire le cuivre du site actif (Kracher et al., 2016). Le rôle biologique de ces différentes enzymes est confirmé par le fait qu'elles soient également co-exprimées avec les LPMO de la famille AA9 (Berrin et al., 2017).

Enfin, plusieurs systèmes activés par la lumière peuvent servir à transférer des électrons aux LPMO. Cannella et al. (2016) ont utilisé des pigments photosynthétiques (thylakoïdes et chlorophylline), combinés avec de l'ascorbate, qui ont multiplié par 100 l'activité catalytique des LPMO lorsque le système était exposé à la lumière. La même année, Bissaro et al. ont prouvé que l'oxydation de l'eau par un photocatalyseur (TiO₂ dopé au vanadium) peut fournir les électrons nécessaires à l'activité des LPMO.

2.4.6 Utilisation des LPMO dans les cocktails cellulolytiques

Dès leur découverte, les LPMO ont éveillé l'attention par leurs possibles applications industrielles dans le domaine des biotechnologies ; en effet, les premières découvertes concernant ces enzymes sont apparues dans des brevets industriels avant d'être publiées dans des journaux scientifiques, et elles ont rapidement été intégrées aux cocktails cellulolytiques commerciaux (Brown et al., 2008; Dotson et al., 2005; Harris et al., 2010; Johansen, 2016). Contrairement aux EG et aux CBH qui agissent sur des chaînes de cellulose isolées, les LPMO peuvent se fixer à la surface de la cellulose cristalline (Eibinger et al., 2014). Elles sont ainsi capables d'introduire des ruptures de chaînes dans des régions récalcitrantes, ce qui fragilise mécaniquement la structure des fibres de cellulose (Villares et al., 2017), et crée de nouveaux points d'attaques pour les hydrolases, ce qui améliore leur efficacité (voir Figure 19). Cette synergie se traduit par une importante amélioration de la saccharification de substrats lignocellulosiques industriels variés, tels que la paille de blé, la rafle de maïs, la bagasse de canne à sucre, le pin, le peuplier, le bouleau ou l'épicéa (Hemsworth et al., 2015; Karnaouri et al., 2017; Müller et al., 2018). Cependant, l'utilisation de LPMO dans des procédés industriels nécessite de repenser ces derniers et de les adapter pour tenir compte des conditions bénéfiques à l'activité des LPMO, notamment la présence de partenaires redox adaptés, ainsi que l'apport suffisant et contrôlé d'oxygène ou de peroxyde d'hydrogène (Müller et al., 2015; Müller et al., 2018).

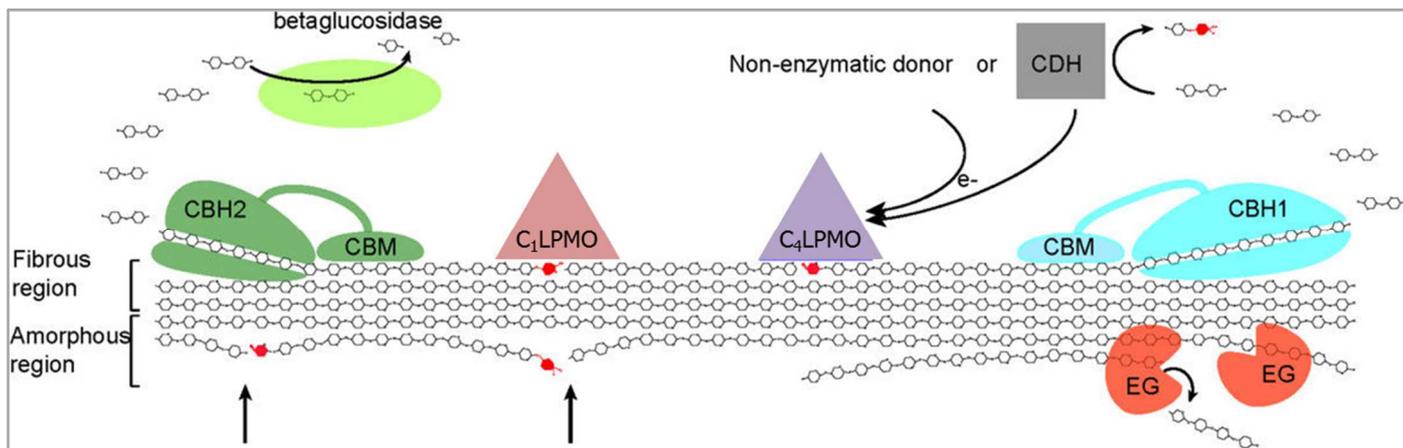


Figure 19 : Représentation de la synergie entre LPMO, hydrolases et cellobiose déshydrogénase (Horn et al., 2012)

Les sucres oxydés, introduits en C1 ou en C4, sont représentés en rouge ; les nouveaux points d'attaque pour les CBH sont montrés par les flèches noires. L'accumulation de clivages oxydatifs peut libérer des oligosaccharides natifs (comme représenté sur la partie supérieure de la cellulose, entre les deux LPMO) ou oxydés (comme en bas de l'image).

Les cellulases, les hémicellulases et les LPMO ont des activités complémentaires, qui leur permettent de dépolymériser efficacement les polysaccharides pariétaux, avec l'aide d'autres enzymes. Parmi les organismes qui les produisent naturellement, les champignons filamenteux font partie des plus efficaces, et sont donc une source très utilisée d'enzymes lignocellulolytiques pour la dégradation de la biomasse végétale.

3 Utilisation de secrétomes fongiques pour l'hydrolyse de biomasse lignocellulosique

3.1 Classifications et modes de vie des champignons

Le règne fongique est constitué d'organismes eucaryotes, hétérotrophes vis-à-vis du carbone, qui ne peuvent pas réaliser de photosynthèse et se nourrissent donc par absorption au travers de leur paroi cellulaire de matière organique digérée de manière extracellulaire. Parmi ceux-ci on distingue souvent d'une part les levures unicellulaires, et d'autre part les champignons filamenteux, dont l'organe végétatif est composé d'hyphes qui s'associent en filaments, formant ce qu'on appelle le mycélium. Ce règne regroupe des organismes très différents les uns des autres, et sa phylogénie est complexe (Figure 20). On distingue communément les « champignons inférieurs » et les « champignons supérieurs » (Silar, 2016) :

- Les champignons dits inférieurs, anciennement regroupés sous les noms de « chytridiomycètes » et « zygomycètes » et aujourd'hui classés dans plusieurs embranchements (Hibbett et al., 2007), ont gardé des caractéristiques ancestrales comme la dispersion à l'aide de zoospores flagellées ou un mycélium simple, rarement cloisonné.
- La lignée des *Dikarya* regroupe les champignons supérieurs, s'organisant en mycélium perfectionné (dont les hyphes dits « septés » ont des cloisons percées d'un pore central associé à des structures permettant le passage sélectif de nutriments et d'organites), et capables d'effectuer des fusions entre deux cellules (anastomose). Parmi eux, on distingue les *Ascomycota* des *Basidiomycota* par leurs structures formant des spores : il peut s'agir soit de sacs (ou asques) contenant les ascospores, soit de cellules spécialisées appelées basides qui bourgeonnent pour donner les basidiospores.

Les champignons, présents dans tous les écosystèmes, s'adaptent parfois à des conditions extrêmes, et jouent des rôles écologiques essentiels, notamment par leur implication dans la dégradation et le recyclage de la matière organique. Les champignons ont acquis au cours de leur évolution différents styles de vie (Rytioja et al., 2014) :

- la symbiose, qui leur permet d'avoir une relation mutualiste avec des algues (pour former des lichens), des animaux (champignons du système digestif) ou encore des plantes (champignons mycorhiziens par exemple) ;
- le parasitisme, lorsqu'ils se développent au détriment d'une autre espèce, de plantes ou d'animaux par exemple ;
- le saprophytisme, lorsqu'ils dégradent le bois mort ou les déchets organiques.

Cependant, ces modes de vie ne sont pas cloisonnés, et beaucoup de champignons peuvent cumuler plusieurs de ces stratégies nutritionnelles ou passer de l'une à l'autre.

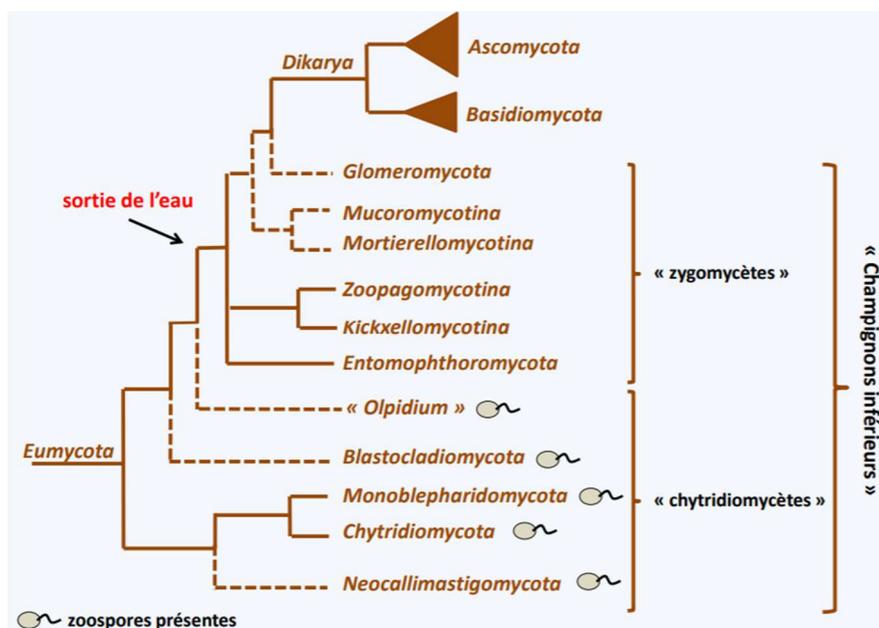


Figure 20 : Phylogénie du règne *Fungi* (ou *Eumycota*) (Silar, 2016)

Ces stratégies nutritionnelles sont souvent associées à des différences dans l'éventail d'activités enzymatiques des différentes espèces fongiques. Par exemple, les champignons ectomycorhiziens tels que *Laccaria bicolor* et *Piriformospora indica*, qui se développent en symbiose avec des plantes grâce à un réseau mycélien entourant les racines, ont un arsenal de CAZymes réduit, qui leur permet de préserver l'intégrité des cellules de leur hôte. Seules quelques GH, ainsi que quelques LPMO, sont exprimées avant ou pendant la colonisation des racines, et potentiellement impliquées dans la modification des parois végétales et fongiques pour l'établissement du contact avec l'hôte et l'échange de nutriments (Martin et al., 2016; Rytioja et al., 2014; Zhang et al., 2018). Les pathogènes de céréales *Magnaporthe grisea* et *Fusarium graminearum* possèdent un grand nombre de gènes codant pour des GH, des polysaccharides lyases et des pectine estérases qui leur servent à dégrader les tissus vivants, puis morts. Contrairement à ces champignons nécrotrophes, les pathogènes biotrophes comme *Ustilago maydis* et *Puccinia graminis*, qui dépendent d'hôtes vivants, possèdent plusieurs GH5 mais peu d'autres CAZymes ; cela leur évite probablement d'infliger des dégâts trop importants à leur hôte, et donc de déclencher des mécanismes de défense de la plante (Kämper et al., 2006).

C'est chez les champignons saprophytes que l'on trouve généralement la plus grande diversité d'enzymes actives sur la lignocellulose, mais là encore leurs stratégies varient. Parmi les dégradeurs du bois, on distingue communément trois catégories, nommées d'après l'aspect visuel des dégradations qu'ils causent : la pourriture blanche, qui regroupe des basidiomycètes capables de dégrader tous les composants de la paroi végétale, avec parfois une préférence pour la lignine ; la pourriture brune, composée de basidiomycètes qui produisent une dépolymérisation des polysaccharides pariétaux auxquels ils accèdent en modifiant la lignine ; et la pourriture molle, causée par des ascomycètes qui attaquent les fibres de cellulose dans la paroi secondaire, laissant un contenu important en lignines (Blanchette, 2000). Les champignons de la pourriture blanche, dont l'organisme

modèle est *Phanerochaete chrysosporium*, possèdent un grand nombre de gènes codant pour des enzymes actives sur la cellulose (LPMO et cellulases, notamment GH6 et GH7), les hémicelluloses et la lignine (laccases, peroxydases), ainsi que des modules de liaison à la cellulose (CBM1). A l'inverse, les champignons de la pourriture brune, tels que *Postia placenta* et *Serpula lacrymans*, possèdent des hémicellulases, mais très peu de cellulases et pas de peroxydase. La dégradation de la cellulose et la modification de la lignine se font essentiellement par voie non-enzymatique, grâce à l'action oxydative de radicaux libres, notamment des radicaux hydroxyles, produits par des réactions de Fenton extracellulaires. Il existe plusieurs espèces possédant des modes d'action intermédiaires (par exemple *Botryobasidium botryosum* et *Jaapia argillacea*), ce qui démontre que la séparation entre pourriture blanche et brune n'est pas si dichotomique, et dont l'étude a mené à l'utilisation du terme de « pourriture grise » (Floudas et al., 2012; Riley et al., 2014; Rytioja et al., 2014). Les mécanismes de colonisation du bois par les ascomycètes de la pourriture molle, dont font partie les genres *Trichoderma* et *Aspergillus*, sont moins bien étudiés que ceux des champignons des pourritures brune et blanche, bien qu'ils soient capables d'attaquer la biomasse dans des environnements particulièrement chauds, froids ou humides qui inhibent la colonisation par les autres dégradeurs (Hamed, 2013). Une étude basée sur une quarantaine de génomes a montré que le contenu en laccases et peroxydases des champignons de la pourriture molle était plus faible que celui des champignons de la pourriture blanche, reflétant leur moindre capacité de dégradation de la lignine ; en revanche, le nombre de gènes codant pour des cellulases et des LPMO était équivalent, et le nombre d'hémicellulases et de pectinases était supérieur (Sista Kameshwar and Qin, 2018).

Certains champignons sont donc extrêmement bien adaptés à la dégradation de la biomasse, et sont capables de produire un grand nombre d'enzymes pour dégrader les matériaux riches en cellulose. Parmi eux, le plus étudié pour ses capacités cellulolytiques est certainement l'ascomycète saprotrophe *Trichoderma reesei*.

3.2 Utilisation du secrétome de *Trichoderma reesei*

3.2.1 Système cellulolytique de *Trichoderma reesei*

Trichoderma reesei a été découvert pendant la Seconde Guerre mondiale, dans les Iles Salomon, lorsqu'il a été isolé sur des tentes et uniformes en coton en train de se décomposer dans les camps américains. Un programme de recherche a été établi par l'armée américaine, visant à comprendre et contrôler cette dégradation, ainsi qu'à étudier les possibles utilisations de ce phénomène. La souche "QM6a", initialement repérée pour son impressionnante capacité à dégrader la cellulose cristalline, est toujours considérée comme une souche de référence. Les travaux effectués sur les enzymes de cette souche sont à l'origine des connaissances que nous avons aujourd'hui des différentes cellulases et de leur action synergique (voir

paragraphe 2.2 de ce chapitre). La CBHI de *T. reesei* a par exemple été la première cellulase eucaryote à être clonée, et la première cellulase dont la structure 3D a été résolue (Divne et al., 1994; Shoemaker et al., 1983).

Afin d'améliorer les performances de la souche sauvage, des stratégies de mutagenèse aléatoire et de criblage ont été mises en place dès les années 1970. En utilisant des techniques d'irradiation ou de mutagenèse chimique (voir Figure 21), des souches hyper-productrices de cellulases avec des niveaux de production largement supérieurs à ceux de la souche sauvage ont pu être obtenues, notamment la souche Rut-C30, qui est depuis des décennies une référence pour la recherche académique, ou la souche CL847, qui fait partie des souches utilisées à l'échelle industrielle (Durand et al., 1988; Eveleigh and Montenecourt, 1979; Peterson and Nevalainen, 2012).

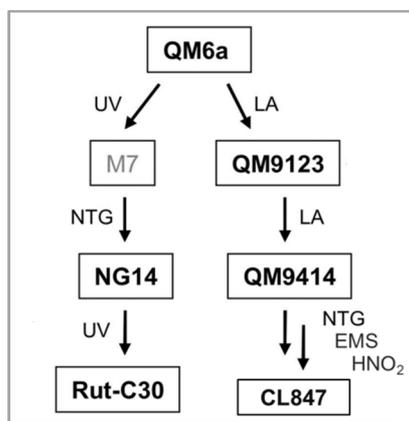


Figure 21 : Illustration de la généalogie de quelques souches mutantes de *Trichoderma reesei* (Vitikainen et al., 2010)
Techniques de mutagène utilisées : accélérateur linéaire (LA), rayons ultra-violetts (UV), N-nitroguanidine (NTG), sulfonate d'éthyle méthane (EMS), acide nitreux (HNO₂).

T. reesei étant très étudié et utilisé, son génome a été séquencé en 2008, révélant un nombre étonnamment faible de gènes impliqués dans la dégradation de la biomasse végétale, étant donné son efficacité. Le nombre et la diversité des gènes de *T. reesei* codant pour des cellulases et des hémicellulases est en effet inférieur à beaucoup d'autres champignons cellulolytiques dont le génome est connu. Il possède seulement 2 gènes codant pour des cellobiohydrolases, 4 gènes codant pour des endoglucanases caractérisées (et 3 autres prédites), 3 gènes codant pour des LPMO de la famille AA9 (dont une seule a été caractérisée), ainsi que 2 gènes codant pour des β -glucosidases (9 autres sont prédits). En plus de ces enzymes actives sur la cellulose, résumées dans le Tableau 4, le génome de *T. reesei* contient une quinzaine de gènes codant pour des hémicellulases caractérisées, et une trentaine d'autres prédites, avec notamment des activités xyloglucanase, endoxygénase, xylosidase, mannanase, mannosidase, arabinofuranosidase, acétyl xylane estérase, galactosidase, glucuronidase, fucosidase (Häkkinen et al., 2012; Martinez et al., 2008). D'autres enzymes "accessoires" portant un CBM1, sont impliquées dans la dégradation de la lignocellulose : CIP1 (« cellulose induced protein » 1), qui a des similarités structurelles avec les lyases mais dont la fonction n'est aujourd'hui pas

connue ; CIP2, qui est une glucuronoyl estérase de la famille CE15 pouvant cliver des liaisons entre la lignine et l'hémicellulose ; et SWO1, une swollenine dont le rôle serait de déstabiliser l'organisation de la cellulose cristalline (Häkkinen et al., 2012; Saloheimo et al., 2002).

Tableau 4 : Enzymes caractérisées de *T. reesei* ayant une activité sur la cellulose

Fonctions	Noms usuels	Familles CAZY	
Cellobiohydrolase	CBH I (Cel7A)	GH7	CBM1
	CBH II (Cel6A)	GH6	CBM1
Endoglucanase	EG I (Cel7B)	GH7	CBM1
	EG II (Cel5A)	GH5	CBM1
	EG III (Cel12A)	GH12	
	EG V (Cel45A)	GH45	CBM1
β-glucosidase	BGL I (Cel3A)	GH3	
LPMO	LPMO9A (Cel61A/EG IV)	AA9	CBM1

Pour dégrader efficacement la cellulose, *T. reesei* doit donc compter sur de hauts niveaux de production de quelques cellulases, ce qui est possible grâce à un nombre important de gènes impliqués dans la détection des substrats et le transport membranaire, ainsi que l'induction et la sécrétion des cellulases (Gupta et al., 2016).

3.2.2 Etudes transcriptomiques et protéomiques

L'expression d'un grand nombre de gènes précédemment cités n'est pas constitutive, mais induite par la présence de substrats lignocellulosique ou de disaccharides dérivés de la cellulose, tels que le cellobiose et son produit de transglycosylation, le sophorose. Ces inducteurs solubles seraient formés à partir de la cellulose grâce à une faible expression basale de quelques cellulases, dont certaines attachées à la paroi cellulaire, pour pouvoir ensuite être assimilés grâce à des cellobiose perméases et activer plusieurs facteurs de transcription ; le lactose (un dimère de glucose et de galactose) permet également d'induire la sécrétion de nombreuses cellulases, par un mécanisme encore mal compris (Kubicek et al., 2009). Le rôle des facteurs de transcription dans l'expression des gènes impliqués dans la dégradation de la biomasse est complexe, et tous les acteurs du réseau de régulation de *T. reesei* ne sont pas connus, cependant plusieurs facteurs de transcription majeurs ont déjà été identifiés. XYR1 est l'un des principaux activateurs de transcription contrôlant la synthèse de cellulases, xylanases et autres enzymes ; ACE2 et ACE3, ainsi que le complexe de protéines HAP2/3/5 ont également été identifiés comme activateurs. CRE1 est un acteur majeur de la répression catabolique par le carbone : il empêche la transcription de ses gènes cibles lorsque des sources de carbone plus favorables que la cellulose, telles que le glucose, sont présentes. ACE1 est également un répresseur de l'expression des cellulases et xylanases (Bischof et al., 2016; Kubicek et al., 2009; Portnoy et al., 2011a; Silva-Rocha et al., 2014).

Grâce à l'utilisation de techniques de transcriptomique (puces à ADN, qPCR, RNA-seq) et de protéomique (électrophorèse 2D, chromatographie en phase liquide, spectrométrie de masse), de nombreuses études ont montré que la composition du transcriptome et du protéome extracellulaire (sécrétome) de *T. reesei* varie en fonction de la source de carbone utilisée (Adav et al., 2012; dos Santos Castro et al., 2014; Häkkinen et al., 2012; Jun et al., 2013) (voir exemple Figure 22). Les inducteurs utilisés sont nombreux (cellobiose, sophorose, lactose, xylose, cellulose, xylane, épicea, paille de blé, bagasse de canne à sucre, rafle de maïs, sciure de bois...) et les souches de *T. reesei* testées ne sont pas toujours les mêmes, mais bien qu'elle varie selon les conditions de production, la composition du cocktail d'enzymes secrétées reste dominée par les cellobiohydrolases, suivies des endoglucanases.

Dans les sécrétomes des souches hyper-productrices Rut-C30 et CL847 cultivées sur lactose, les cellobiohydrolases Cel6A et Cel7A représentent entre 70% et 80% du total des protéines ; les endoglucanases Cel7B et Cel5A représentent environ 12%, alors que la β -glucosidase Cel3A représente seulement entre 0.5 et 1.5%, et la AA9 Cel61A moins de 1% du total (He et al., 2014; Herpoël-Gimbert et al., 2008). L'activité β -glucosidase des sécrétomes est généralement assez faible, probablement parce qu'une partie des enzymes ayant cette activité reste fixée à la paroi cellulaire du champignon (Bischof et al., 2016). Or les β -glucosidases sont essentielles car elles permettent de lever l'inhibition des autres cellulases par le cellobiose ; l'amélioration de cette activité a donc depuis longtemps fait partie des stratégies prioritaires d'amélioration du sécrétome de *T. reesei* (voir paragraphes suivants).

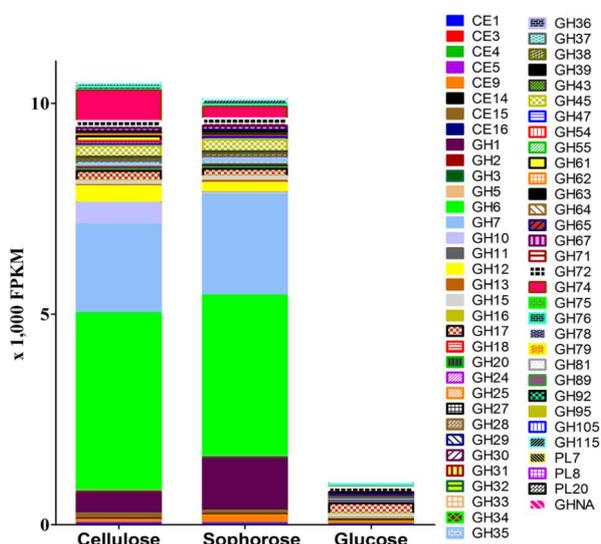


Figure 22 : Expression des gènes codant pour différentes familles de CAZymes par *T. reesei* Rut C30 cultivée sur glucose, sophorose et glucose (dos Santos Castro et al., 2014). Les données sont issues d'une analyse RNA-seq, et exprimées en fragments par kilobase d'exon par million de fragments cartographiés (FPKM).

3.2.3 Ingénierie génétique des souches

L'utilisation de techniques de génomique, transcriptomique et sécrétomique comparatives permet d'étudier les différentes générations de souches obtenues par mutagenèse aléatoire, afin de mieux comprendre les mutations impliquées dans le phénomène d'hyperproduction de cellulases, et de pouvoir s'en inspirer pour des approches de génie génétique plus ciblées (Herpoël-Gimbert et al., 2008; Kubicek, 2013; Le Crom et al., 2009; Poggi-Parodi et al., 2014; Portnoy et al., 2011b; Vitikainen et al., 2010). L'analyse du génome de la souche Rut-C30 a montré qu'un total de 74 gènes étaient affectés par des mutations, notamment des gènes impliqués dans le transport nucléocytoplasmique, le métabolisme des ARN messagers, la sécrétion des protéines et le tri vacuolaire, ainsi que de nombreux facteurs de transcription (Le Crom et al., 2009) ; une des premières mutations à avoir été identifiée est un raccourcissement du répresseur de transcription CRE1, qui le transforme en un activateur (Ilmén et al., 1996; Rassinger et al., 2018).

Les facteurs de transcription représentent une cible logique pour l'ingénierie génétique dans le but d'améliorer la production d'enzymes ; d'autres approches actuellement explorées incluent la modification des protéines impliquées dans le remaniement de la chromatine des gènes d'intérêt, ainsi que la modification directe de leurs promoteurs (Gupta et al., 2016). Ces approches s'accompagnent du développement d'outils spécifiques pour la transformation génétique de *T. reesei*, notamment l'utilisation de marqueurs d'auxotrophie/prototrophie et de systèmes de recyclage des marqueurs, l'amélioration de l'intégration spécifique par recombinaison homologue grâce à la délétion de gènes impliqués dans la jonction d'extrémités non homologues, ainsi que la mise au point de techniques de croisement sexué et l'utilisation du système CRISPR/Cas9 (Derntl et al., 2015; Liu et al., 2015; Seidl and Seiboth, 2010).

Un bon exemple d'amélioration de *T. reesei* par ingénierie génétique est la production d'une souche recombinante sur-exprimant le gène *bgl1*, grâce au remplacement par recombinaison homologue de son promoteur par des promoteurs plus efficaces ; cette modification a permis d'augmenter la quantité de β -glucosidase I sécrétée, et de multiplier l'activité β -glucosidase par 4.0 et 7.5 en utilisant respectivement les promoteurs de *egl3* et de *xyn3* (Rahman et al., 2009).

3.2.4 Amélioration des enzymes

Afin d'améliorer les performances du sécrétome de *T. reesei*, il est également possible d'améliorer l'efficacité des enzymes individuelles et de mieux les adapter aux conditions industrielles dans lesquelles elles seront utilisées, notamment en modifiant leur activité spécifique sur les substrats insolubles, leur thermostabilité, leur pH optimal, ou encore leur tolérance aux inhibiteurs (Zhang et al., 2006). Deux types de techniques d'ingénierie d'enzymes peuvent être utilisées pour cela : le design rationnel de protéines, basé sur des

mutations ciblées de résidus impliqués dans le mécanisme catalytique et la liaison aux substrats ; ou l'évolution dirigée des protéines, basée sur des cycles de modifications aléatoires suivies de criblage des mutants (voir Figure 23).

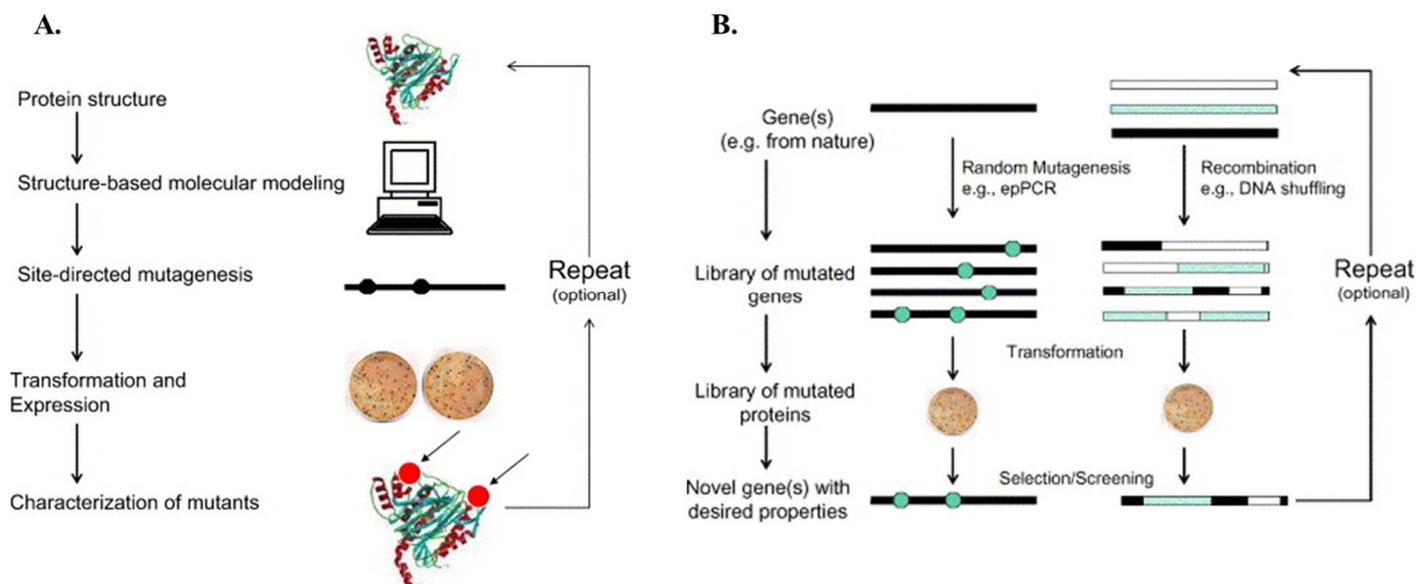


Figure 23 : Différentes approches d'ingénierie pouvant être appliquées aux enzymes de *T. reesei* (Zhang et al., 2006)

A. Design rationnel ; **B.** Evolution dirigée.

En utilisant la première approche, et grâce à la mise au point d'une souche spécifique pour l'expression des mutants et de plusieurs tests de criblage à haut débit, Lantz et al. (2010) ont obtenu des cellobiohydrolases Cel7A et Cel7B ayant une meilleure thermostabilité, et ont réussi à améliorer l'activité spécifique de Cel6A sur biomasse lignocellulosique tout en gardant ces nouvelles propriétés de stabilité. Plusieurs approches d'évolution dirigée ont également été tentées. Wang et al. (2005) ont utilisé la technique de la PCR sujette à erreur, ou "error prone PCR" (epPCR) pour améliorer la tolérance de l'endoglucanase EGIII aux pH alcalins. Dans une autre étude, Ayrinhac et al. (2011) ont généré des variants du gène *bgl1* grâce à des recombinaisons avec des gènes proches issues de *Chaetomium globosum* et *Neurospora crassa* grâce à une technique appelée « *L-shuffling* ». Deux cycles de recombinaisons et de criblage ont permis d'identifier un nouveau gène, qui une fois exprimé dans *T. reesei* a conduit à l'obtention d'un cocktail cellulolytique plus efficace, permettant de diviser par quatre la quantité d'enzymes utilisées en gardant la même efficacité pour la dégradation de biomasse que le cocktail de la souche parente. Afin d'améliorer la thermostabilité des cellobiohydrolases par évolution dirigée, le groupe de Frances Arnold a développé une approche computationnelle nommée SCHEMA, qui permet d'identifier des blocs de séquences qui stabilisent la structure des protéines chimériques après recombinaison. Cette approche leur a permis de développer des familles de CBHI et CBHII actives à 63 voire 70°C (Heinzelman et al., 2009b; Heinzelman et al., 2010) et une étude plus approfondie leur a permis d'identifier certaines mutations responsables de cette thermostabilité accrue (Heinzelman et al., 2009a; Komor et al., 2012).

Les pistes à explorer pour l'amélioration du cocktail cellulolytique de *T. reesei* ne manquent donc pas, mais l'efficacité du cocktail n'est pas le seul paramètre à prendre en compte dans l'amélioration des rendements de l'étape d'hydrolyse enzymatique.

3.2.5 Mise en œuvre de l'hydrolyse enzymatique

Le cocktail cellulolytique de *T. reesei* est certes efficace, mais son utilisation en vue d'une application industrielle nécessite de prendre en compte un certain nombre de facteurs qui peuvent influencer la vitesse initiale de l'hydrolyse ou le taux de conversion final du substrat en sucres. Une hydrolyse enzymatique se déroule en effet typiquement selon trois phases : elle débute par une première phase, assez courte, de conversion rapide, puis la réaction ralentit dans la phase intermédiaire, et dans la dernière phase le taux de conversion n'évolue quasiment plus. Malgré des temps d'hydrolyse longs et des doses de cellulases élevées, ce taux maximum reste très souvent inférieur à 100%, ce qui signifie qu'une partie de la cellulose ne peut pas être hydrolysée (Figure 24).

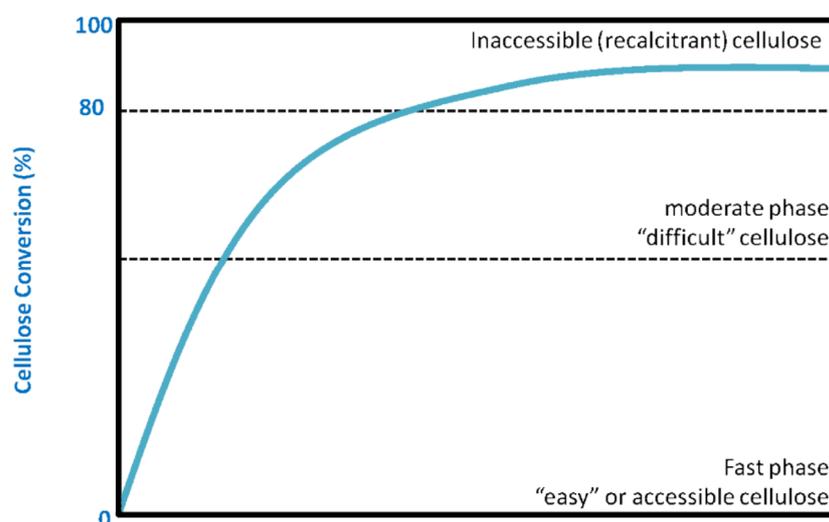


Figure 24 : Déroulement typique de l'hydrolyse enzymatique de la cellulose (Arantes and Saddler, 2011)

Le ralentissement de l'hydrolyse et son arrêt prématuré sont des problèmes majeurs, et plusieurs facteurs en lien avec la composition et la structure physique du substrat peuvent jouer un rôle important. Pour Arantes and Saddler (2011), le facteur limitant est celui de la surface de cellulose accessible aux cellulases, à la fois à l'extérieur des fibres et à l'intérieur, grâce à des pores et des fissures. Cette surface accessible dépend largement de la méthode de prétraitement utilisé, puisqu'elle peut faire varier la taille des particules de substrat et le volume des pores, ainsi que le degré de polymérisation de la cellulose et sa cristallinité. La topologie de la surface de la cellulose est également importante, puisque des obstacles peuvent bloquer l'avancée des cellulases processives sur la chaîne, créant des embouteillages qui ralentissent l'hydrolyse (Igarashi et al., 2011). D'autres facteurs de ralentissement sont liés à la quantité d'hémicelluloses et de lignine restantes après

le prétraitement, celles-ci pouvant couvrir les fibres de cellulose et diminuer leur accessibilité ; les cellulases peuvent également s'adsorber sur des particules de lignine et perdre ainsi leur activité. Enfin, selon l'efficacité du lavage qui a suivi le prétraitement, des produits issus de la lignine tels que l'acide formique peuvent être présents et inhiber les cellulases (Jørgensen et al., 2007a).

Un paramètre essentiel de l'hydrolyse est la concentration de substrat solide utilisée. Il est en effet essentiel pour la viabilité économique du procédé de travailler à des fortes concentrations de solide, qui permettent une économie d'eau et d'énergie, ainsi qu'une augmentation de la concentration de sucres puis d'éthanol dans le milieu, ce qui facilite sa récupération et sa distillation. Cependant il a été observé que le taux final de conversion décroît de manière linéaire lorsque l'on augmente la concentration de solide, ce qui est un frein à l'utilisation de telles conditions dans les procédés industriels (Kristensen et al., 2009). Plusieurs hypothèses ont été formulées pour expliquer cette limitation de l'hydrolyse. La phase aqueuse étant essentielle au transport des enzymes vers leur substrat et à l'évacuation de leurs produits, une limitation du transfert de masse dû à la viscosité importante du mélange pourrait limiter l'efficacité de l'hydrolyse. Cette viscosité rend l'agitation du milieu réactionnel dans un réacteur classique difficile au-delà d'une concentration de 10 à 15% de solide ; plusieurs stratégies ont donc été développées afin d'améliorer le transfert de masse, notamment en changeant le mode d'agitation (Du et al., 2017; Jørgensen et al., 2007b).

Une autre hypothèse largement acceptée est celle de l'inhibition des enzymes par leurs produits. En effet, une forte concentration de glucose peut inhiber l'activité des β -glucosidases, menant à une accumulation de cellobiose, qui est un inhibiteur puissant des cellobiohydrolases ; les sucres issus de la dégradation des hémicelluloses peuvent également inhiber les cellulases (Jørgensen et al., 2007a; Olsen et al., 2014). En plus de cette inhibition par les produits, Kristensen et al. (2009) ont également démontré l'influence d'une inhibition de l'adsorption des cellulases sur leur substrat, probablement du fait du glucose et du cellobiose, mais son mécanisme exact est encore inconnu.

Afin de surmonter ces limitations, il peut être tentant d'augmenter les doses d'enzymes utilisées, mais cela augmente également le coût du procédé ; il est donc nécessaire de trouver un compromis et de déterminer une dose minimale d'enzymes permettant une hydrolyse satisfaisante, sans que le temps nécessaire ne soit trop long. Les conditions optimales de fonctionnement des enzymes doivent également être déterminées ; pour les enzymes des genres *Trichoderma* et *Aspergillus*, la température optimale se situe entre 40 et 55°C, et le pH optimal entre 3,8 et 5,4 (Tengborg et al., 2001). Il faut également prendre en compte la possible dégradation des enzymes par protéolyse, ou par des températures trop élevées s'ils ont une faible stabilité thermique.

Tous ces facteurs sont absolument déterminants pour le rendement et le coût de l'opération, et le choix des conditions de mise en œuvre de l'hydrolyse doit être fait avec soin. En plus des facteurs permettant d'optimiser les conditions d'hydrolyse, il peut être avantageux d'envisager une amélioration des enzymes utilisées.

Cependant, celles-ci ayant déjà fait l'objet de nombreuses modifications, la marge d'amélioration restante est assez faible. En dehors des techniques de modification des souches et d'ingénierie des enzymes du sécrétome de *T. reesei* évoquées précédemment, la recherche d'autres enzymes permettant une hydrolyse enzymatique performante, issues d'autres organismes, peut aussi être envisagée.

3.3 Stratégies de supplémentation

3.3.1 Limitations du sécrétome de *T. reesei*

Certaines activités nécessaires à la dégradation complète de la biomasse lignocellulosique sont absentes du génome de *Trichoderma reesei* (notamment les activités pectate lyase, pectine estérase ou féruloyle estérase) ; d'autres sont présentes seulement à travers un nombre limité de gènes et/ou ne sont que très faiblement exprimées dans les sécrétomes. C'est le cas notamment des LPMO et de plusieurs hémicellulases, alors que leur action est essentielle pour une saccharification complète de la biomasse (Bischof et al., 2016; Martinez et al., 2008). Les différents prétraitements existants n'étant pas capables de dégrader complètement les hémicelluloses et les pectines, la portion restante est susceptible d'empêcher l'accès à la cellulose. Ainsi, de nombreuses études ont montré l'existence de synergies entre des cellulases et plusieurs enzymes "auxiliaires" (xyloglucanase, xylanase, β -xylosidase, β -mannosidase, α -glucuronidase, α -L-arabino-furanosidase, α -galactosidase, acétyle xylane estérase, acide férulique estérase, féruloyle estérase...) qui améliorent la dégradation de diverses biomasses prétraitées, les combinaisons d'enzymes les plus efficaces dépendant évidemment de la composition du substrat et du prétraitement utilisé (Hu et al., 2013; Świątek et al., 2014; van Dyk and Pletschke, 2012). Dans le cas du cocktail de *T. reesei*, plusieurs études ont montré que l'ajout de préparations commerciales contenant des activités pectinases, xylanases, β -xylosidase et autres hémicellulases permettait d'améliorer les performances sur diverses biomasses (Berlin et al., 2007; Qing and Wyman, 2011; Sills and Gossett, 2011; Świątek et al., 2014). Suite à la découverte des LPMO et de leur action synergique avec les cellulases, des AA9 ont également été ajoutées aux nouvelles générations de cocktail enzymatiques commerciaux produits par *T. reesei*, tels que le Cellic CTec2 (Cannella et al., 2012).

3.3.2 La biodiversité fongique comme source d'inspiration

La stratégie de *Trichoderma reesei* consistant à produire un nombre limité d'enzymes semble être une exception au sein du règne fongique (Rytioja et al., 2014), et il est facile de trouver des espèces exprimant un répertoire d'enzymes plus varié. De plus, dans la nature, les champignons dégradant la biomasse végétale vivent en communauté avec d'autres organismes, et une espèce n'est pas destinée à assurer seule cette dégradation. Les champignons filamenteux ont donc développé des stratégies complémentaires, en sécrétant

des enzymes agissant de manière séquentielle sur différentes parties de la biomasse, afin de dégrader efficacement les polymères (Benoit et al., 2015). Il peut donc être avantageux d'explorer la biodiversité fongique dans le but de trouver des activités enzymatiques capables d'améliorer l'hydrolyse de la lignocellulose. Des enzymes lignocellulolytiques ont par exemple été identifiées chez des organismes fongiques isolés dans la forêt tropicale (Berrin et al., 2012), sur des arbres de mangrove (Arfi et al., 2013), sur des feuilles en décomposition (Shrestha et al., 2015), dans le système digestif d'herbivores ou d'insectes (Rojas-Jiménez and Hernández, 2015; Thareja et al., 2006), et même sur des éponges marines (Batista-García et al., 2017).

Dans les démarches de supplémentation du cocktail de *Trichoderma reesei*, les espèces du genre *Aspergillus* ont souvent été envisagées. Il s'agit de l'un des genres fongiques les plus connus et étudiés, dont plusieurs espèces sont déjà utilisées à l'échelle industrielle pour la production d'acides organiques et d'enzymes impliquées dans la dégradation ou la modification des polysaccharides végétaux, avec des applications dans l'alimentation humaine et animale, ou les industries textile et papetière. Ces ascomycètes sont en effet capables de se développer dans des habitats très différents, avec un mode de vie généraliste adapté à tout type de biomasse, grâce à la sécrétion d'une grande variété d'enzymes impliquées dans la dégradation de la lignocellulose (Culleton et al., 2013; de Vries et al., 2017; de Vries and Visser, 2001). Les souches d'*Aspergillus* sont beaucoup utilisées pour la production de β -glucosidases, pouvant servir à compléter les cocktails de *T. reesei*. ; des préparations commerciales de β -glucosidases sont issues d'*Aspergillus niger*, qui a également été étudié pour sa production de pectinases, xylanases, mannanases et estérases (van den Brink et al., 2014). L'étude des transcriptomes et des sécrétomes de *Aspergillus niger* et *Trichoderma reesei* révèle quelques différences dans les profils d'enzymes sécrétées par les deux organismes, et l'utilisation de mélanges des deux sécrétomes a un effet positif sur l'hydrolyse de la biomasse (Borin et al., 2017; van den Brink et al., 2014).

3.3.3 Supplémentation par des sécrétomes fongiques

Des études génomiques peuvent permettre d'identifier, chez des champignons ayant un répertoire de CAZymes plus varié que celui de *T. reesei*, des gènes codant pour des enzymes absentes de *T. reesei* ou ayant de faibles homologues avec celles qu'il produit. Ainsi, deux mannanases des familles GH5 et GH26 ainsi qu'une xylanase (GH11) ont été identifiées chez *Podospira anserina* et se sont révélées être individuellement capables de compléter le sécrétome de la souche industrielle CL847 (Couturier et al., 2011). Une féruloyl estérase issue d'*Aspergillus niger* a également permis d'améliorer le rendement d'hydrolyse de paille de blé lorsqu'elle a été ajoutée à un sécrétome de *T. reesei* contenant principalement des cellulases et des xylanases (Tabka et al., 2006).

Afin de ne pas se limiter à l'ajout d'enzymes fongiques isolées, il peut être intéressant d'employer une stratégie consistant à tester une variété de secrétomes fongiques pour déterminer lesquels sont capables de compléter les cocktails cellulolytiques existants, puis à les explorer pour trouver la ou les enzyme(s) responsable(s) (Poidevin et al., 2014). L'utilisation de secrétomes, constitués de toutes les enzymes sécrétées par un organisme dans un environnement particulier, offre un niveau de diversité élevé. En effet, leur composition varie en fonction des conditions dans lesquelles l'organisme a été cultivé, sous le contrôle de mécanismes de signalisation et de régulation transcriptionnelle responsables de l'induction ou de la répression de l'expression de certains gènes (Aro et al., 2005; Brown et al., 2014). Cependant, des méthodes de plus en plus poussées d'identification des protéines facilitent leur exploration et leur caractérisation (Bouws et al., 2008; Di Cologna et al., 2017).

Cette stratégie de supplémentation a permis d'identifier plusieurs organismes fongiques dont le secrétome est capable d'améliorer l'efficacité du cocktail de *T. reesei* : *Chaetomium thermophilum* et *Penicillium funiculosum* (Rosgaard et al., 2006), *Aspergillus awamori* (Gottschalk et al., 2010), *Ustilago maydis* (Couturier et al., 2012), *Fusarium verticillioides* (Ravalason et al., 2012), *Podospora anserina* (Poidevin et al., 2014), *Laetisaria arvalis* (Navarro et al., 2014), *Sclerotium rolfsii* (Kracher et al., 2014), un mutant d'*Aspergillus niger* (Peciulyte et al., 2017), *Aspergillus terreus* et *Achaetomium strumarium* (Brar et al., 2018). Dans plusieurs cas, l'étude de l'activité et de la composition des secrétomes a permis d'émettre des hypothèses sur l'identité des enzymes responsables de l'amélioration de la saccharification.

4 Enjeux et stratégie du projet de recherche

L'étude de la littérature montre que le développement d'un procédé rentable d'utilisation de biomasse lignocellulosique pour la production de biocarburants passe notamment par la diminution des quantités d'enzymes utilisées (NREL, 2014). Cela implique d'améliorer les performances du cocktail cellulolytique de *Trichoderma reesei*. A cet effet, l'une des stratégies envisageables est l'exploration des répertoires enzymatiques de différents organismes fongiques, sécrétés en réponse à diverses conditions de croissance, et qui offrent potentiellement une grande diversité d'enzymes encore inconnues.

En 2012, dans le cadre d'un projet mené notamment par IFP Energies Nouvelles et l'unité BBF (Biodiversité et Biotechnologie Fongiques) de l'INRA, les secrétomes de nombreuses souches fongiques récoltées en Guyane et en France métropolitaine avaient été criblés pour leur capacité à compléter un cocktail de *T. reesei* pour la saccharification de paille de blé prétraitée (Berrin et al., 2012). Une souche d'*Aspergillus japonicus* (conservée dans la collection du CIRM-CF sous le numéro BRFM405) avait été repérée pour ses bonnes performances et avait fait l'objet d'un brevet (Berrin et al., 2013). Le secrétome avait été fractionné et

les performances des différentes fractions testées, mais cela n'avait pas permis de trouver les enzymes responsables du « boost » de saccharification observé.

Ce projet de thèse s'appuie sur les résultats de cette étude préliminaire. Il vise à **identifier, produire et caractériser de nouvelles enzymes susceptibles d'améliorer les performances du cocktail cellulolytique de *Trichoderma reesei***, à partir des sécrétomes de plusieurs souches du genre *Aspergillus*.

En effet, afin de pouvoir mener une étude comparative sur plusieurs sécrétomes, en plus de la souche BRFM405, trois autres souches d'*A. japonicus* ont été incluses dans ce projet, ainsi qu'une souche d'*Aspergillus niger*, récoltée elle aussi en Guyane. Certains gènes pouvant être exprimés de manière spécifique sur certains inducteurs, en fonction de leur composition, il a été décidé de cultiver les cinq souches sur trois inducteurs distincts : du son de maïs, de la pulpe de betterave et de la cellulose microcristalline.

Une fois ces sécrétomes produits, une étude protéomique était envisagée afin d'annoter les différentes protéines sécrétées à partir du génome des souches. En parallèle, une méthode devait être développée afin d'évaluer la capacité des sécrétomes à améliorer l'hydrolyse de plusieurs biomasses d'intérêt industriel de compositions différentes (paille de blé, Miscanthus et peuplier), en supplémentation d'un cocktail de *T. reesei* de d'IFP Energies Nouvelles. Ces deux approches devaient permettre de sélectionner une ou plusieurs protéines d'intérêt, pouvant être clonées et exprimées individuellement chez *Pichia pastoris*, afin de caractériser leur activité sur différents substrats, et d'évaluer leur apport à la saccharification de biomasse.

Les objectifs de ce projet étaient donc les suivants :

- Obtenir des cocktails enzymatiques de compositions variées
- Mettre au point et utiliser un test de criblage des performances de ces cocktails enzymatiques
- Rassembler et analyser les données collectées pour identifier des protéines cibles
- Produire ces protéines de manière hétérologue et les purifier
- Caractériser les enzymes sélectionnées et comprendre leur mécanisme d'action
- Obtenir un effet synergique avec les enzymes de *T. reesei*

Chapitre II : Matériel et méthodes

1 Préparation des secrétomes

1.1 Souches fongiques

Les souches d'*Aspergillus japonicus* et d'*Aspergillus niger* utilisées dans cette étude sont maintenues dans la collection du Centre International de Ressources Microbiennes - Champignons Filamenteux (CIRM-CF, INRA, Marseille, France; <http://www.inra.fr/crb-cirm/>) sous les numéros d'accèsion CIRM BRFM 405, 430, 1487, 1489 et 1490.

1.2 Milieux de culture

Le milieu gélosé utilisé, nommé MYA2, a été préparé avec 20 g/L d'extrait de malt, 1 g/L d'extrait de levure et 20 g/L d'agar. Trois milieux de culture liquides ont été utilisés, contenant chacun une source de carbone différente, utilisée en tant qu'inducteur de sécrétion de protéines : soit une fraction autoclavée de son de maïs (ASM), soit de la pulpe de betterave (SBP), soit de la cellulose microcristalline (Avicel). Les milieux contenant du son de maïs ou de la pulpe de betterave ont été préparés de la manière suivante : 15 g/L d'inducteur (ASM ou SBP, fournis par ARD, Pomacle, France); 2,5 g/L de maltose; 1,84 g/L de tartrate diammonium comme source d'azote; 0,5 g/L d'extrait de levure; 0,2 g/L de KH_2PO_4 ; 0,0132 g/L de $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ et 0,5 g/L de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$. Le troisième milieu inducteur a été préparé en utilisant 4 g/L d'Avicel PH-101 (Sigma-Aldrich, Saint-Quentin Fallavier, France); 10 g/L de xylose; 1,8 g/L de tartrate diammonium; 0,5 g/L d'extrait de levure; 0,2 g/L de KH_2PO_4 ; 0,0132 g/L de $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ et 0,5 g/L de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$.

1.3 Culture et récolte des secrétomes

Les souches choisies dans la collection du CIRM-CF ont été repiquées sur milieu gélosé MYA2, et incubées à 30°C pendant 12 à 15 jours, jusqu'à la sporulation. Les spores ont été récoltées et leur concentration a été évaluée en utilisant une cellule de Thoma. Chacune des cinq souches a été inoculée à une concentration de 2×10^5 spores/mL dans neuf fioles bafflées contenant 100 mL de milieu liquide, de manière à obtenir un triplicat pour chacun des trois milieux décrits plus haut (ASM, SBP, Avicel). Les fioles ont été incubées pendant 7 jours à 30°C, dans l'obscurité et sous agitation rotative à 105 rpm (Infors, Massy, France). Les cultures ont ensuite été arrêtées, les triplicats ont été réunis, et le surnageant liquide a été séparé du mycélium à l'aide de Miracloth (Merck-Millipore, Darmstadt, Allemagne). 40 mL de chaque surnageant ont été filtrés sur des membranes de polyéthersulfone 0,22 μm (Merck-Millipore) puis diafiltrés et concentrés sur des membranes de polyéthersulfone avec un seuil de coupure de 10kDa (Vivaspin, Sartorius, Göttingen, Allemagne) dans du tampon acétate de sodium 50mM pH5 jusqu'à un volume final de 2 mL. Les secrétomes ont été conservés à -20°C, et leur concentration totale de protéines a été évaluée par les méthodes de Bradford (Bio-Rad, Ivry, France) et du BCA (Bicinchoninic Acid Protein Assay, Sigma-Aldrich) en utilisant une gamme standard d'albumine sérique bovine (BSA).

2 Tests de saccharification en supplémentation

2.1 Substrats utilisés

La paille de blé, le Miscanthus et le peuplier prétraités ont été obtenus auprès d'IFP Energies Nouvelles (Rueil-Malmaison, France). Ces trois biomasses ont été prétraitées par explosion à la vapeur en conditions acides, lavées à l'eau chaude pour éliminer les produits libres, et séchées respectivement à 55°C, à 37°C et à température ambiante. Elles ont ensuite été broyées et tamisées afin de ne retenir que les particules de taille inférieure à 0,8 mm, et leur teneur en eau (% m/m) a été déterminé par séchage à 105°C en étuve à convection naturelle (VWR, Fontenay-sous-Bois, France). Leur composition en sucres a été déterminée par Estelle Bonnin (BIA, INRA, Nantes) par chromatographie en phase gazeuse après hydrolyse acide, selon une méthode décrite par Morris et al. (2010). Cette composition est présentée dans le Tableau 7 (p. 97).

2.2 Cocktail enzymatique de référence

Le cocktail K975, fourni par IFP Energies Nouvelles, est constitué du sécrétome de la souche CL847 de *Trichoderma reesei*, produit en présence de lactose, dont l'activité β -glucosidase spécifique est de 0,8 UI/mg. Il a été complété par le cocktail commercial de β -glucosidases d'*Aspergillus niger* SP188 (Novozyme, Bagsværd, Danemark). Le principe des tests de supplémentation repose sur la comparaison des performances de ce cocktail de référence seul avec celles du cocktail supplémenté par les différents sécrétomes étudiés.

2.3 Tests de saccharification en tubes

Les tests de saccharification en tubes de 2 mL ont été réalisés en triplicats en présence de paille de blé, de Miscanthus ou de peuplier à 5% de matière sèche (m/v), dans un volume réactionnel total de 1 mL. Les substrats ont été pesés, puis incubés dans du tampon acétate de sodium 50 mM (pH 4,8) et du chloramphénicol (0,1 g/L), pendant au moins 1h à 45°C sous agitation rotative à 850 rpm, afin que la biomasse soit bien imprégnée. On a ensuite ajouté le cocktail K975, à raison de 5 mg/g de matière sèche (MS), et le cocktail SP188, dans des quantités permettant d'obtenir une activité β -glucosidase totale de 80 UI/g MS, et enfin 10 μ L de sécrétomes (pour les conditions de supplémentation) ou 10 μ l de tampon (pour les conditions de référence). L'hydrolyse a été réalisée à 45°C sous agitation rotative à 850 rpm. A chaque temps de prélèvement, les triplicats ont été ébouillantés à 100°C pendant au moins 5 min pour inactiver les enzymes, puis refroidis et centrifugés à 15000xg pendant 5 min, et le surnageant a été prélevé.

2.4 Tests de saccharification en microplaques

Les tests de saccharification miniaturisés ont été réalisés en microplaques de 96 puits, dans lesquelles ont été distribués 100 μ L de suspension de biomasse à 6,3% de matière sèche (m/v) dans du tampon acétate de sodium 50 mM (pH 4,8) en présence de chloramphénicol (0,1 g/L). Pendant la distribution, la suspension de biomasse a été prélevée à l'aide d'une pipette multicanaux dans un bécher sous agitation magnétique constante, afin que

la répartition de la biomasse dans les puits soit la plus homogène possible. Les plaques ont ensuite été scellées et conservées à -20°C jusqu'à leur utilisation. Lors des tests de supplémentation, les enzymes ont été distribués à l'aide d'un robot Tecan Genesis Evo 200 (Tecan, Lyon, France) : les cocktails K975 et SP188 ont été distribués dans les concentrations indiquées plus haut, dans un volume de 15 µL, puis 10 µL de sécrétomes dilués au 10^{ème} ont été ajoutés (ou 10 µL de tampon dans les conditions de référence, présentes sur chaque plaque) pour atteindre un volume total de 125 µL. Chaque condition a été reproduite dans sept puits identiques, auxquels s'ajoute un puits contrôle contenant les enzymes seuls (sans biomasse). Une colonne de chaque plaque a été consacrée aux conditions contrôles contenant la biomasse seule (voir Figure 25). Les plaques ont ensuite été scellées et incubées à 45°C sous agitation rotative à 850 rpm, pendant 24 à 96h. A chaque temps de prélèvement, les plaques correspondantes ont été centrifugées après ajout de 120 µL de tampon, puis le surnageant a été filtré.

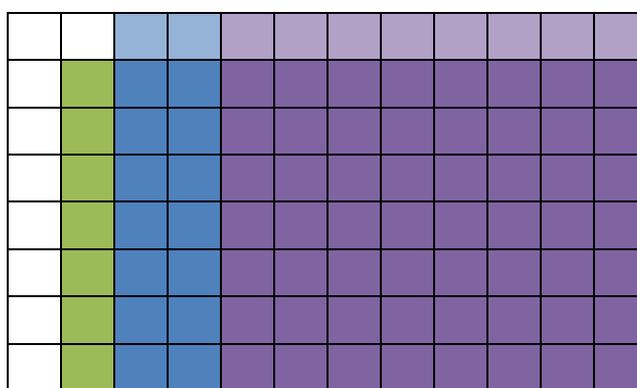


Figure 25 : Schéma d'une plaque utilisée pour les tests de saccharification miniaturisés.

Les puits représentés en vert contiennent la suspension de biomasse seule ; les puits en bleu foncé contiennent la suspension de biomasse et le cocktail de référence ; les puits en bleu clair contiennent le cocktail de référence seul ; les puits en violet foncé contiennent la suspension de biomasse, le cocktail de référence, et les sécrétomes (un différent dans chaque colonne) ; les puits violet clair contiennent uniquement le cocktail de référence et les sécrétomes (un différent dans chaque colonne). Chaque puits a été complété avec du tampon acétate de sodium 50 mM (pH 4,8) jusqu'à un volume de 125µL.

2.5 Quantification des produits et calcul des rendements

Pour les deux types de tests de saccharification, la concentration de glucose dans les échantillons a été mesurée par dosage spectrophotométrique à l'aide du réactif Glucose GOD-PAP (Biolabo, Maizy, France) en utilisant une gamme standard de glucose, et les rendements ont été calculés en tenant compte de la quantité de glucose cellulosique initialement présente. Plusieurs tests ont été réalisés pour chaque condition, et afin de pouvoir combiner les résultats, les rendements obtenus en présence de sécrétomes ont été traduits en pourcentage d'amélioration par rapport à la référence interne de chaque test de saccharification. Un test t de Student a été réalisé pour chaque condition afin de déterminer si la moyenne des résultats était statistiquement différente de la moyenne des références, en utilisant la *p*-value comme critère.

3 Analyse du contenu protéique des sécrétomes

3.1 Identification des protéines par LC-MS/MS

Pour chaque sécrétome, 15 µg de protéines ont été déposés sur gel SDS-PAGE précoulé 10% Tris-glycine (Mini-PROTEAN TGX, BioRad) pour une migration de 0,5 cm, puis les gels ont été colorés au bleu de Coomassie (BioRad). Chaque piste d'électrophorèse a été découpée en deux bandes de 2 mm de largeur et l'identification des protéines a été réalisée à la Plateforme d'Analyse Protéomique de Paris Sud-Ouest (PAPSSO, INRA, Jouy-en-Josas, France; <http://pappso.inra.fr/>), selon un protocole décrit par Navarro et al. (2014). Brièvement, la digestion des protéines contenues dans les bandes de gel a été effectuée selon un procédé standard de trypsinolyse, en utilisant de la trypsine modifiée (Promega, Charbonnières-les-Bains, France). L'analyse des peptides obtenus a été effectuée par chromatographie liquide Ultimate 3000 RSLCnano (Thermo Fisher Scientific, Waltham, Massachusetts, USA) couplée à un spectromètre de masse Q-exactive (Thermo Fisher Scientific) utilisant une source ionisation électrospray (ESI-MS). L'attribution des peptides et l'annotation des protéines a été réalisée par comparaison des données de spectrométrie de masse aux protéines prédites dans les génomes d'*Aspergillus japonicus* CIRM BRFM 405 (non publié), d'*Aspergillus aculeatus* ATCC16872 (<https://genome.jgi.doe.gov/Aspac1/Aspac1.home.html>) et d'*Aspergillus niger* CBS 513.88 / FGSC A1513 (<https://www.uniprot.org/proteomes/UP000006706>), en utilisant le logiciel X!Tandem (X!Tandem Cyclone, Jouy-en-Josas, France). L'annotation des protéines a été complétée manuellement par BLAST en utilisant la base de données NCBI nr (<http://www.ncbi.nlm.nih.gov/blast>), et les familles CAZy ont été annotées par Elodie Drula (AFMB, CNRS, Marseille) d'après la base de données CAZy.

3.2 Analyse des données

Les protéines identifiées par un seul peptide ont été retirées de la base de données de résultats. Les autres ont été regroupées en 501 groupes et 580 sous-groupes par le logiciel X!Tandem, selon les règles suivantes : les groupes sont formés de protéines ayant au moins un peptide en commun, et les sous-groupes sont composés de protéines d'un même groupe partageant tous leurs peptides détectés. Le nombre de protéines différentes identifiées est donc égal au nombre de sous-groupes. Cependant, parmi les 66 groupes contenant plusieurs sous-groupes, 59 sont composés uniquement de deux sous-groupes correspondant à deux protéines codées par des gènes homologues dans les différents génomes utilisés. Il a donc été décidé d'approximer le nombre de protéines différentes identifiées au nombre de groupes dans les différentes analyses présentées. Des diagrammes de Venn regroupant les nombres de protéines identifiées dans les différents sécrétomes ont été réalisés grâce à l'outil du groupe « Bioinformatics & Evolutionary Genomics » de l'Université de Gent (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). Pour l'estimation de l'abondance des protéines, la valeur du PAI (*Protein Abundance Index*) calculée par le logiciel X!Tandem a été utilisée ; le PAI est défini comme le nombre de spectres de fragmentation assignés avec un score significatif à une protéine identifiée, divisé par

le nombre de peptides tryptinolytiques théoriquement observables pour cette protéine. Une carte thermique (ou « *heatmap* ») de l'abondance des différentes familles de CAZymes dans les 15 sécrétomes, représentée par les valeurs de PAI, a été réalisée par double classification hiérarchique à l'aide de l'outil ClustVis (Metsalu and Vilo, 2015), en utilisant des variances normalisées, des matrices de distance calculées par corrélation de Pearson et des partitions regroupées par la méthode de Ward. Des comparaisons d'abondance des protéines entre les différents sécrétomes, normalisées par les quantités de sécrétomes utilisées pour les tests de saccharification, ont été réalisées en utilisant les fonctions du logiciel Microsoft Excel. En parallèle, une analyse manuelle des données focalisée sur les protéines de fonction inconnue a été réalisée. Elle a révélé la présence d'une protéine présentant des similitudes de séquence avec les LPMO de la famille AA10 d'après les outils en ligne BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) et Phyre 2 (Kelley et al., 2015, <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi>), et appartenant dans la base de données CAZy à une famille putative portant le numéro X273.

4 Analyse bioinformatique

Une recherche des séquences contenant un module X273 a été réalisée dans les génomes pris en compte par la base de données CAZy, issus principalement du NCBI (<https://www.ncbi.nlm.nih.gov/genome>) et du JGI (<https://mycocosm.jgi.doe.gov/mycocosm/home>). Cette recherche a également permis d'identifier les séquences contenant un autre module (CBM ou glycosylphosphatidylinositol (GPI)). Parmi les 1065 séquences trouvées, 208 séquences appartenant à 107 organismes représentatifs des principaux taxons ont été choisies afin de construire un arbre phylogénétique. Pour éviter les interférences dues à la présence ou à l'absence de résidus ou de domaines supplémentaires, les peptides signaux et les extensions C-terminales ont été enlevés avant alignement. Les séquences d'acides aminés correspondant au domaine catalytique ont été alignées à l'aide de l'algorithme MUSCLE (Edgar, 2004) en utilisant les paramètres par défaut. Une matrice de distance a été déduite de l'alignement à l'aide des paramètres de substitution Blosom62 (Henikoff and Henikoff, 1992) et utilisée pour construire l'arbre en utilisant une version améliorée de la méthode de *Neighbour Joining* (Gascuel, 1997). La représentation de l'arbre a été réalisée en utilisant Dendroscope (Huson and Scornavacca, 2012), et la visualisation de la séquence d'acides aminés consensus du module catalytique a été réalisée à l'aide de WebLogo (Crooks et al., 2004).

5 Production de protéines recombinantes

5.1 Construction et transformation des vecteurs d'expression

Les séquences protéiques sélectionnées (voir séquences en Annexe), incluant les peptides signaux natifs, ont été utilisées pour générer des séquences nucléotidiques codantes optimisées pour une expression chez *P. pastoris*. La synthèse totale des gènes a été réalisée (Genewiz, South Plainfield, New-Jersey, USA) et ils ont été clonés dans le vecteur d'expression pPICZαA (Invitrogen, Carlsbad, California, USA), de manière à être en phase avec la séquence codant pour l'étiquette poly-histidine C-terminale. Les plasmides ont été amplifiés via une transformation de cellules compétentes DH5α d'*Escherichia coli*, puis purifiés à l'aide d'un kit HiSpeed Plasmid Midi (Qiagen, Courtaboeuf, France) et linéarisés par l'enzyme de restriction *PmeI*. Ils ont ensuite été transformés par électroporation dans des cellules de la souche SuperMan₅ de *Pichia pastoris* (BioGrammatics, Carlsbad, California, USA) ; cette souche exprime des protéines portant des N-glycosylations de motif Man₅GlcNAc au lieu de motifs hyperglycosylés. Les transformants ont été isolés sur un milieu gélosé contenant de la zéocine (100 et 500 mg/L).

5.2 Production des protéines

5.2.1 Milieux de culture

Pour les cultures en fioles, trois milieux sont utilisés : YPD (10 g/L d'extrait de levure, 20 g/L de peptone, 20 g/L de glucose), BMGY (10 g/L d'extrait de levure, 20g/L de peptone, 3,4 g/L de YNB (*Yeast Nitrogen Base*) sans acides aminés ni sulfate d'ammonium, 10g/L de sulfate d'ammonium, 10 g/L de glycérol, dans du tampon phosphate pH6 100mM) et BMGY (10 g/L d'extrait de levure, 20g/L de peptone, 3,4 g/L de YNB sans acides aminés ni sulfate d'ammonium, 10g/L de sulfate d'ammonium, dans du tampon phosphate pH6 100mM). Pour les fermentations en bioréacteur, on a utilisé un milieu BSM (*Basal Salts Medium*) contenant 26,7 mL/L d'acide phosphorique à 85%, 0,93 g/L de sulfate de calcium, 18,2 g/L de sulfate de potassium, 14,9 g/L de sulfate de magnésium heptahydraté, 4,13 g/L d'hydroxyde de potassium, et 40 g/L de glycérol. Des traces de minéraux sont apportés à l'aide de sels PTM₁ (*Pichia trace minerals 1*) ou PTM₄ (*Pichia trace minerals 4*). La composition des sels PTM₁ est la suivante : 6 g/L de sulfate de cuivre, 3 g/L de sulfate de manganèse, 0,2 g/L de molybdate de sodium, 0,02 g/L d'acide borique, 0,5 g/L de chlorure de cobalt, 20 g/L de chlorure de zinc, 65 g/L de sulfate de fer, 0,08 g/L d'iodure de sodium et 1 mL/L d'acide sulfurique. La composition des sels PTM₄ est la suivante : 2 g/L de sulfate de cuivre, 3 g/L de sulfate de manganèse, 0,2 g/L de molybdate de sodium, 0,02 g/L d'acide borique, 0,5 g/L de sulfate de calcium, 0,5 g/L de chlorure de cobalt, 12,5 g/L de sulfate de zinc, 22 g/L de sulfate de fer, 0,08 g/L d'iodure de sodium et 1 mL/L d'acide sulfurique.

5.2.2 Production en plaques Deepwell

Pour tester l'expression de chaque vecteur, plusieurs transformants résistants à la zéocine ont été cultivés en plaques Deepwell de 24 puits, dans 5 mL de milieu de croissance BMGY contenant 0,2 g/L de biotine, à 30°C et 250 rpm pendant environ 16h, jusqu'à atteindre une densité optique à 600 nm comprise entre 2 et 6. Après centrifugation, les cellules ont été reprises dans 1 mL de milieu BMMY et incubées à 20°C et 200 rpm pendant 3 jours ; le milieu a été supplémenté chaque jour par du méthanol à 3% (vol/vol) afin d'induire l'expression des protéines. Les surnageants de culture ont été purifiés par chromatographie d'affinité avec des billes Ni-NTA selon une procédure automatisée décrite par Haon et al. (2015), et les protéines purifiées ont été déposées sur un gel SDS-PAGE sans coloration avec révélation par photo-activation et fluorescence (Bio Rad, France) pour vérifier leur production.

5.2.3 Production en fioles

Les transformants ayant le meilleur taux de sécrétion ont été repiqués sur milieu gélosé YPD, puis ensemencés dans plusieurs fioles non bafflées de 1 L contenant 200 mL de milieu BMGY, 1 mL/L de sels PTM₄ et 0,2 g/L de biotine. Les fioles ont été incubées à 30°C et 250 rpm pendant environ 16h, jusqu'à atteindre une densité optique à 600 nm comprise entre 2 et 6. L'expression des protéines a été induite en incubant les cellules dans des fioles contenant 40 mL de milieu BMMY, 1 mL/L de sels PTM₄ et 0,2 g/L de biotine, à 20°C et 200 rpm pendant 3 jours, durant lesquels le milieu a été supplémenté par du méthanol à 3% (vol/vol) chaque jour. Les surnageants ont ensuite été récupérés par centrifugation (4000 g, 4°C, 10 min) et filtrés sur des membranes de 0,45 µm (Merck-Millipore) pour éliminer les cellules restantes.

5.2.4 Production en bioréacteurs

Les protéines ont également été produites dans des bioréacteurs commandés par le système BioFlo 115 ou le système de bioréacteurs parallèles DASGIP (Eppendorf, Montesson, France), en suivant, avec quelques modifications, les recommandations d'Invitrogen pour la fermentation de *Pichia pastoris*. Les transformants repiqués sur milieu gélosé YPD ont d'abord été cultivés dans 5 mL de milieu liquide YPD à 30°C et 160 rpm pendant 6h, puis 200 µL ont été transférés dans 100 mL de milieu BMGY et incubés à 30°C et 200 rpm pendant 16h. Cette préculture a servi à ensemercer à hauteur de 10% les bioréacteurs contenant du milieu BSM ajusté à pH 5 et 4 mL/L de sels PTM₁. La consigne de pH a été fixée à 5, celle de température à 30°C, et le minimum du taux d'oxygène dissout a été fixé à 20%. La phase de croissance sur glycérol a été maintenue environ 24h, jusqu'à ce que le taux d'oxygène dissout remonte, signifiant que le glycérol était épuisé. Du sorbitol (50 g par litre de milieu BSM initial) et du méthanol (50 g/L de milieu initial) ont été ajoutés, pour une phase de transition d'environ 4h, jusqu'à ce que le sorbitol soit épuisé. La phase de production avec alimentation continue en méthanol a ensuite été démarrée, avec un flux de méthanol de 3,6 mL/h/L de milieu initial pendant 24h, puis de 7,2 mL/h/L pendant environ 64h. Après arrêt de la fermentation, le surnageant de

culture a été récupéré par centrifugation (10 min à 6500 g) et filtration sur membrane de 0.45 µm (Merck-Millipore).

5.3 Purification des protéines

5.3.1 Purification par chromatographie d'affinité

La colonne utilisée est une colonne de nickel chélaté HisTrap FF (GE Healthcare, Bus, France) ; elle a été connectée à un appareil de FPLC et équilibrée avec un tampon contenant du Tris-HCl 50 mM (pH 7,8), du NaCl 50 mM et de l'imidazole 10 mM. Après avoir ajusté leur pH à 7,8, les surnageants ont été chargés sur la colonne à 4°C. La colonne a été lavée avec le tampon d'équilibrage (Tris-HCl 50 mM pH 7,8, NaCl 50 mM, imidazole 10 mM) et les protéines portant une étiquette poly-histidine ont été éluées avec le même tampon contenant 150 mM d'imidazole. Les fractions d'éluion contenant des protéines recombinantes ont été rassemblées, concentrées et dialysées sur des membranes de polyéthersulfone (seuil de coupure de 10kDa, Vivaspin, Sartorius). Un échantillon de chaque protéine a été chargé sur un gel SDS-PAGE Tris-glycine 10% précoulé (Bio Rad) pour vérifier leur pureté. La concentration en protéines a été déterminée par absorption à 280 nm à l'aide d'un spectrophotomètre Nanodrop ND-2000 (Thermo Fisher Scientific) et des masses et coefficients molaires théoriques calculés à partir des séquences protéiques. Afin d'être chargés en cuivre, les LPMO ont été ensuite incubés en présence de sulfate de cuivre CuSO₄ (1:10 équivalent molaire) pendant 16h, puis l'excédent de cuivre a été éliminé par diafiltration à l'aide de membranes de polyéthersulfone avec un seuil de coupure de 5 kDa (Vivaspin, Sartorius) dans du tampon acétate 50 mM (pH 5,2).

5.3.2 Purification par chromatographie d'échange d'anions

Deux colonnes différentes ont été utilisées : une colonne HiPrep DEAE FF 16/10 (GE Healthcare) pour la purification d'AaAA16, et une colonne 6MonoQ HR 16/10 (GE Healthcare) pour les autres protéines. Les surnageants ont été concentrés et diafiltrés dans du tampon Tris-HCl 20mM (pH 7,8), puis chargés sur la colonne préalablement équilibrée à l'aide du même tampon. La colonne a été lavée avec le tampon d'équilibrage, puis les protéines ont été éluées en utilisant un gradient linéaire (de 0 à 50%) de NaCl 1M. Plusieurs fractions d'éluion ont été chargées sur un gel SDS-PAGE Tris-glycine 10% précoulé (Bio Rad) pour vérifier leur pureté. Les fractions d'éluion contenant des protéines d'intérêt ont été rassemblées, concentrées et dialysées sur des membranes de polyéthersulfone (seuil de coupure de 10kDa, Vivaspin, Sartorius). Les concentrations en protéines ont été déterminées par un dosage de Bradford (BioRad) ainsi que par absorption à 280 nm à l'aide d'un spectrophotomètre Nanodrop ND-2000 (Thermo Fisher Scientific) et des masses et coefficients molaires théoriques calculés à partir des séquences protéiques.

5.4 Caractérisation biochimique des protéines produites

5.4.1 Analyse ICP-MS

L'analyse par spectrométrie de masse à plasma à couplage inductif (ICP-MS), visant à vérifier la présence de cuivre dans le site actif des protéines produites et à en mesurer la concentration, a été réalisée par Florence Chaspoul (IMBE, CNRS, Marseille) comme décrit par Couturier et al. (2018). Brièvement, les échantillons ont été minéralisés, puis dilués dans de l'eau ultra-pure, et analysés par un appareil ICAP Q (Thermo Electron, Les Ullis, France). La concentration en cuivre a été déterminée en utilisant le logiciel Plasmalab (Thermo Electron), à une masse $m/z=63$.

5.4.2 Séquençage N-terminal

La séquence des acides aminés N-terminaux des protéines purifiées a été déterminée à la Plateforme Protéomique de l'Institut de Microbiologie de la Méditerranée (IMM, CNRS, Marseille, France ; <https://www.imm.cnrs.fr/proteomique/>) selon la technique de la dégradation d'Edman. Pour chaque échantillon, 500 picomoles de protéines dans une solution d'acide trifluoroacétique à 0,1% dans l'eau ont été déposées sur une membrane en polyfluorure de vinylidène (Prosorb, Thermo Fisher Scientific) traitée avec 10 μ L de méthanol. Trois lavages par 100 μ l d'eau à 0.1 % d'acide trifluoroacétique (TFA) ont été successivement appliqués pour fixer et dessaler les protéines sur ces membranes. Les membranes ont ensuite été séchées et déposées dans la cartouche d'un séquenceur PPSQ 31B (Shimadzu, Noisiel, France). Le programme standard a été appliqué pour 5 à 10 cycles.

5.4.3 Analyse de masse globale

La masse globale des protéines purifiées a été déterminée à la Plateforme Protéomique de l'IMM, par analyse de temps de vol par désorption-ionisation laser assistée par matrice (*Matrix-assisted laser desorption ionization time-of-flight*, MALDI-ToF). Les échantillons ont été préparés par acidification à l'aide de TFA à 12,5 %, avant l'addition d'une solution saturée de matrice acide 2,5 dihydroxybenzoïque dans 50 % d'acétonitrile dans l'eau, 0,1 % de TFA (v/v). L'analyse a été réalisée sur un spectromètre de masse Microflex II (Bruker Daltonics, Brême, Allemagne) dans la gamme de 2 à 100 kDa, dans un mode linéaire et positif, en utilisant les méthodes d'acquisition prédéterminées par le constructeur (Lp12, Lp66). L'étalonnage de masse externe a été effectué à l'aide des calibrants protéiques ProtI, pour la gamme de 5 à 16 kDa, et ProtII, pour la gamme de 20 à 100 kDa (Bruker Daltonics).

5.4.4 Tests Amplex Red

Un test fluorimétrique décrit par Kittl et al. (2012) permet de mesurer l'activité des LPMO grâce à la quantification du peroxyde d'hydrogène produit par les LPMO en présence d'oxygène et d'un réducteur, à l'aide d'une réaction en présence d'Amplex Red (Sigma-Aldrich) et de peroxydase de raifort. Des plaques 96 puits noires (Greiner Bio One, Kremsmünster, Autriche) contenant dans un volume total de 100 μ L du tampon

citrate-phosphate 50 mM (pH6), 50 μ M d'Amplex Red, 7,1 U/mL de peroxydase, 0,01 à 20 μ M de LMPO et 1 mM de réducteur (L-cystéine ou ascorbate) ont été incubées à 30°C pendant 30 minutes. La fluorescence a été suivie à raison d'un point par minute, à une longueur d'onde d'excitation de 560 nm et une longueur d'onde d'émission à 595 nm, à l'aide d'un lecteur de plaque Tecan Infinite M200 (Tecan).

6 Caractérisation de l'activité des LPMO

6.1 Substrats utilisés

Les substrats utilisés pour les tests de caractérisation sont les suivants : Avicel PH-101 (Sigma-Aldrich), PASC (*Phosphoric Acid Swollen Cellulose*, préparée à partir d'Avicel selon la méthode décrite par Wood (1988)), β -1,3/1,4-glucane d'orge, glucomannane de konjac, cellohexaose, xylohexaose, oligosaccharides de xyloglucane de structures XXXG, XXLG et XLLG selon la nomenclature de Fry et al. (1993) (Megazyme, Bray, Irlande), et NFC (nanofibrilles de cellulose, obtenues à partir de cellulose par traitement aux endoglucanases et microfluidisation, fournies par le Centre Technique du Papier, Grenoble, France).

6.2 Tests de dégradation de substrats

Les tests de dégradation de substrats ont été réalisés en triplicats, dans un volume de 400 μ L d'eau contenant 0,1% à 1% (m/v) de substrats solides (PASC, Avicel, β -1,3/1,4-glucane d'orge, glucomannane de konjac) ou 1 mM de substrats solubles (cellohexaose, xylohexaose, oligosaccharides de xyloglucane). Les LPMO AA16 (2 μ M à 20 μ M) ont été ajoutées et incubées avec le substrat pendant au moins 20 min à 45°C avant ajout d'un donneur d'électrons, choisi parmi la L-cystéine, l'ascorbate, l'acide *p*-coumarique, l'acide sinapique, l'acide vanillique, l'acide cafféique, l'acide syringique, l'acide gallique, l'acide 3-hydroxyanthranilique, l'acide tannique, le pyrogallol et la ménadione (1 mM), ou encore la cellobiose déshydrogénase de *Podospora anserina* PaCDHB (Bennati-Granier et al., 2015) (2 nM). Les échantillons ont été incubés 15 min à 72h dans un thermomixer (Eppendorf) à 45°C et 850 rpm. Les réactions ont été arrêtées par ajout de 40 μ L de soude à 1 M, et les échantillons ont été filtrés pour séparer la fraction soluble de la fraction insoluble.

6.3 Tests de synergies sur cellulose

Des tests de synergie entre les AA16 et la CBHI de *T. reesei* ont été réalisés en triplicats, dans un volume total de 800 μ L, en présence de 0,1 ou 1% (m/v) de PASC ou de NFC dans l'eau. Les substrats ont d'abord été incubés en présence de LPMO (10 mg/g de substrat) et de L-cystéine (1 mM), à 45°C et 850 rpm, pendant 16 à 24h. La réaction a été arrêtée par ajout de soude (100 mM) et les échantillons ont été centrifugés à 15000xg pendant 10 min. Le surnageant a été retiré, et la fraction insoluble restante du substrat a été lavée deux fois dans de l'eau. L'hydrolyse par la CBHI de *T. reesei* (1 mg/g ; fournie par IFP Energies Nouvelles) a été réalisée dans un volume de 800 μ L contenant du tampon acétate 50 mM pH5,2, pendant 30 min à 45°C et 850 rpm.

Les réactions ont été arrêtées par ajout de 80 μL soude à 1 M, et les échantillons ont été filtrés pour séparer la fraction soluble de la fraction insoluble.

6.4 Analyse des produits solubles de dégradation

6.4.1 HPAEC-PAD

Les mono- et oligosaccharides issus de la dégradation des substrats cités plus haut ont été détectés par chromatographie haute-performance par échange d'anions avec détection ampérométrique (*High Performance Anion Exchange Chromatography coupled with Pulsed Amperometric Detection*, HPAEC-PAD, Dionex, Thermo Fisher Scientific) selon la méthode décrite par Westereng et al. (2013). Les éluents utilisés contiennent 0,1 M de soude (éluent A) et 1 M d'acétate de sodium dans de la soude à 0,1 M (éluent B). L'élution a été réalisée à un débit constant de 0,25 mL/min à 30°C, en utilisant un gradient linéaire de 0 à 10% d'éluent B en 10 min, puis de 10 à 30% d'éluent B en 25 min, et enfin un gradient exponentielle de 30 à 100% d'éluent B en 5 min. Les conditions initiales (100% d'éluent A) ont ensuite été restaurées en 1 min et maintenues pendant 9 min pour reconditionner la colonne. Les produits ont été identifiés en utilisant comme standard des cello-oligosaccharides natifs de degré de polymérisation (DP) compris entre 1 et 6 (Megazyme), ainsi que des cello-oligosaccharides oxydés en C1, préparés par incubation avec la CDH de *Podospora anserina* (PaCDHB, (Bennati-Granier et al., 2015)).

6.4.2 Spectrométrie de masse

L'analyse par spectrométrie de masse tandem (MS/MS) a été réalisée par Mathieu Fanuel (BIA, INRA, Nantes) sur un spectromètre de masse haute définition Synapt G2Si (Waters Corp., Manchester, Royaume-Uni) équipé d'une source ionisation électrospray (ESI). Après sélection des précurseurs ioniques et avant l'étape de fragmentation, la mobilité ionique a été utilisée pour réduire les interférences dues aux impuretés de l'échantillon, et effectuée dans une cellule de mobilité ionique à ondes progressives (*Traveling-Wave Ion Mobility*, TWIM). Le débit d'hélium a été maintenu à 180 mL/min et le débit d'azote a été ajusté à 90 mL/min dans la cellule de mobilité. La hauteur de l'onde mobile a été réglée à 40 V et la vitesse de l'onde a été réglée à 300 $\text{m}\cdot\text{s}^{-1}$. Après avoir traversé la cellule de mobilité, les espèces oxydées ont été fragmentées par dissociation induite par collision dans la cellule de transfert de l'instrument (MS/MS). Dans ces expériences, les échantillons ont été dilués dix fois dans du MeOH/H₂O (1:1, v/v) et infusés à un débit de 5 $\mu\text{L}/\text{min}$. Les acquisitions ont été réalisées en polarité négative, ainsi qu'en mode "sensibilité".

6.5 Spectroscopie du site actif

La géométrie du cuivre (II) dans le centre actif des AA16 a été étudiée par résonance paramagnétique électronique (RPE). Cette technique de spectroscopie permet de détecter les électrons non appariés au sein d'une espèce moléculaire, suite à l'application d'un champ magnétique. La recherche de résonance est généralement réalisée à une fréquence fixe d'irradiation micro-ondes, en faisant varier le champ magnétique externe appliqué. La mesure de l'absorbance en fonction du champ magnétique permet d'enregistrer un spectre, qui donne des informations sur la structure électronique du composé étudié et les interactions existant entre l'électron non apparié et les noyaux qui l'entourent. L'analyse du spectre permet notamment de déterminer les tenseurs g (facteur de Landé, caractéristique de la molécule paramagnétique étudiée) et A (tenseur hyperfin d'anisotropie, qui caractérise le couplage entre spin électronique et spins nucléaires). Les spectres RPE des AA16 (85 à 165 μM) en solution dans l'eau ou dans un tampon acétate de sodium 50 mM pH5,2 ont été enregistrés à une fréquence de 9388,2 MHz (bande X) par un spectromètre Bruker EMX (Bruker, Wissembourg, France) à des températures de 100 et 110K dans l'azote liquide. Des spectres ont également été enregistrés après incubation des échantillons avec de l'acide ascorbique en excès pendant 10 min à 24h. Des simulations spectrales ont été réalisées par Nadia Touati (IRCP, CNRS, Paris, France) en utilisant le logiciel EasySpin 5.0.3 (Stoll and Schweiger, 2006) intégré au logiciel MATLAB R2016a, afin de déterminer les paramètres g_x , g_y , g_z et $|A_x|$, $|A_y|$, $|A_z|$.

6.6 Modélisation des structures 3D par homologie

La modélisation des structures tridimensionnelles des quatre protéines produites a été obtenue à l'aide de leurs séquences, en utilisant les outils en ligne I-TASSER (Yang and Zhang, 2015, <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) et Phyre2 (Kelley et al., 2015, <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi>). Les modèles utilisés sont les structures de protéines issues des familles AA10 et AA15, portant les numéros suivants dans la base de données PDB : 5FTZ (*S*/AA10), 4OY7 (*Sc*AA10C), 5WSZ (*Bt*AA10A), 5MSZ (*Td*AA15). La visualisation des structures a été réalisée avec le logiciel PyMol.

Chapitre III : Identification d'une nouvelle famille de LPMO dans des secrétomes fongiques

1 Résumé

La première étape du projet a consisté en une étude des cocktails enzymatiques secrétés par les cinq souches sélectionnées d'*Aspergillus*, cultivées sur trois milieux distincts. Les 15 secrétomes ainsi préparés ont été utilisés pour compléter un cocktail cellulolytique de *T. reesei* lors de tests de saccharification de trois biomasses prétraitées (paille de blé, Miscanthus et peuplier). L'obtention de nombreux réplicats grâce à une méthode miniaturisée robotisée a permis d'identifier les secrétomes qui améliorent de manière statistiquement significative les rendements de saccharification de chacune des biomasses. Les performances des secrétomes diffèrent selon les biomasses, et dépendent de la souche et du substrat utilisés pour les produire. L'analyse protéomique des secrétomes a également montré que les 15 cocktails enzymatiques ont des compositions très différentes les uns des autres, et qu'il existe une importante diversité au sein de la plupart des secrétomes, notamment en termes de CAZymes ; on peut noter aussi la présence de nombreuses protéines de fonction inconnue, qui laissent envisager la découverte de potentielles nouvelles activités.

Le contenu des secrétomes ayant montré de bonnes performances de saccharification sur Miscanthus a été analysé plus en détail, ce qui a permis d'identifier une protéine de fonction inconnue, sécrétée par la souche BRFM 1490, qui possède une séquence proche de certaines LPMO tout en ne faisant partie d'aucune famille CAZy déjà existante. Il s'est avéré que cette protéine faisait partie d'une famille putative de LPMO, identifiée précédemment par analyse génomique par l'équipe CAZy et par Voshol et al. (2017), mais n'ayant pas encore de caractérisation biochimique. Une analyse bioinformatique a été réalisée, révélant que cette famille est présente chez au moins 580 microorganismes (champignons et oomycètes) ayant des modes de vie impliquant des interactions avec la biomasse végétale. L'alignement des séquences révèle de nombreux résidus conservés, dont deux histidines pouvant former l'« *histidine brace* » caractéristique des LPMO.

Une séquence de cette nouvelle famille, issue d'*Aspergillus aculeatus* a été exprimée chez *Pichia pastoris*, et après une mise au point des conditions de production et de purification, la protéine a été obtenue sous forme fonctionnelle, avec une *histidine brace* intacte contenant du cuivre. L'enzyme a été testée sur plusieurs substrats, mais n'a montré d'activité que sur la cellulose et les cello-oligosaccharides. L'utilisation de chromatographie HPAEC-PAD et de spectrométrie de masse a permis de détecter des produits de dégradation solubles, sous forme de sucres natifs et de sucres oxydés à leur extrémité C1. Des tests d'activité sur cellulose amorphe en association avec une cellobiohydrolase de *T. reesei* ont révélé que l'enzyme pouvait améliorer les performances de cette cellulase, ce qui pourrait avoir des applications biotechnologiques intéressantes. La nouvelle famille ainsi caractérisée a été nommée AA16 dans la base de données CAZy. Ces résultats ont été publiés en mars 2019 dans *Biotechnology for Biofuels* et sont présentés ici en anglais dans la forme de l'article accepté ; seule la numérotation des figures a été modifiée.

Article 1: AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes

Camille Filiatrault^{1,2}, David Navarro¹, Mireille Haon¹, Sacha Grisel¹, Isabelle Herpoël¹,
Didier Chevret³, Mathieu Fanuel⁴, Bernard Henrissat⁵, Senta Heiss², Antoine Margeot² and
Jean¹ Guy Berrin¹

¹ Biodiversité et Biotechnologie Fongiques (INRA, Aix Marseille Université), Marseille, France.

² IFP Energies Nouvelles, 1 et 4 avenue de Bois¹Préau, 92852 Rueil¹Malmaison, France.

³ Plateforme d'Analyse Protéomique de Paris Sud¹Ouest (INRA, Agro-ParisTech), Jouy¹En¹Josas, France

⁴ Biopolymères Interactions Assemblages (INRA), Nantes, France

⁵ Architecture et Fonction des Macromolécules Biologiques (CNRS, Aix Marseille Université, INRA), Marseille, France

2 Background

Lignocellulosic biomass is a renewable and abundant feedstock, considered as a promising alternative to non-sustainable fossil resources for the production of biofuels, biomaterials and bio-based chemicals (Cherubini, 2010; de Bhowmick et al., 2018). Its transformation requires an extensive deconstruction of the plant cell wall polymers, namely the cellulose and hemicellulose polysaccharides, that are intermeshed together with lignin to form a highly resistant structure. Efficient hydrolytic enzyme cocktails are therefore required but are still at the present time a major bottleneck for cost-effective industrial processes.

The degradation of plant polysaccharides into simple sugars can be achieved using enzymes secreted by biomass-degrading organisms such as bacteria and filamentous fungi. These carbohydrate-active enzymes (or CAZymes) are classified in the CAZy database (<http://www.cazy.org/>) into several families based on their amino acid sequence similarities (Lombard et al., 2014). Cellulases and hemicellulases, which belong to the glycoside hydrolases (GH), have been studied for more than 60 years for their ability to depolymerize cellulose and hemicelluloses. More recently, novel enzymes called LPMOs (Lytic Polysaccharide Monooxygenases) were discovered (Vaaje-Kolstad et al., 2010) and raised a high interest due to their ability to boost the hydrolysis of lignocellulosic biomass (Harris et al., 2010). LPMOs are able to cleave glycosidic bonds using an oxidative mechanism, in the presence of oxygen atoms (coming from dioxygen or hydrogen peroxide) and an extracellular electron donor (Bissaro et al., 2017; Bissaro et al., 2018). Electrons can be provided by small organic compounds such as lignin fragments and plant- or fungus-derived phenols, by enzymatic systems such as cellobiose dehydrogenase (CDH) or by photocatalytic systems (Cannella et al., 2016). All LPMOs share some common features, such as a copper-containing active site, in which the metal is coordinated by three nitrogen atoms from two histidine side chains and the N-terminal amine group that form the so-called "histidine brace" (Ciano et al., 2018; Quinlan et al., 2011).

LPMOs are currently grouped into six CAZy auxiliary activities (AA) families (AA9-AA11, AA13-AA15), based on bioinformatic analyses of their amino-acid sequences similarities. These families are found in several taxonomic groups, and have been shown to be active on numerous substrates (cellulose, hemicelluloses, chitin and starch). The fungal AA9 family includes enzymes active on cellulose and cello-oligosaccharides, but also polysaccharides containing β -1,4 linked glucose units such as xyloglucans, glucomannans and β -glucans (Bennati-Granier et al., 2015; Fanuel et al., 2017; Isaksen et al., 2014). LPMOs from the AA10 family, which are found mainly in bacteria and some viruses, are active on both chitin and cellulose (Vaaje-Kolstad et al., 2012). AA11 and AA13 families are found exclusively in fungi; only a few of their members, which are respectively active on chitin and starch components, have been characterized so far (Hemsworth et al., 2014; Lo Leggio et al., 2015; Vu and Marletta, 2016). The fungal AA14 family was discovered recently, and two of its members showed activity on recalcitrant xylan coating cellulose fibers (Couturier et al., 2018). Finally, the discovery of the AA15 family revealed the existence of LPMOs of animal origin (invertebrates), active on both cellulose and chitin (Sabbadin et al., 2018). These enzymes have only recently been described, and the discovery of more families and new enzyme specificities can be expected.

Cellulases, hemicellulases and LPMOs are key components of industrial cocktails dedicated to biomass degradation (Johansen, 2016). The vast majority of these cocktails are based on enzymes secreted by *Trichoderma reesei*, a fungal species that is known for its efficient cellulolytic enzymes secretion, and benefits from decades of strain improvement (Peterson and Nevalainen, 2012; Seidl and Seiboth, 2010). However, the sequencing of its genome revealed a surprisingly small number of genes involved in polysaccharides degradation (Häkkinen et al., 2012; Martinez et al., 2008), and several missing accessory activities (Bischof et al., 2016). With only three AA9 LPMOs encoded in its genome, *T. reesei* has one of the smallest set of LPMOs among fungal saprotrophs, which are usually rich in AA9 with up to 40-50 genes (Lenfant et al., 2017). To compensate this lack of diversity, which may be a bottleneck for *T. reesei* cocktails improvement, one strategy is to upgrade *T. reesei* cocktails with enzymes from other biomass-degrading filamentous fungi isolated from biodiversity. Such enzymes can be found by mining the sets of secreted enzymes (*i.e.* secretomes) produced by fungal strains on various carbon sources, thanks to the development of protein identification tools such as LC-MS/MS and the increasing number of available annotated fungal genomes (Bouws et al., 2008; Di Cologna et al., 2017). Since a large panel of CAZymes and LPMOs are differentially secreted by fungal saprotrophs (Berrin et al., 2017; Nekiunaite et al., 2016), comparative secretomics is a promising approach to discover novel enzymes playing a role in plant biomass deconstruction. Herein, we report the identification of a new LPMO family among *Aspergillus* spp. secretomes improving the conversion of plant biomass. One member of this new LPMO family, termed AA16, was characterized in terms of substrate specificity, regioselectivity of oxidation and contribution to cellulose degradation.

3 Results

3.1 Exploration of fungal secretomes to improve biomass saccharification

During previous screenings of fungal strains for their biomass-degradation ability, several strains of the *Aspergillus* genus appeared promising (Berrin et al., 2015). In the present study, four strains of *Aspergillus japonicus* and one strain of *Aspergillus niger* were grown in the presence of three inducers (maize bran (MB), sugar beet pulp (SBP) and cellulose (Avicel)) to favor the secretion of diverse lignocellulose-active enzymes. After 7 days of growth, the 15 different secretomes (Table 5) were tested for their ability to improve a reference *T. reesei* cocktail obtained from strain CL847 (Durand et al., 1988; Herpoël-Gimbert et al., 2008) for the saccharification of three dilute-acid steam-exploded biomasses: wheat straw, Miscanthus and poplar.

Table 5 : Protein concentration (determined by BCA assay) of the secretomes of five *Aspergillus* spp. strains grown on three inducers, after seven days of culture.

Code	Species	Strain number	Inducer	Concentration (mg/mL)
405_Avi	<i>Aspergillus japonicus</i>	CIRM-BRFM 405	Avicel	0.55
405_MB	<i>Aspergillus japonicus</i>	CIRM-BRFM 405	Maize bran	12.56
405_SBP	<i>Aspergillus japonicus</i>	CIRM-BRFM 405	Sugar beet pulp	4.96
430_Avi	<i>Aspergillus niger</i>	CIRM-BRFM 430	Avicel	0.27
430_MB	<i>Aspergillus niger</i>	CIRM-BRFM 430	Maize bran	17.63
430_SBP	<i>Aspergillus niger</i>	CIRM-BRFM 430	Sugar beet pulp	7.98
1487_Avi	<i>Aspergillus japonicus</i>	CIRM-BRFM 1487	Avicel	0.25
1487_MB	<i>Aspergillus japonicus</i>	CIRM-BRFM 1487	Maize bran	17.24
1487_SBP	<i>Aspergillus japonicus</i>	CIRM-BRFM 1487	Sugar beet pulp	8.43
1489_Avi	<i>Aspergillus japonicus</i>	CIRM-BRFM 1489	Avicel	0.29
1489_MB	<i>Aspergillus japonicus</i>	CIRM-BRFM 1489	Maize bran	13.47
1489_SBP	<i>Aspergillus japonicus</i>	CIRM-BRFM 1489	Sugar beet pulp	12.37
1490_Avi	<i>Aspergillus japonicus</i>	CIRM-BRFM 1490	Avicel	0.34
1490_MB	<i>Aspergillus japonicus</i>	CIRM-BRFM 1490	Maize bran	12.31
1490_SBP	<i>Aspergillus japonicus</i>	CIRM-BRFM 1490	Sugar beet pulp	8.50

After 48h of biomass saccharification, three secretomes (produced by *A. japonicus* BRFM 405 on MB, *A. japonicus* BRFM 1487 on SBP and *A. niger* BRFM 430 on Avicel) were able to significantly improve the cellulose conversion yield of wheat straw (Figure 27A), which is noteworthy given that the yields using the

reference cocktail alone were already high, reaching 60% glucose release from cellulose after only 24h, and over 90% at the plateau (Figure 26). The *T. reesei* cocktail was less efficient on poplar (60% of maximum saccharification yield, see Figure 26), and this woody biomass proved to be also recalcitrant to the *Aspergillus* secretomes: only one of them (produced by *A. japonicus* BRFM 405 on MB) was able to significantly improve the yield after 96h of reaction (Figure 27B).

Concerning Miscanthus, a total of 10 secretomes among the 15 tested were able to significantly improve the cellulose conversion yield (Figure 27C), after 24h and/or 96h of reaction. Interestingly, the secretomes of *A. japonicus* BRFM 405 and *A. niger* BRFM 430 produced on all of the 3 inducers were able to boost Miscanthus saccharification, while none of the secretomes of the *A. japonicus* BRFM 1489 strain had an effect on yield. As for the two remaining strains (*A. japonicus* BRFM 1487 and 1490), their secretomes produced on lignocellulosic substrates (MB and SBP) improved the glucose yield by 9 to 16%, while their secretomes produced on Avicel did not lead to any significant improvement. This demonstrates that not only the strain, but also the inducer has an effect on the biomass degradation ability of the produced enzyme cocktails. Concerning the time effect, it can be noticed that all secretomes produced on MB and SBP have a lower effect after 96h of reaction than after 24h, which can be explained by the fact that the higher the yield, the more it becomes difficult to improve it when approaching the saccharification plateau. The secretomes of strains BRFM 405 and 430 produced on Avicel show the opposite effect, with a greater yield improvement at 96h, which could mean that the enzymes responsible for the boost act more slowly than those of other secretomes. Overall, these results show that several *Aspergillus* secretomes are able to improve biomass saccharification, and that the enzymes responsible for the boost are probably different from one secretome to another.

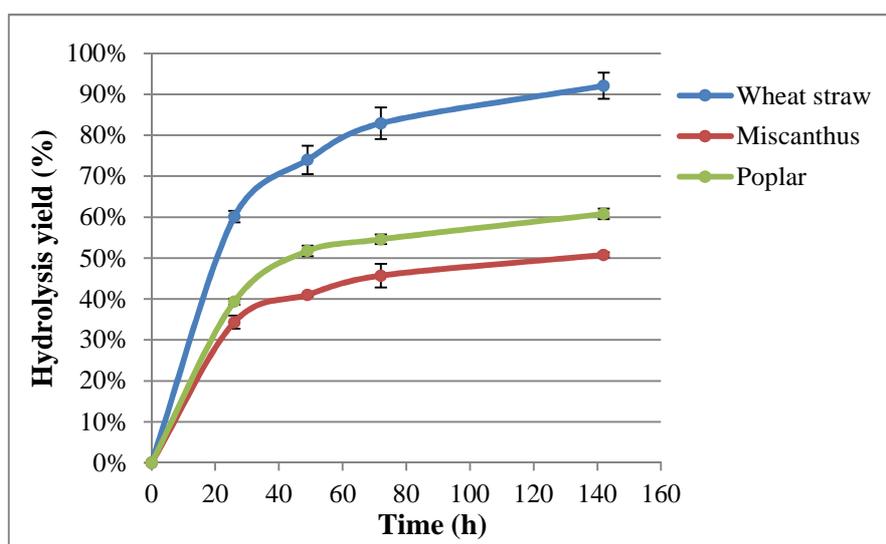


Figure 26 : Hydrolysis of three pretreated lignocellulosic substrates by a cellulolytic cocktail of *Trichoderma reesei* strain CL847 produced on lactose. Error bars represent standard deviations calculated on 3 replicates

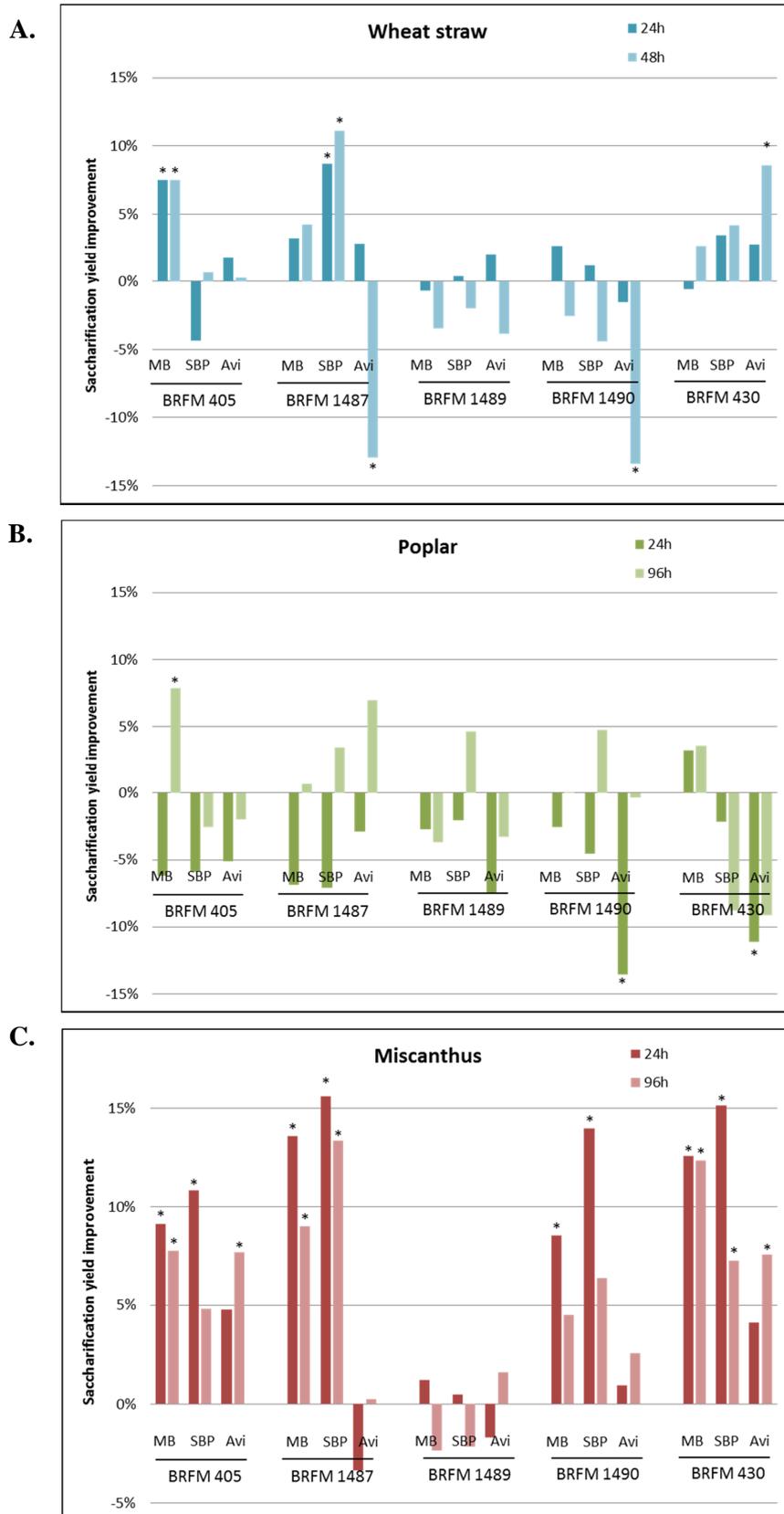


Figure 27 : Biomass saccharification performances of 15 *Aspergillus* spp. secretomes

Effect on the saccharification of pretreated biomass (5% w/v) of the addition to a *T. reesei* cellulolytic cocktail of *Aspergillus* secretomes produced by strains BRFM 405, 430, 1487, 1489 and 1490 on maize bran (MB), sugar beet pulp (SBP) and Avicel (Avi). A. Wheat straw; B. Poplar; C. Miscanthus. The bars show the improvement of cellulose conversion yields in the presence of secretomes compared to cellulose conversion yields obtained with *T. reesei* cocktail alone, after 24h and 96h of reaction. Means were calculated on more than 10 replicates; the * indicate a Student test *p*-value lower than 0.05.

3.2 Comparative proteomic analysis of fungal secretomes

To understand the differences in terms of enzyme composition between these fungal secretomes, proteomic analyses were performed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), and the detected peptides were assigned using the public genomes of *A. niger* and *A. aculeatus*. This latter species is very close to *A. japonicus*, and the two species are identical in some classifications (Abarca et al., 2004), which explains our decision to use the *A. aculeatus* genome to identify and annotate *A. japonicus* proteins.

The number of different proteins present in each secretome is highly variable, ranging from 33 to over 200. Most of the produced secretomes have a larger diversity of enzymes than the CL847 reference cocktail, which contains approximately 30 proteins. However, it must be noted that this *T. reesei* cocktail was produced under different conditions than the *Aspergillus* secretomes. As expected, the *Aspergillus* secretomes are rich in CAZymes, accounting for 25% to 67% of the total number of proteins, but they also contain enzymes from other classes (oxidases, esterases, proteases, nucleases, etc.) and proteins of unknown function (Figure 28).

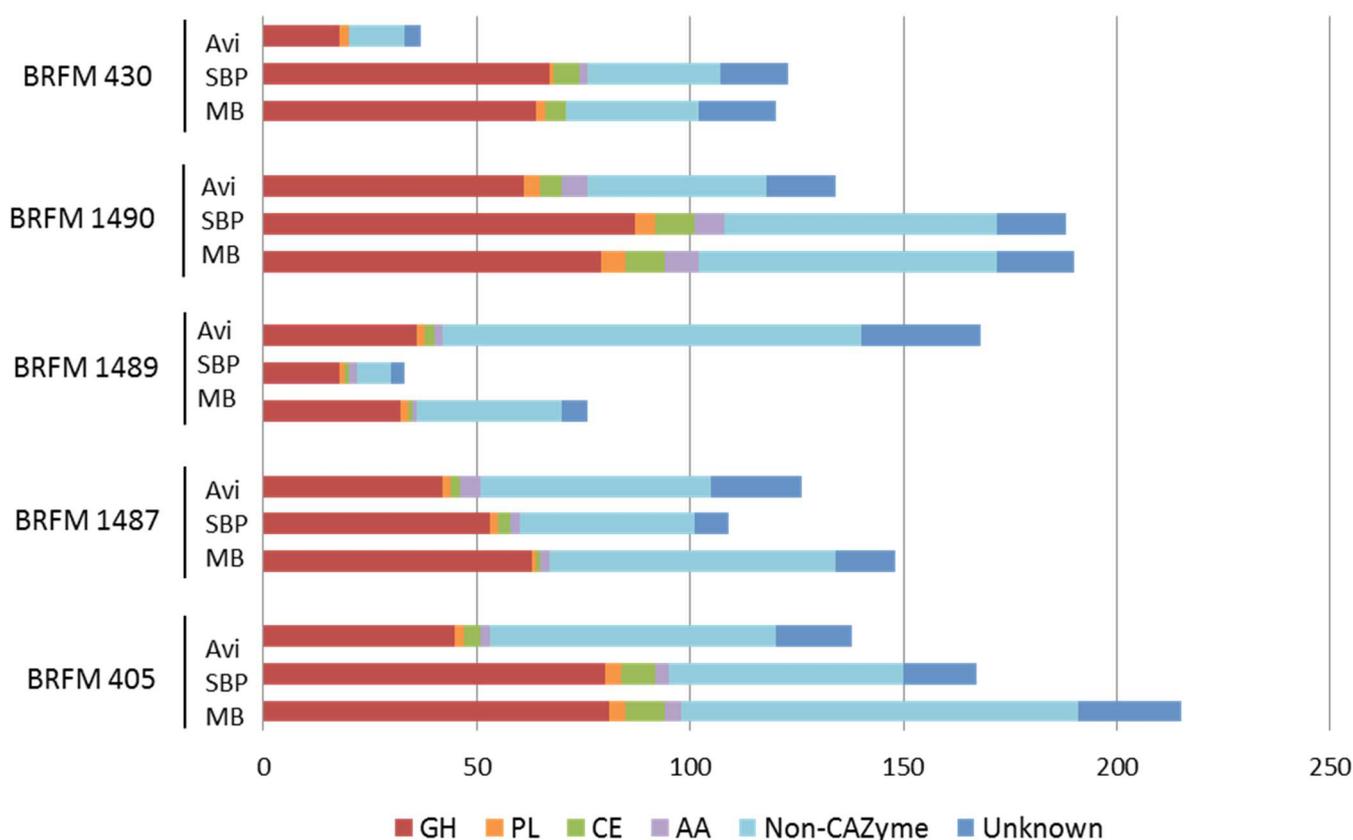


Figure 28 : Protein content of the *Aspergillus* spp. secretomes

Composition of the 15 secretomes in number of identified CAZymes (GH: Glycoside Hydrolases; PL: Polysaccharide Lyases; CE: Carbohydrate Esterases; AA: Auxiliary Activities), non-CAZyme proteins and proteins of unknown function.

The composition of each secretome that showed an improvement of saccharification was further examined, with special attention for proteins of putative or unknown function. Among other proteins, our attention was dragged to a protein of unknown function (ortholog of *A. aculeatus* XP_020060743.1), which was found only in MB and SBP secretomes of *A. japonicus* BRFM 1490 that increased *Miscanthus* saccharification (Figure 27C). Interestingly, orthologs of this protein are not found in the *T. reesei* genome. As revealed by sequence analyses, this protein displays some common structural features with known LPMOs. For instance, its N-terminal residue after signal peptide cleavage is a histidine residue. A deeper analysis of the mass spectrometry data obtained from *A. japonicus* BRFM 1490 secretome revealed that the N-terminal peptide displays a methylated histidine (mass +14.0157 Da), a feature commonly observed in fungal LPMOs.

3.3 Bioinformatic analysis of a new LPMO family

The protein of unknown function (ortholog of *A. aculeatus* XP_020060743.1) identified in the fungal secretomes is part of a putative LPMO family that has been reported in a genome mining approach using a HMM model developed by Voshol et al. (Voshol et al., 2017). Without any experimental demonstration of activity, this family could not be added to the CAZy database, which contains only biochemistry-based families. In absence of biochemical characterization, this family was temporarily called X273.

An analysis of non-redundant protein database (in August 2018) revealed a total of 1,065 sequences containing a X273 module, belonging to 580 microorganisms presenting a variety of life styles oriented toward interactions with plant biomass: a majority of the sequences belong to saprophytes and phytopathogens, but plant endophytes and symbionts are also represented. These organisms are mainly fungal (with about 80% of Ascomycetes, 18% of Basidiomycetes and a few Chytridiomycetes), but X273 modules are also found in some Oomycetes, which are fungus-like protists, and more specifically in known phytopathogens of the genera *Phytophthora* and *Pythium*. The Oomycetes present a slight family expansion, with an average number of 5.7 X273 genes per species compared to 1.8 for the Fungi. The X273 module is preceded in all these sequences by a signal peptide and is often followed by a C-terminal extension of varying length. In a few cases, this extension is composed of a linker and a CBM1 module or a glycosylphosphatidylinositol (GPI) anchor (Figure 29A).

The sequences of 208 X273 modules (edited to remove the CBM1, C-terminal extension or GPI anchor) from 107 organisms representative of major taxons were aligned and used to build a phylogenetic tree (Figure 29B), in which five major clades can be distinguished. Clades 1 and 2 group together enzymes from Ascomycetes and Basidiomycetes that have likely been kept from a common ancestor of these two phyla. Conversely, clade 3 gathers only Ascomycetes sequences (except two sequences from *Cryptococcus* sp. that could have been acquired by horizontal gene transfer (Butler et al., 2006)) and clade 4 consists of Basidiomycetes sequences

only, indicating that these sequences evolved after the split between these two phyla. Finally, clade 5 displays sequences from Oomycetes and one sequence from the basidiomycete *Cronartium quercuum*, which could mean that these sequences were kept from a common ancestor of Oomycetes and Fungi, or could also be the result of a horizontal gene transfer. Interestingly, the sequences belonging to proteins containing a CBM1 or a GPI anchor are not spread across the tree but are gathered respectively in clades 1 and 4, suggesting that the X273 modules of these groups may have evolved for distinct functions.

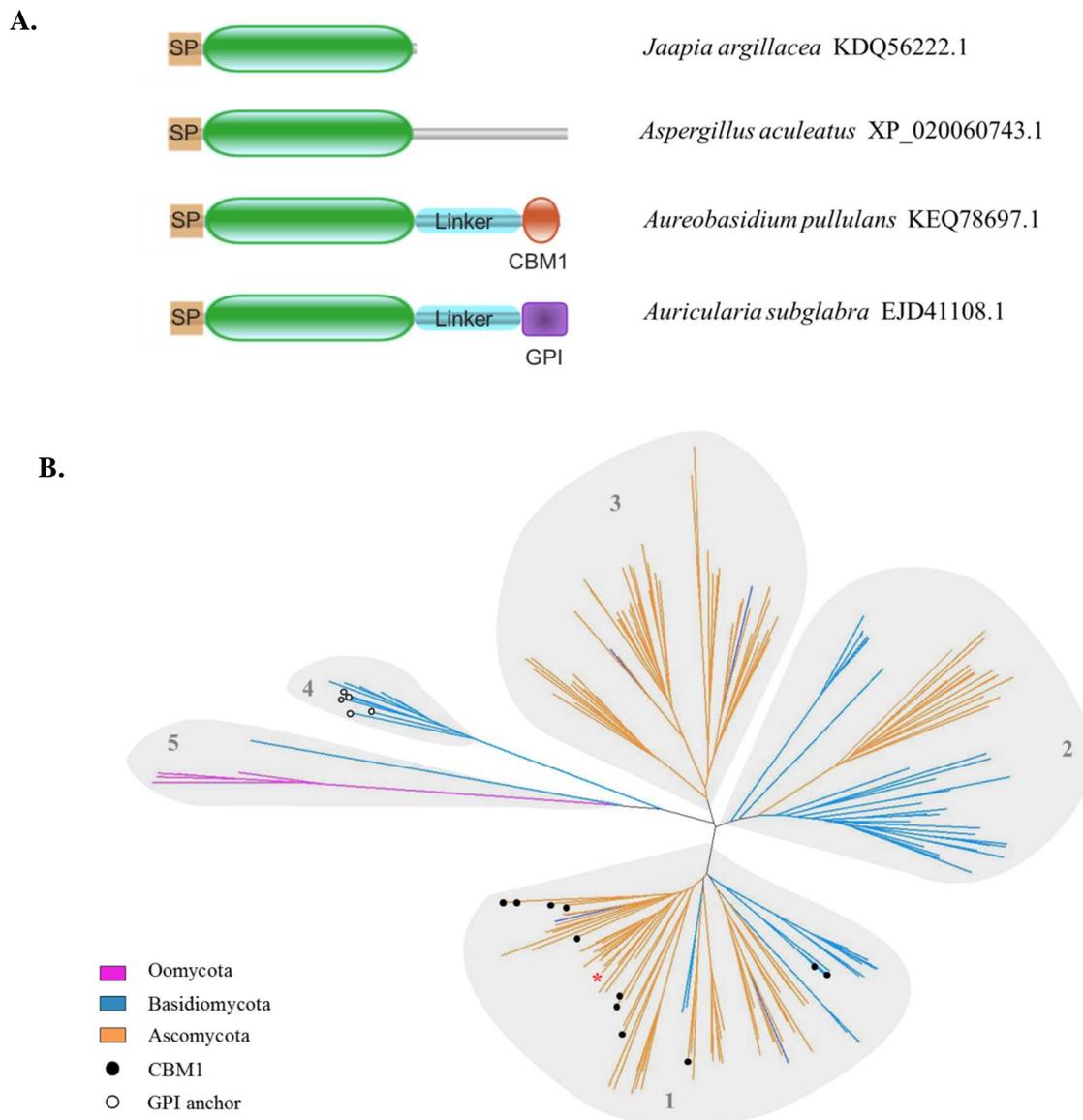


Figure 29 : Phylogeny and modularity of the new LPMO family

A. Graphical representation of the different types of modularity existing among the family, citing examples of proteins from various species with their Genbank ID. The catalytic module, shown in green, is preceded by a signal peptide (SP), and can be followed by a C-terminal extension, a CBM1 module or a GPI anchor.

B. Radial phylogram of 208 selected sequences, based on the alignment of their catalytic module alone. Edges are colored according to the taxonomic division of the corresponding organisms. Full circles indicate CBM1-containing sequences and empty circles indicate GPI anchor-containing sequences. The sequence from *Aspergillus aculeatus* is indicated by a red star.

Sequence alignment of the modules reveals several well-preserved regions within the family (Figure 30). All sequences include an N-terminal histidine, as well as a second strictly conserved histidine, which is a characteristic of the copper binding site (Histidine brace) of all LPMOs characterized so far. There are about twenty other highly conserved residues, including a conserved Q-T/N-Y motif that fits the N/Q/E-X-F/Y motif identified by (Vu et al., 2014b) *et al.* in families AA9, AA10, AA11 and AA13, containing the tyrosine or phenylalanine residue taking part in the coordination of copper in the active site.

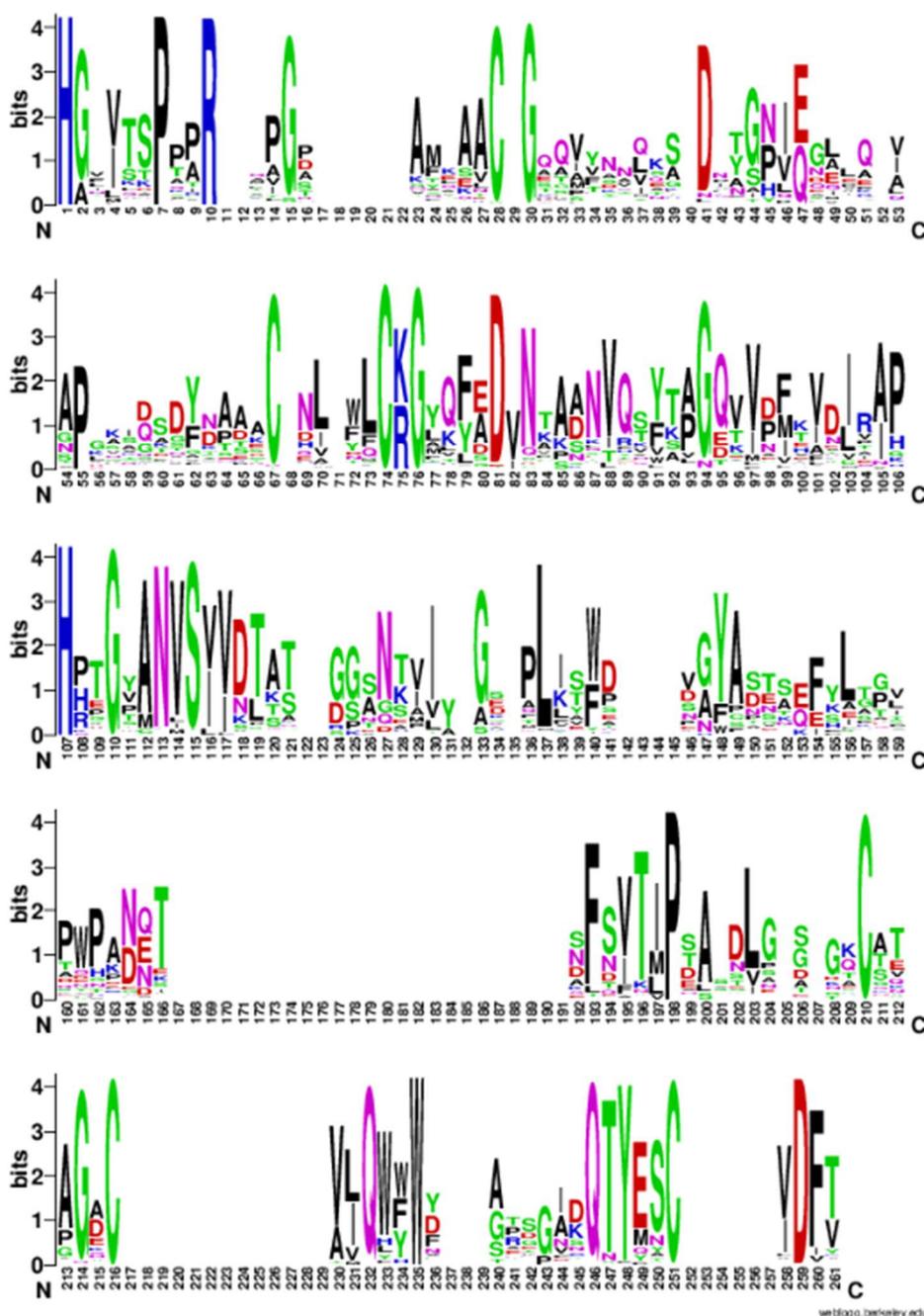


Figure 30 : Graphical representation of AA16 module consensus amino acids, based on the alignment of 213 sequences, generated using the WebLogo application (Crooks et al., 2004).

The strictly conserved histidine residues are shown in positions 1 and 107.

3.4 Heterologous expression and purification

The X273 protein of *A. aculeatus* (Genbank ID XP_020060743.1) was produced in *Pichia pastoris* without its C-terminal extension. The recombinant protein bearing a C-terminal polyhistidine-tag was first produced in flasks in the presence of trace metals including copper and purified from the culture supernatant by immobilized metal ion affinity chromatography (IMAC), following the same protocol used for AA9 LPMOs (Bennati-Granier et al., 2015). The produced protein had a size of ~38 kDa, higher than the expected 21 kDa, attributable to N- and/or O-glycosylation. N-terminal sequencing revealed that in approximately 30% of the sample, the protein was deprived of the N-terminal residue, an heterogeneity that is frequently encountered in the production of recombinant proteins and could be due to peptide bond cleavage by locally produced reactive oxygen species (Kim et al., 1985; McMahon et al., 2001). The rest of the sample did not contain the expected N-terminal histidine, which was replaced either by an aspartate residue or by other non-identified modifications. Recent reports have shown that in LPMOs the residues surrounding the copper are particularly prone to oxidation (Bissaro et al., 2017; Loose et al., 2018). The N-terminal histidine of *A. aculeatus* X273 could have been modified into aspartate, 2-oxo-histidine or other intermediate products due to metal-catalyzed oxidation in the presence of copper as described in Uchida and Kawakishi (1989), Stadtman (1993), or Urresti et al. (2018), most probably leading to an inactive enzyme.

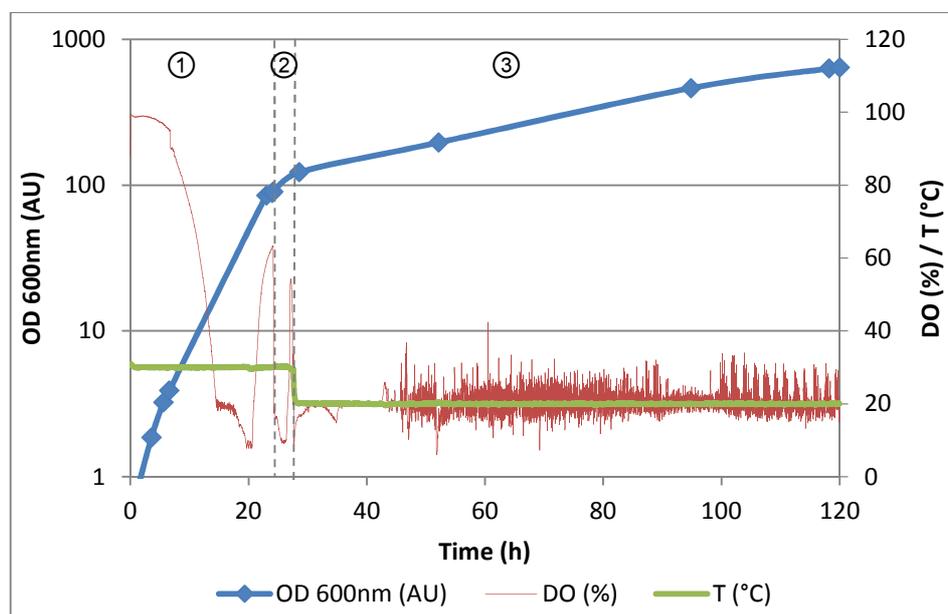


Figure 31 : Time profiles of dissolved oxygen (DO), temperature and optical density (OD) at 600 nm during the production of AaAA16 by *Pichia pastoris* in a 1.3-L bioreactor. 1: Glycerol batch phase; 2: Sorbitol and methanol transition phase; 3: Methanol fed-batch phase.

Alternatively, the same recombinant X273 protein was produced in a bioreactor in the presence of trace metals including copper. A glycerol batch phase was followed by an induction fed-batch phase using methanol

(Figure 31). An aliquot of the culture supernatant was purified by IMAC, to reach a final purified protein production yield of 0.5 g per liter of culture. N-terminal sequencing showed the expected sequence with an intact N-terminal histidine in over 90% of the produced protein, while the cleaved sequence represented only less than 10%.

3.5 Substrate specificity and regioselectivity of cleavage

Preliminary enzyme activity assays were performed using the *P. pastoris* bioreactor supernatant that contains a major band at ~38 kDa and the purified *A. aculeatus* X273 on phosphoric acid swollen cellulose (PASC). The soluble degradation products were characterized using high performance anion exchange chromatography coupled with amperometric detection (HPAEC-PAD), and the assay carried out using the bioreactor supernatant revealed an important release of soluble cello-oligosaccharides (DP2-DP6, see Figure 32), which was a good starting point for further investigations. However, the protein purified by IMAC displayed reduced activity on cellulose, even tested at high protein loading (Figure 32), suggesting that the purification using the standard IMAC column was detrimental for the activity of this enzyme.

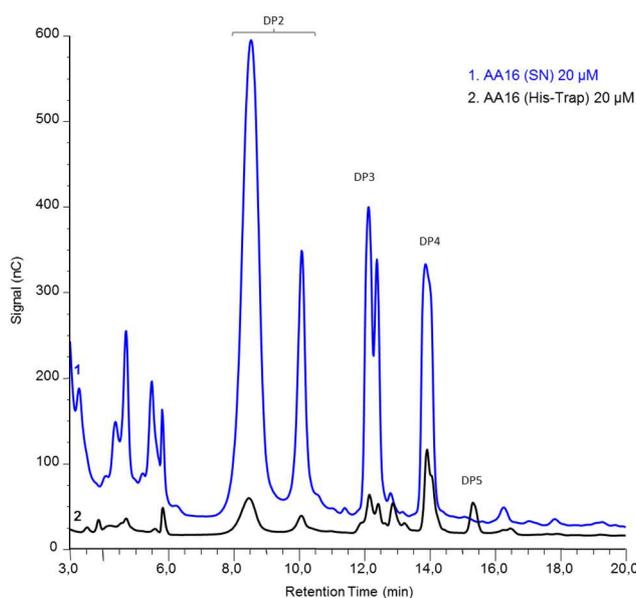


Figure 32 : HPAEC-PAD chromatograms of 0.1% PASC soluble degradation products, after 70h incubation with *P. pastoris* bioreactor supernatant (SN, blue chromatogram) or with AaAA16 protein purified by affinity chromatography using a nickel HisTrap column (black chromatogram).

Therefore, the protein was purified to homogeneity by ion exchange chromatography (Figure 33). ICP-MS analysis was used to confirm the presence of copper in the active site, with ~1 copper atom per protein molecule. The purified X273 protein from *A. aculeatus* was able to produce hydrogen peroxide without substrate in the presence of L-cysteine and ascorbate (Figure 34) which is a common feature for LPMOs (Kittl et al., 2012).

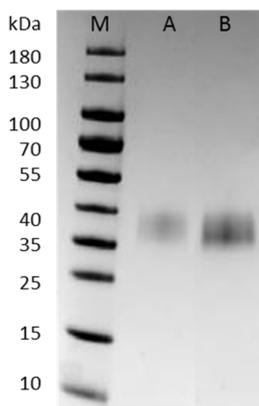
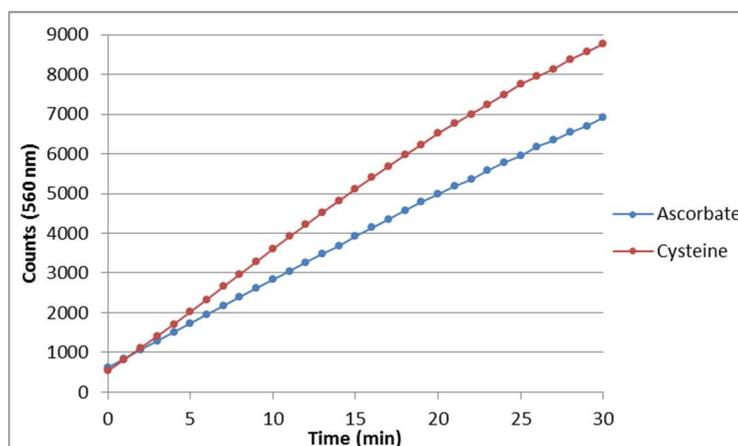


Figure 33 : SDS-PAGE analysis of AaAA16 before and after purification. Lane A, *Pichia pastoris* bioreactor supernatant; lane B, AaAA16 purified by ion exchange chromatography; lane M, Molecular weight protein ladder.

Figure 34 : Production of H₂O₂ by AaAA16 in the presence of ascorbate or L-cysteine, visualized using the Amplex red coupled assay. The produced H₂O₂ is used by horse radish peroxidase to transform Amplex Red into fluorescent resorufin, which is measured by fluorimetric counts at 560 nm.



Further enzymatic assays were performed at moderate enzyme loading on several substrates containing β -linked glucose units (phosphoric acid swollen cellulose (PASC), microcrystalline cellulose (Avicel), β -glucan, glucomannan) as well as cello-, xylo- and xyloglucan-oligosaccharides. Significant activity was detected on celohexaose (Figure 35) and cellulosic substrates (Figure 36A) with the release of a majority of non-oxidized products (DP2-DP5), and small peaks eluting at the same retention time as C1-oxidized products (DP2ox-DP4ox). Mass spectrometry was used to confirm the presence of oxidized oligosaccharides at a mass of m/z +16 compared to the non-modified oligosaccharides (Figure 36B). These peaks could correspond either to an aldonic acid form (C1-oxidized) or a gemdiol form (C4-oxidized), but the ketone form which is usually observed at m/z -2 in the case of a C4-oxidation was not found. The MS/MS analysis of the DP3 oxidized product, taken as an example, was consistent with the presence of an oxidation at the reducing end, providing further evidence that the X273 protein from *A. aculeatus* is a C1-oxidizing LPMO active on cellulose (Figure 36C).

Altogether, the results suggest that the characterized protein is a copper-containing enzyme able to produce hydrogen peroxide in the absence of substrate and able to oxidatively cleave cellulose with C1 regioselectivity. The enzymatic characterization of *A. aculeatus* enzyme establishes that the enzyme has LPMO activity and therefore family X273 was renamed AA16 in the CAZy database. Consequently, *A. aculeatus* X273 is named AaAA16.

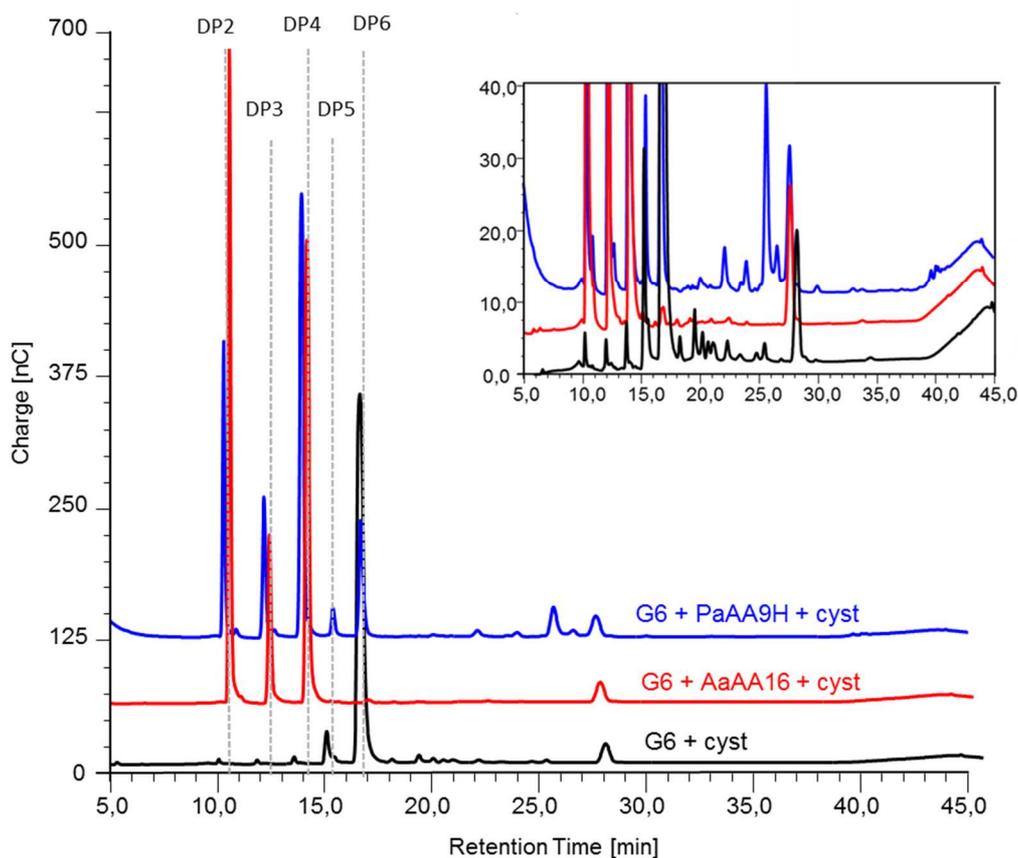


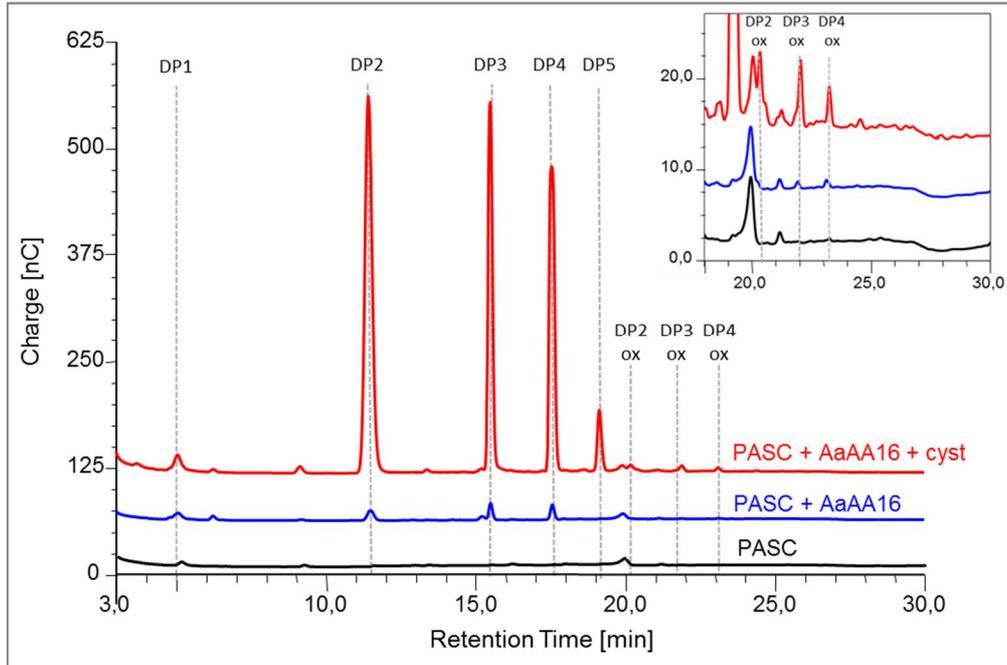
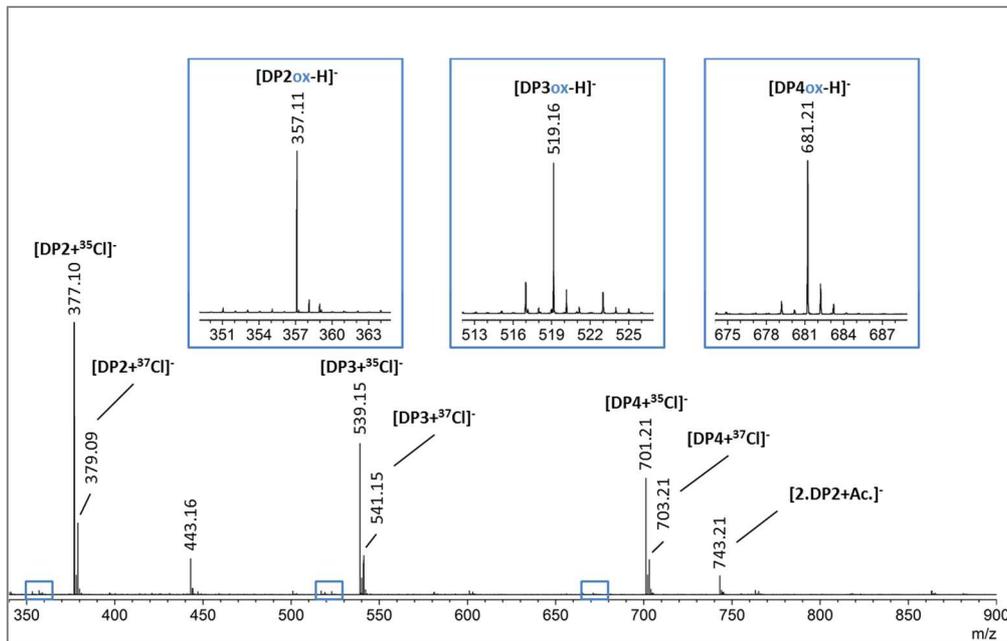
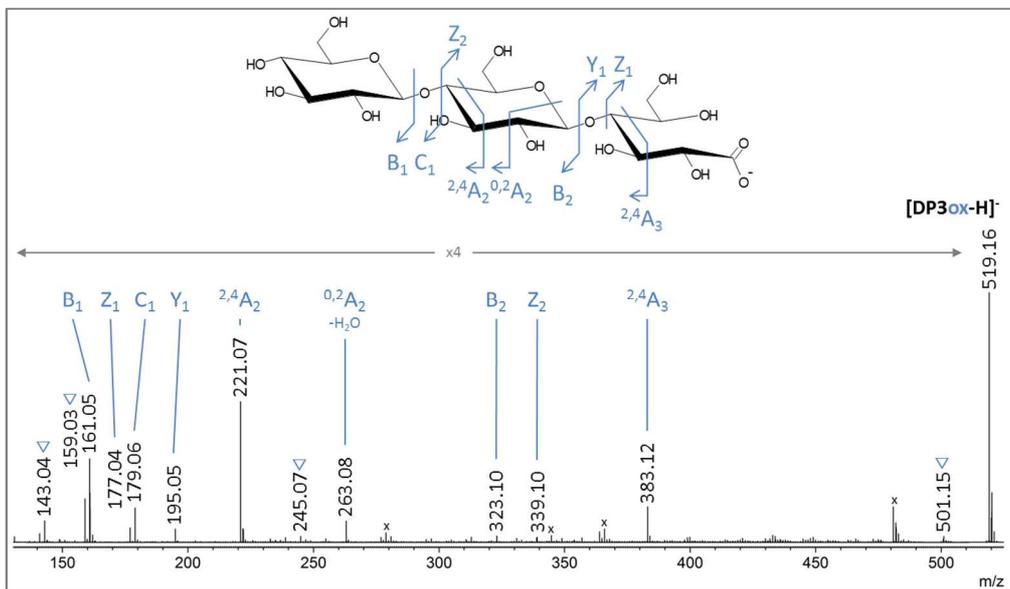
Figure 35 : HPAEC-PAD chromatograms showing soluble products generated from 0.5 mM cellohexaose (G6) using 4.4 μM AaAA16 and 1 mM L-cysteine. The LPMO9H from *Podospora anserina* (PaAA9H) (Bennati-Granier et al., 2015) was used as a reference of LPMO acting on cello-oligosaccharides (producing native sugars and C4-oxidized products, visible at 26 and 40 min in the box at the top right).

Figure 36 : Oxidative cleavage of cellulose by AaAA16 (see next page)

A. HPAEC-PAD chromatograms showing soluble products generated from 0.1% PASC cellulose using 4.4 μM AaAA16, with or without L-cysteine (1 mM). The peak annotations are based on comparison with oligosaccharides standards, native (DP1-DP5) or oxidized at the C1 position (DP2_{ox}-DP4_{ox}). The box at the right top show an enlargement of the C1-oxidized product region.

B. MS spectrum of soluble degradation products generated by AaAA16 from PASC. Mass spectrometry analysis of soluble degradation products generated by AaAA16 from PASC. The main panel shows the full-scale spectrum of the sample containing native and oxidized cello-oligosaccharides, and the boxed regions are magnified to show the peaks corresponding to oxidized cello-oligosaccharides.

C. MS/MS spectrum of the DP3 oxidized product peak (m/z 519.16) observed on panel B. The fragmentation pattern corresponds to a C1 oxidized species with an aldonic acid at the reducing end. Observed fragments are depicted on the structure. Blue triangles: Water losses. Black stars: contaminant peak from an ion co-isolated during the MS/MS precursor selection.

A.**B.****C.**

3.6 Synergy assays

Many LPMOs are able to act in synergy with cellulases and in particular with cellobiohydrolase I (CBHI) (Eibinger et al., 2014; Harris et al., 2010; Langston et al., 2011; Selig et al., 2015; Song et al., 2018), which is mainly responsible for the degradation of crystalline cellulose (Payne et al., 2015). To determine whether it is the case of the AaAA16 enzyme, a sequential assay was performed on nano-fibrillated cellulose (NFC) and PASC. The cellulosic substrates were first treated for 24h with AaAA16 in the presence of L-cysteine; the control condition contained L-cysteine but no AA16 enzyme. After inactivating enzymes by boiling, the soluble products were removed and the insoluble fraction was washed. *T. reesei* CBHI was then added for 2h, and the produced cellobiose was quantified by HPAEC-PAD. As shown on Figure 37, the product release from both cellulosic substrates was more important after AaAA16 treatment than in the condition without prior AA16 treatment, probably due to the new chain ends produced by LPMO action that allowed an increased CBHI activity. *T. reesei* CBHI being a major component of industrial cellulolytic cocktails, the ability of AaAA16 to improve its activity on cellulose is an interesting feature indicating that the new AA16 family could be explored for biotechnological applications.

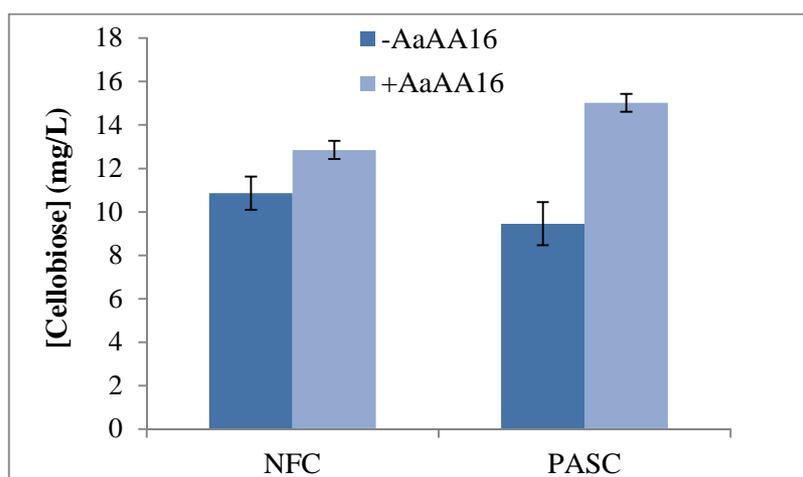


Figure 37 : Synergy between AaAA16 and CBHI for the degradation of cellulosic substrates.

Cellobiose released from 0.1% PASC or NFC by *T. reesei* CBHI (1 mg/g of substrate), with or without prior 24 h treatment by AaAA16 (10 mg/g), was quantified by HPAEC-PAD. Error bars represent the standard deviation calculated on three replicates

4 Discussion

The *A. aculeatus* AA16 LPMO characterized in this study shares many features with fungal LPMOs from other families, including (i) two strictly conserved histidines forming the copper histidine-brace, (ii) methylation of the first histidine residue in fungal secretomes, (iii) presence of one copper atom per protein and production of hydrogen peroxide (iv) ability to cleave cellulose and cello-oligosaccharides while producing C1-oxidized products, (v) synergy with CBHI for the degradation of cellulose. Although its

sequence and its modularity have long suggested that the family may have LPMO activity, none of its members has been biochemically characterized so far, possibly due to recombinant production and handling problems. AaAA16 is indeed a challenging enzyme and we faced difficulties in obtaining a functional protein due to self-oxidation of the recombinant enzyme produced in flasks and inactivation during the affinity chromatography purification. Oxidation of the N-terminal histidine of LPMOs is a frequently encountered issue due to the formation of reactive oxygen species at the active site in the presence of copper. Recently, it was hypothesized that the methylation of this residue in fungi could be a protective mechanism against self-inactivation of the enzyme (Petrovic et al., 2018). This post-translational modification is lacking in *P. pastoris*, which could lead to consider the use of another host organism such as *A. niger* for recombinant expression of AA16 enzymes. However, in the case of AaAA16, we managed to circumvent this problem with the set-up of a bioreactor production which led to a functional enzyme.

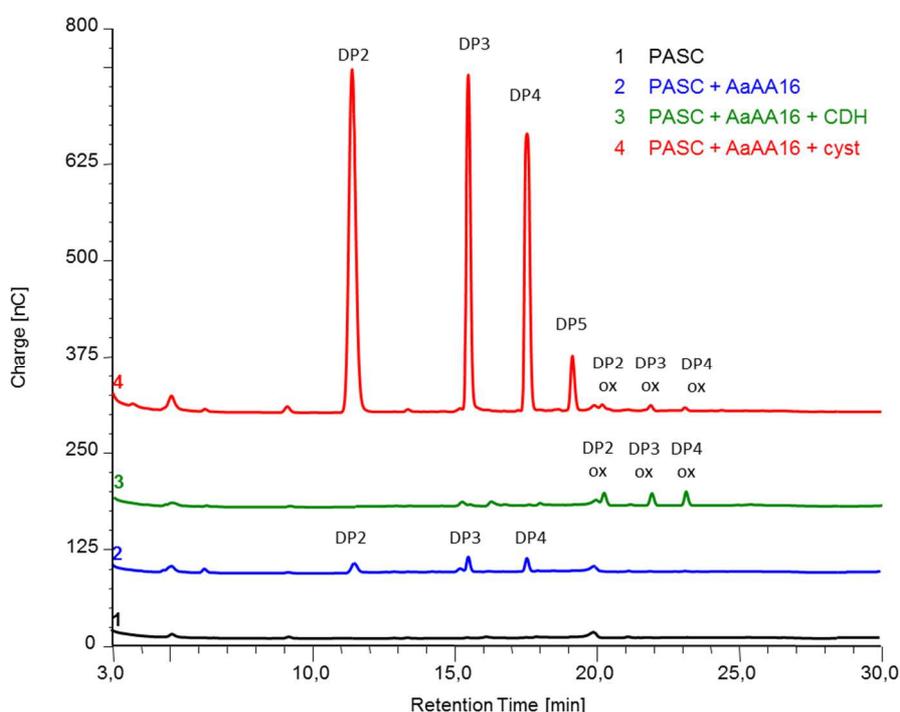


Figure 38 : HPAEC-PAD chromatograms of PASC soluble degradation products, after incubation with AaAA16 alone or using *Podospira anserina* cellobiodehydrogenase (PaCDHB) (Bennati-Granier et al., 2015) or L-cysteine as electron donors. The oxidized products visible in the CDH condition (green chromatogram) are the result of the action of CDH on cellodextrins released from PASC by AaAA16 (blue chromatogram). The proper LPMO activity is only visible after activation by L-cysteine (red chromatogram).

Another noticeable feature of this new LPMO is its atypical product profile, which differs from those of other fungal LPMOs active on cellulose. Indeed, the ratio of C1-oxidized vs. non-oxidized products is much lower than those of so far characterized AA9 enzymes, suggesting that the two families may have a different reactivity. More generally, the differences existing between AA9 and AA16 enzymes are of interest, considering that these two fungal LPMO families seem to be targeting the same substrate. It can be noted that

the gene expansion of the AA16 family, with an average of 5.7 sequences per genome for Oomycetes and 1.8 for Fungi, is strikingly lower than in the AA9 family, many fungal species having over 30 AA9 genes in their genome (Lenfant et al., 2017). Moreover, cellobiose dehydrogenase (CDH), which is a well-established enzymatic redox partner of AA9 LPMOs, was not found in the *A. japonicus* BRFM 1490 secretomes. Interestingly, *in vitro* assays with a recombinant CDH from *Podospora anserina* did not seem to promote AA16 activity (Figure 38). This observation could reveal a difference in the activation of this LPMO family compared to AA9 LPMOs. The secretomic analysis also revealed that two AA9 enzymes, including one with a CBM1, are co-secreted with *A. japonicus* AA16, which could indicate that they have a different biological role or a possible synergistic effect.

5 Conclusions

Using a comparative secretomic approach, we were able to identify and characterize the founding member of a new LPMO family that plays a role in plant biomass deconstruction, and could be of interest for biotechnological applications. Although *Aspergillus* species have been deeply investigated for decades and a large number of their CAZymes have been characterized, the activity of this new LPMO family had not been unveiled until now.

In order to better understand the action of members of this new family on plant biomass and their possible biological roles, it will be necessary to characterize other AA16 enzymes, including some multi-modular enzymes bearing a CBM1 or a GPI anchor. Further structural and spectroscopic investigations are also required to provide insights into the mechanism of the AA16 family.

Chapitre IV : Caractérisation de trois nouvelles LPMO de la famille AA16

1 Résumé

Suite à la caractérisation d'une première protéine ayant entraîné la création de la famille AA16 dans la base de données CAZy, il était nécessaire de poursuivre les travaux sur cette famille afin de mieux comprendre son activité. Pour cela, trois nouvelles cibles ont été sélectionnées. La première cible (appelée AjAA16) est la protéine AA16 issue d'*Aspergillus japonicus* BRFM 405 ; elle possède une séquence assez proche de celle d'AaAA16, issue d'*A. aculeatus*, mais représente potentiellement mieux la protéine présente dans les sécrétomes d'*A. japonicus* étudiés dans la première partie du projet. Les deux autres cibles sélectionnées (appelées ApAA16-CBM1 et PFAA16A-CBM1) sont issues d'*Aureobasidium pullulans* et de *Pestalotiopsis fici*, et comportent toutes les deux, en plus du module catalytique AA16, un module de liaison à la cellulose (CBM1).

Ces trois cibles ont été produites de manière recombinante chez *Pichia pastoris* en bioréacteur. Bien que leur caractérisation après purification ait révélé que l'histidine N-terminale impliquée dans le site actif n'était pas toujours intacte, ces trois protéines ont montré une activité de dégradation de la cellulose semblable à celle d'AaAA16, et même supérieure dans le cas d'AjAA16. Les trois nouvelles AA16, tout comme AaAA16, sont capables d'améliorer l'hydrolyse de cellulose par la CBHI de *T. reesei* lorsqu'elles sont utilisées pour prétraiter le substrat, et là encore AjAA16 est la plus performante. Cependant, ces quatre enzymes ont toujours un profil de produits solubles composé de très peu de produits oxydés, et ce phénomène reste inexpliqué. Les différents donneurs d'électrons testés n'ont pas permis la libération de produits oxydés supplémentaires. Afin d'avoir des informations sur la configuration du cuivre dans l'*histidine brace*, des expériences de spectroscopie par résonance paramagnétique électronique ont été réalisées. Les spectres confirment la présence de cuivre (II) dans le site actif, dans une symétrie de coordination axiale, avec des paramètres semblables à ceux déjà observés pour les autres familles de LPMO. Les expériences en présence de réducteur n'ont cependant pas montré de diminution du signal, ce qui aurait dû se produire si le cuivre (II) était correctement réduit en cuivre (I). L'une des pistes pour expliquer la faible production de produits oxydés pourrait donc être liée à l'état redox des AA16, et la RPE serait un bon moyen d'apporter des réponses aux questions qui se posent à ce propos.

D'autres informations essentielles sur la famille AA16 pourraient être déduites de la résolution de leur structure tridimensionnelle. Les cristaux testés n'ayant pas pour l'instant permis de collecter des données de diffraction, les structures 3D des quatre protéines étudiées ont été modélisées par homologie avec des protéines de séquences proches. Ces modèles révèlent une structure typique des LPMO, présentant des caractéristiques retrouvées dans les autres familles, notamment concernant les résidus proches du site actif.

Ces résultats étant destinés à être publiés, ils sont présentés ici sous la forme d'un article scientifique en anglais, à compléter et à soumettre dans les prochains mois.

Article 2: Functional characterization of three novel AA16 LPMOs from *Aspergillus japonicus*, *Aureobasidium pullulans* and *Pestalotiopsis fici*

Camille Filiatrault Chastel^{1,2}, Paul Jacottin², Simon Arragain², Nadia Touati³, Kristian Frandsen⁴, Senta Heiss-Blanquet², Antoine Margeot² and Jean-Guy Berrin¹

¹ Biodiversité et Biotechnologie Fongiques (INRA, Aix Marseille Université), Marseille, France.

² IFP Energies Nouvelles, 1 et 4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France.

³ Institut de Recherche de Chimie Paris (CNRS, Chimie ParisTech), Paris, France

⁴ Department of Chemistry, University of Copenhagen, Copenhagen, Denmark

2 Background

The discovery of lytic polysaccharide monooxygenases (LPMOs) approximately 10 years ago was a major breakthrough for the understanding of recalcitrant biomass degradation in Nature, and for applications in industrial biotechnological processes (Harris et al., 2010; Johansen, 2016; Vaaje-Kolstad et al., 2010). LPMOs are copper-containing enzymes, able to cleave various polysaccharides by oxidizing their glycosidic bonds, either at the C1 or the C4 carbon. This oxidizing mechanism requires an oxygen atom (provided either by dioxygen or hydrogen peroxide) and an extracellular electron donor (Bissaro et al., 2017; Bissaro et al., 2018).

LPMOs are classified as “Auxiliary Activity” enzymes in the CAZy database since 2013 (Levasseur et al., 2013), and until now seven different LPMO families have been defined, based on sequence and structural similarities. Some discrepancies in substrate specificities and regioselectivities can be found within the best characterized families (i.e. AA9 and AA10). Indeed, the scope of substrates found for the fungal AA9 enzymes was enlarged from cellulose to cello-oligomers and hemicellulose polymers (Fanuel et al., 2017; Simmons et al., 2017). In the bacterial AA10 family, chitinolytic and/or cellulolytic enzymes are found (Forsberg et al., 2014b). According to CAZy, these two families have respectively 29 and 22 characterized members, with respectively 16 and 19 of them having a solved 3D structure. It is far from being the case for more recently discovered families, with 3 characterized members in the AA13 family and only 2 in AA11, AA14 and AA15 (in September 2019). These low numbers can be explained by the fact that recombinant production of correctly processed and active LPMOs can be particularly challenging. Moreover, many pitfalls can be encountered during LPMO assays, such as multiple possible side reactions, enzyme self-inactivation, reducing agent self-oxidation, instability of some soluble reaction products, or difficulty to characterize insoluble reaction products (Eijsink et al., 2019). Nonetheless, more detailed characterization of enzymes belonging to each of the existing families is needed to better understand the reaction mechanism of LPMOs and their possible biological roles.

Concerning the last family created, AA16, only one of its members has been produced and characterized so far (Filiatrault-Chastel et al., 2019), and no structural data are yet available. The AA16 from *Aspergillus aculeatus*, AaAA16, is able to cleave cellulose and to produce both native and C1-oxidized sugar. In this study, we heterologously produced and characterized three other AA16 enzymes in order to gain a better understanding of this family. Their activity on cellulose was assessed and Electron Paramagnetic Resonance (EPR) spectrometry was used to study the environment of the copper center and its redox properties. Homology models were also built to get some information on the copper environment and substrate binding surface.

3 Results and discussion

3.1 Selection and production of three novel AA16 enzymes

The sequence of AaAA16, the first characterized AA16, has been retrieved from the publicly available genome of *Aspergillus aculeatus* ATCC16872. *Aspergillus japonicus* is a closely related species and the genome of strain *A. japonicus* CBS 114.5 is publicly available. The AA16s from these two strains (GenBank IDs XP_020060743.1 and XP_025526839.1) have almost identical sequences, with only one amino acid differing in their catalytic module. However, one of the *A. japonicus* strains maintained in the CIRM-CF collection (Marseille, France) has a gene encoding for a quite different AA16, since it displays a difference of 20 amino acids with the catalytic module of AaAA16. To assess the influence of these differences, it was decided to produce and characterize this enzyme from the CIRM-CF strain, further called AjAA16.

AaAA16 and AjAA16 are part of the same subgroup of the AA16 phylogenetic tree (Figure 39). This subgroup is the only one that contains sequences associated to a cellulose binding module (CBM1) in their C-terminal extension. However, like many enzymes in that group, AjAA16 does not bear a CBM itself. To assess the effect of such a module on the AA16 activity, two CBM1-bearing sequences were selected. The first one, ApAA16 (GenBank ID KEQ78697.1), stems from the ubiquitous black mold *Aureobasidium pullulans*, and the other one (PFAA16A-CBM1) originates from the endophytic fungus *Pestalotiopsis fici*. The genome of *P. fici* encodes three AA16s (denoted PFAA16A-CBM1, PFAA16B and PFAA16C) that belong to different subgroups in the AA16 family (Figure 39B), but only one of them (PFAA16A-CBM1, Genbank ID XP_007832240.1) bears a CBM1 and was therefore selected for biochemical characterization.

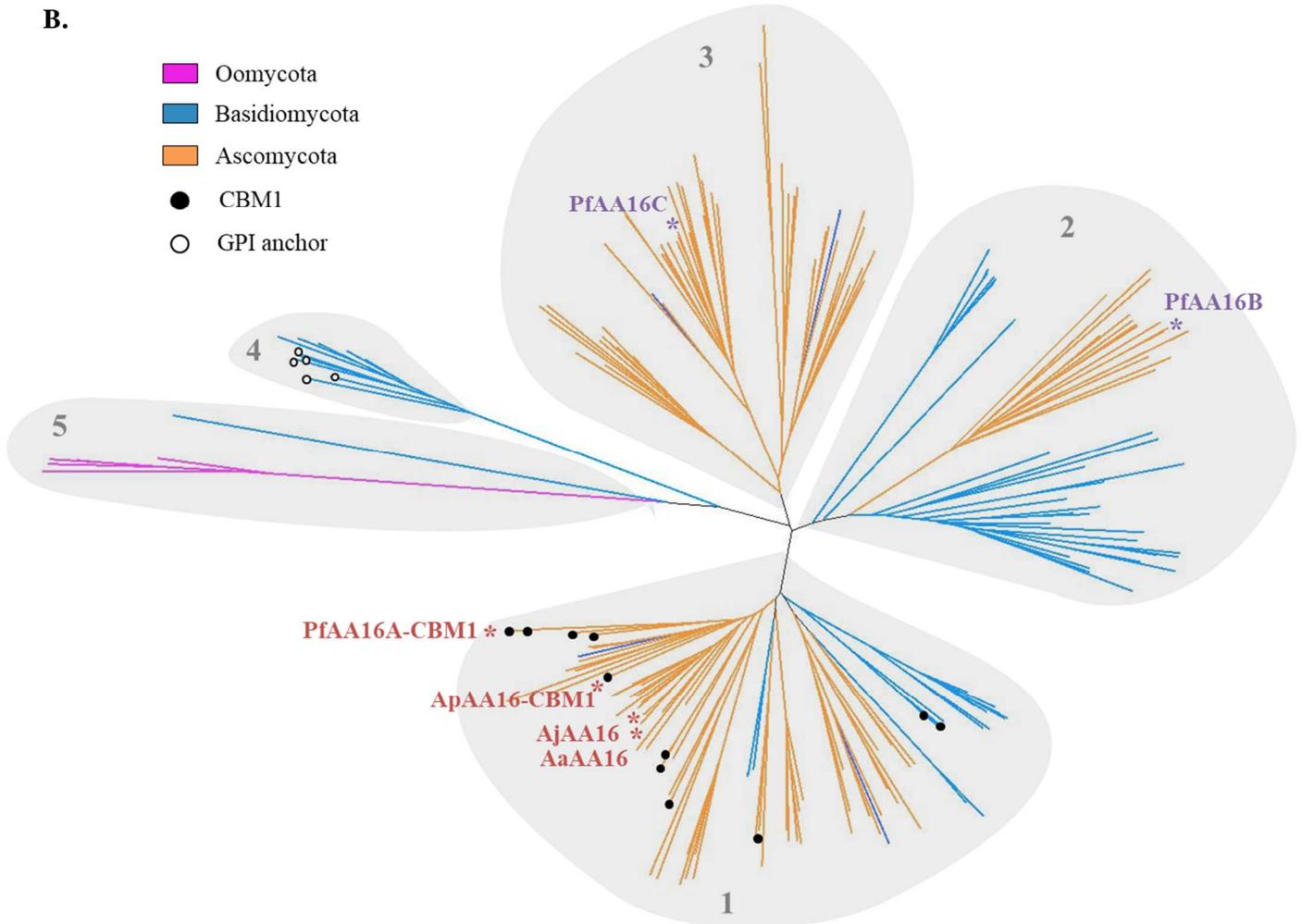
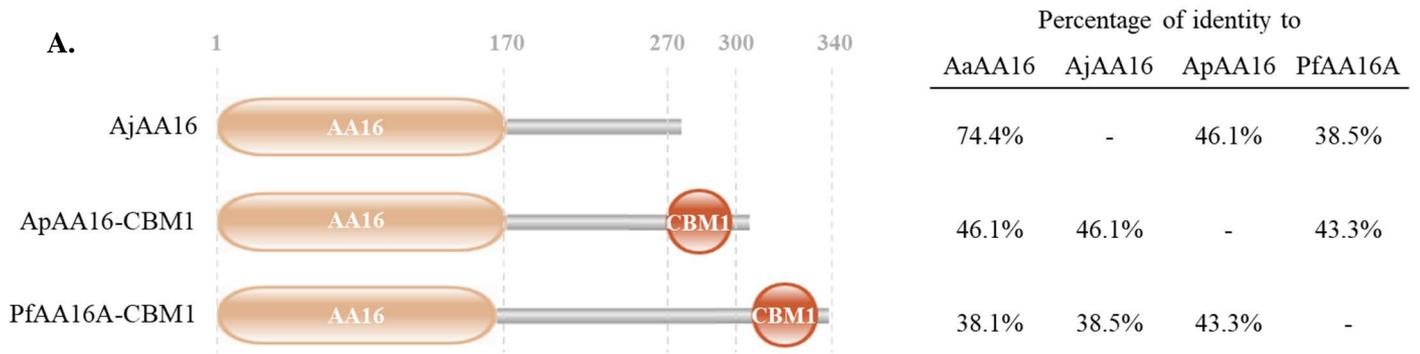


Figure 39 : Modularity and phylogeny of the produced AA16 LPMOs

A. Modular representation of the newly produced AA16 (AjAA16, ApAA16-CBM1, PfAA16A-CBM1), indicating the percentage of amino acid identity they share with each other and with previously produced AaAA16.

B. Radial phylogram of 208 selected sequences from the AA16 family, based on the alignment of their catalytic modules, showing the sequences from *Aspergillus aculeatus* (AaAA16), *Aspergillus japonicus* (AjAA16), *Aureobasidium pullulans* (ApAA16) and *Pestalotiopsis fici* (PfAA16A-CBM1, PfAA16B and PfAA16C; only PfAA16A-CBM1 was recombinantly produced).

These three targets (AjAA16, ApAA16-CBM1 and PfAA16A-CBM1) were expressed in *Pichia pastoris* using native signal sequences. In the case of AjAA16, the catalytic module is followed by a long C-terminal extension (110 amino acids), rich in serine, alanine and threonine residues, which was predicted as disordered. For recombinant production, this extension was removed to leave only the catalytic module, while for the two other targets we used the full length version including the CBM1 domain. The three recombinant proteins (AjAA16, ApAA16-CBM1 and PfAA16A-CBM1) were produced in 1L-bioreactors, in the presence of trace metals including copper (Figure 40). The proteins were then purified from the culture supernatant using anion exchange chromatography, and the fractions containing eluted protein were analyzed by SDS-PAGE. All fractions from the elution peak of ApAA16-CBM1 contained proteins of the same size, ~70 kDa, higher than the expected 34 kDa, probably due to N- and O-glycosylation (Figure 41A, B and C; Table 6). These fractions were thus pooled for subsequent tests. In contrast, different fractions from AjAA16 elution contained proteins of different sizes, and only the fractions corresponding to the second part of the peak showed an increased cellulose degradation activity in the presence of L-cysteine, with traces of oxidized products (Figure 41D, E and F); these fractions were pooled together and used for subsequent tests. A similar strategy was applied to PfAA16A; only the fractions from the second part of the elution peak allowed the formation of small amounts of cellulose degradation products in the presence of L-cysteine (Figure 41, D, G and H) and were thus pooled for subsequent use. After dialysis and concentration, the final production yields were estimated at 32, 55 and 3 mg of purified protein per liter of culture for AjAA16, ApAA16-CBM1, and PfAA16A-CBM1, respectively.

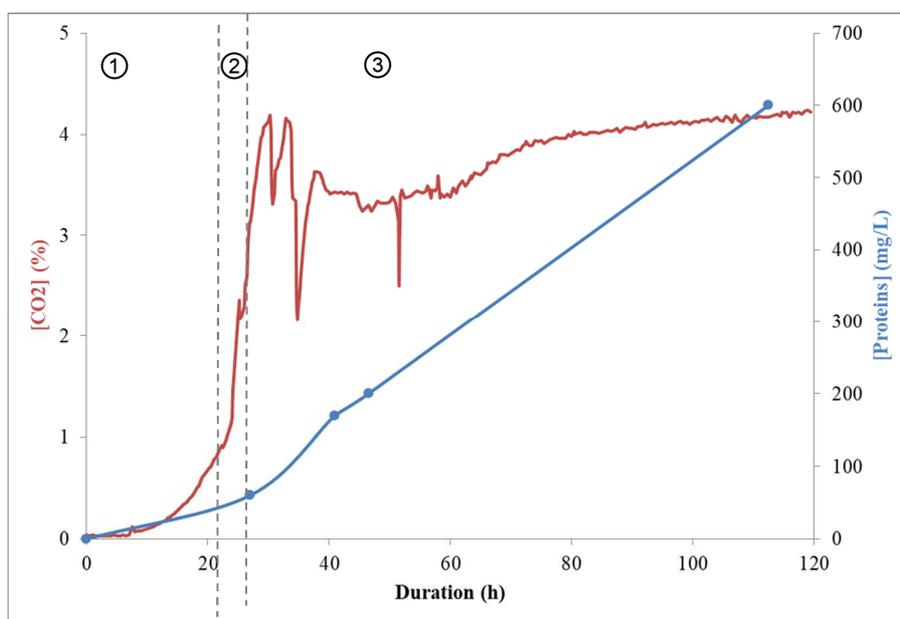


Figure 40 : Time profiles of dissolved CO₂ (showing the biomass growth) and protein production during the production of AjAA16 in a *Pichia pastoris* bioreactor. 1: Glycerol batch phase; 2: Sorbitol and methanol transition phase; 3: Methanol fed-batch phase.

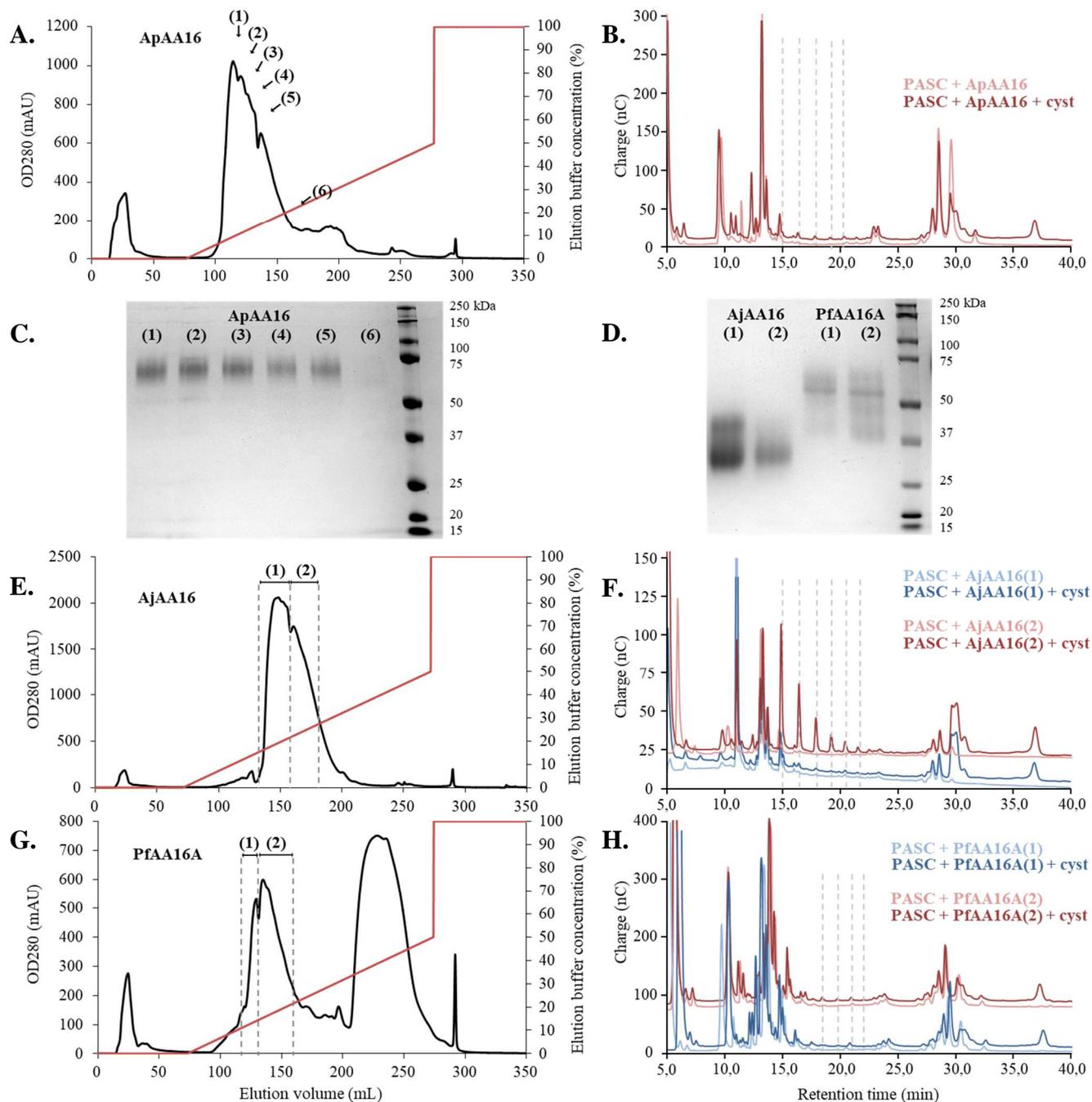


Figure 41 : Purification of AA16 enzymes

A, E, G. Anion exchange chromatograms of the 3 produced proteins (**A.** ApAA16-CBM1; **E.** AjAA16; **G.** PfAA16A-CBM1), showing the optical density at 280 nm (OD280) used to detect the proteins, and in red the slope of the elution buffer gradient.

C,D. SDS-PAGE analysis of selected fractions from the 3 protein elutions (**C.** ApAA16-CBM1; lanes (1) to (6) contains the 6 fractions marked in **A** by numbered arrows ; **D.** AjAA16 and PfAA16A-CBM1; lane marked as (1) and (2) contains the pooled fractions identified by (1) and (2) in **E** and **G**). In both gels the last lane contains a molecular weight protein ladder.

B, F, H. HPAEC-PAD chromatograms showing soluble products generated from PASC cellulose using pooled fractions from AA16 enzyme purifications (**B.** ApAA16-CBM1; **F.** AjAA16; **H.** PfAA16A-CBM1) with or without L-cysteine. The blue chromatograms were obtained using pooled fractions from the first part of the elution peaks of AjAA16 and PfAA16A-CBM1 (identified by a (1) in **E** and **G**); the red chromatograms were obtained using all fractions of ApAA16, or pooled fractions from the second part of the elution peaks of AjAA16 and PfAA16A-CBM1 (identified by a (2) in **E** and **G**). The dashed lines indicate the position of the oligosaccharide peaks that are seen only in the presence of fractions (2) and L-cysteine.

Global mass determination by MALDI-ToF mass spectrometry confirmed that the experimental mass of AjAA16 (23.56 kDa) was higher than the expected 21.79 kDa, probably due to glycosylations. On the other end, both ApAA16-CBM1 and PfAA16A-CBM1 had experimental masses lower than the theoretical ones, with a difference of ~10 kDa (Table 6). This could be due to the fact that both these proteins bear a very long linker, with about 100 and 130 amino acids, which could entail folding problems and possibly some proteolytic cleavage in this flexible region during recombinant protein production. However, given that their mass is higher than the expected mass of the AA16 domain alone, we can expect their catalytic domains to be intact and to have some activity, provided that the N-terminus is intact. N-terminal sequencing revealed that unfortunately, not all of the proteins were correctly processed. In 30% of the purified AjAA16 sample, either the first residue or the first five residues were cleaved off, and this was also the case for ApAA16-CBM1 and PfAA16A-CBM1 samples with respectively 5 and 11% of cleaved proteins. When the N-terminal residue was present, it was not detected as a histidine, but as a non-identified modified amino acid. Histidine is often detected with lower yields than other amino acids, and its coordination with a copper ion could impair even more its detection but in this case, the most probable explanation is the presence of oxidative modifications, as already observed for the AaAA16A LPMO (Filiatrault-Chastel et al., 2019). Proteins produced in yeasts such as *Pichia pastoris* do not bear a methylation on the N-terminal histidine, which is thought to protect the histidine against oxidative damage (Petrovic et al., 2018); in order to obtain proteins with an intact histidine brace, other production hosts such as *Aspergillus niger* could be considered. However, since the purified proteins displayed some activity on cellulose (see Figure 41 and next paragraph), some of the proteins are expected to harbor an intact N-terminal histidine residue coordinating the copper ion.

Table 6 : Main features of the produced LPMOs.
Glycosylation sites were predicted using the Hirst group glycosylation prediction server (Hamby and Hirst, 2008).

	AjAA16	ApAA16	PfAA16A
Presence of a linker and a CBM1	No	Yes	Yes
Theoretical molecular weight (kDa)	21.79	33.88	36.52
Experimental mass (kDa)	23.56	23.52	26.92
Number of theoretical N-glycosylation sites	4	6	8
Number of theoretical O-glycosylation sites	21	55	59
Theoretical pI	4.43	6.2	3.87

3.2 Activity of AA16 enzymes on cellulose

The three newly produced AA16 enzymes are able to degrade PASC cellulose into native and C1-oxidized soluble sugars; as already observed with AaAA16 in our previous work, the amount of visible oxidized sugars is very low. However, some differences can be spotted between the enzymes: the amounts of native and oxidized sugars released by AjAA16 are higher than those released by ApAA16-CBM1 and PfAA16A-CBM1, which are comparable to those obtained using AaAA16 (Figure 42). The role of CBMs on LPMO activity is still poorly understood, but in some cases their presence improved the binding of LPMOs to their substrate and modified the profiles of released products (Chalak et al., 2019; Crouch et al., 2016). However, given the MS results and the probable cleavage of CBM1 from the catalytic modules of ApAA16 and PfAA16A (see above), it is not surprising that these proteins do not have an increased activity compared to AaAA16.

Since AjAA16 was the most active enzyme, the release of soluble sugars produced from PASC was followed over 24 hours using HPAEC-PAD. The amount of short native sugars (DP2-DP4) regularly increased over this period, while the amount of longer native sugars (DP5-DP7) increased in the first 1 to 4 hours and then decreased, probably due to degradation by AjAA16 into shorter products. Even though the amount of detected C1-oxidized products was very low, an increase of DP3 and DP4 oxidized sugars is visible over time (Figure 43).

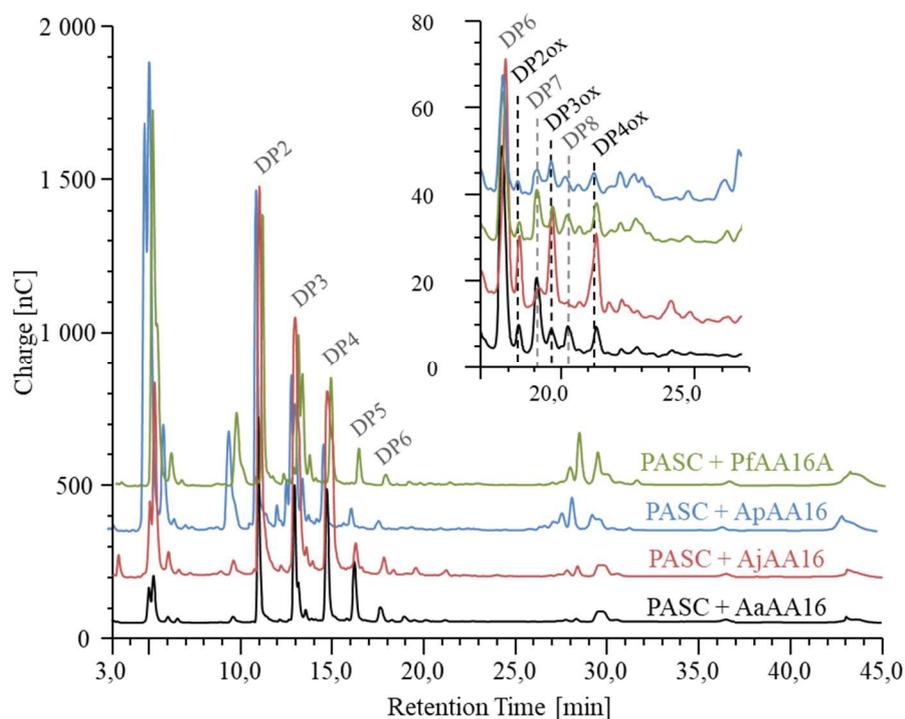


Figure 42 : HPAEC-PAD chromatograms showing soluble products generated from 1% PASC, after a 24h reaction, using AA16 enzymes (5 μ M) and L-cysteine (1mM). The peak annotations are based on comparison with oligosaccharides standards, native or oxidized at the C1 position. The box at the right top shows an enlargement of the C1-oxidized product region.

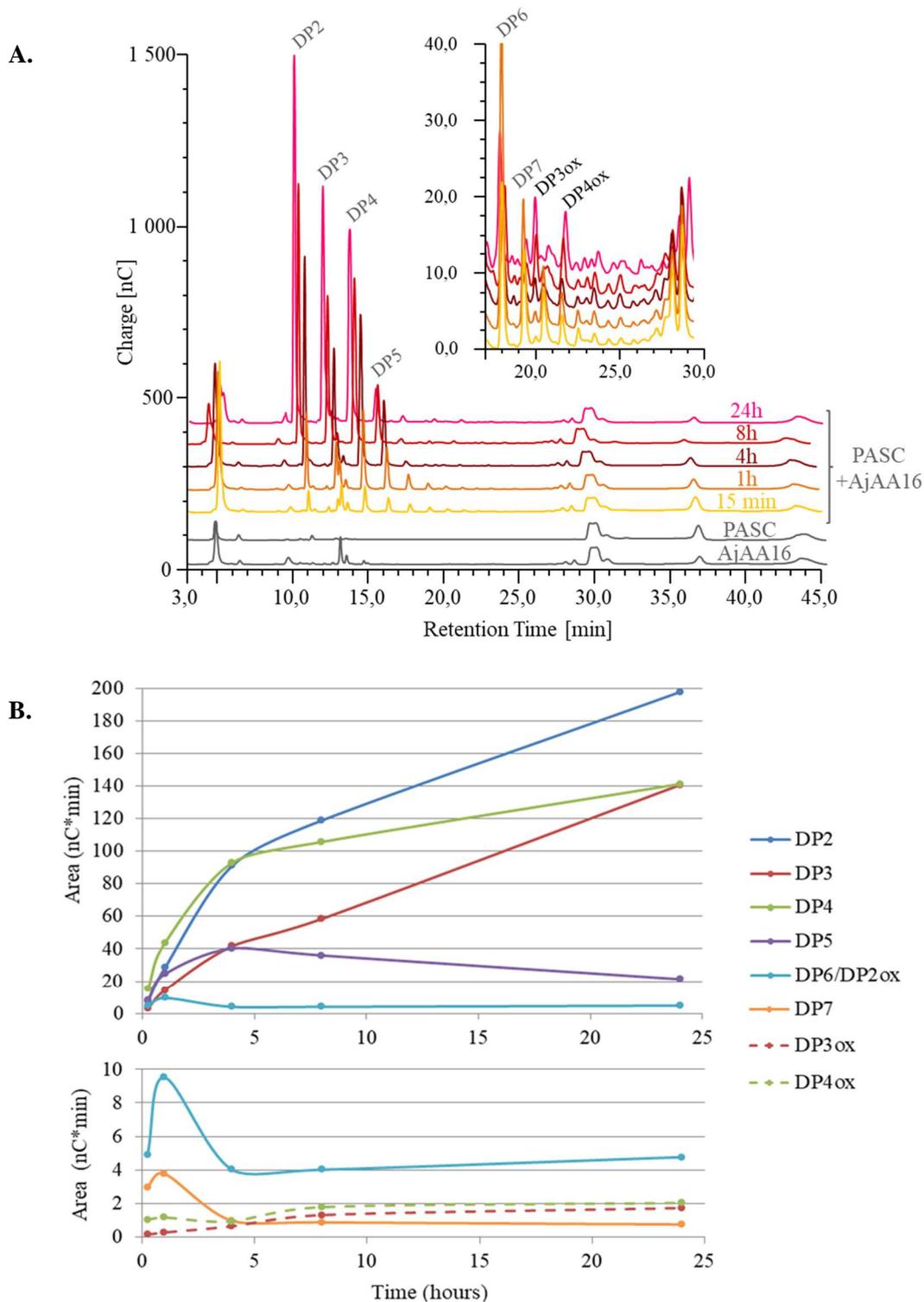


Figure 43 : Time course of cellulose degradation using AjAA16

A. HPAEC-PAD chromatograms showing soluble products generated from 0.1% PASC cellulose using AjAA16 ($2\ \mu\text{M}$) and L-cysteine ($1\ \text{mM}$). The peak annotations are based on comparison with oligosaccharides standards, native or oxidized at the C1 position. The box at the right top shows an enlargement of the C1-oxidized product region.

B. Evolution over time of the HPAEC-PAD peak areas, for native sugars (DP2-DP7) and C1-oxidized sugars (DP2-DP4). Due to the lack of standards for celloheptaose (DP7) and to the difficulty to differentiate the cellohexaose (DP6) and cellobionic acid (DP2ox) peaks, the areas were not converted into product concentrations.

3.3 Redox partners

Given the low amount of oxidized products that were released from PASC using the AA16 LPMOs with L-cysteine, we tested several other electron donors in order to check whether they were able to increase the oxidizing activity of the enzymes. Among the twelve compounds that were tested, only three were not able to activate AaAA16: tannic acid, pyrogallol and menadione. All the others (L-cysteine, ascorbic acid, p-coumaric acid, sinapic acid, vanillic acid, caffeic acid, syringic acid, gallic acid and 3-hydroxyanthranilic acid) triggered the activity of AaAA16 and allowed the release of soluble products from PASC (Figure 44). However, the majority of these products were native sugars, and none of the tested reductants helped to release more oxidized sugars than what had been observed with L-cysteine.

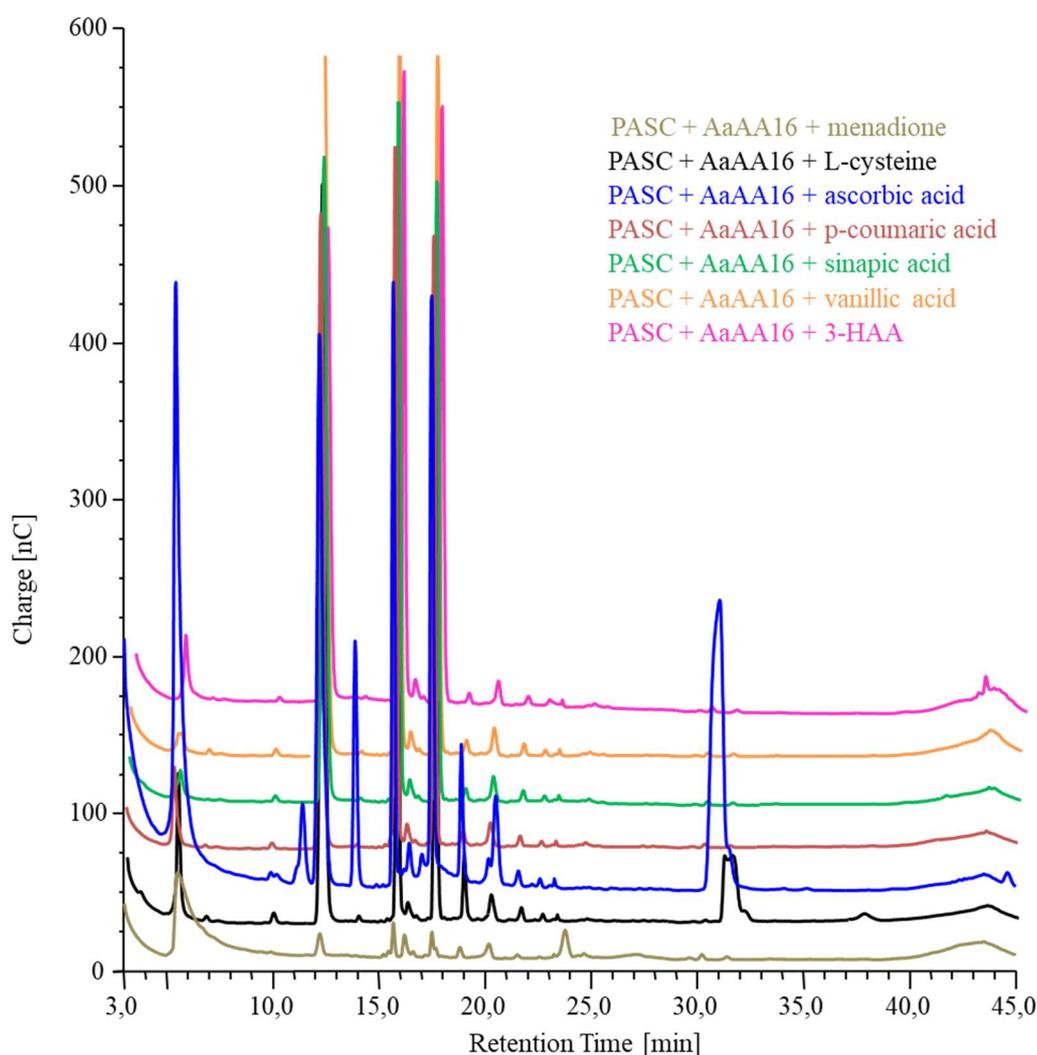


Figure 44 : Influence of different electron donor on the products released by AaAA16.

HPAEC-PAD chromatograms showing soluble products generated from 0.1% PASC cellulose using AaAA16 (4.4 μM) and 1mM of various reductants. Products profiles using tannic acid and pyrogallol are not shown but were very similar to the one obtained with menadione; products profiles using caffeic acid, syringic acid and gallic acid were very similar to the 6 others.

3-HAA: 3-hydroxyanthranilic acid.

3.4 Synergy assays

In our previous work, a 24h treatment of cellulosic substrates with AaAA16 showed to be beneficial for the hydrolytic activity of *T. reesei* cellobiohydrolase I (CBHI), with a 60% increase of released cellobiose compared to the condition without LPMO treatment (Filiatrault-Chastel et al. 2019). The reason for using a sequential assay with washing steps, and not adding the two enzymes simultaneously as in a classical synergy assay, is that the high quantity of native sugars released by AA16 enzymes would make the results difficult to interpret. A similar assay was used to test the newly produced AA16 enzymes, with a 24h LPMO treatment and several washes of the released soluble sugars, followed by hydrolysis using CBHI. All AA16 enzymes increased the amount of released sugars, with a 75% improvement using AjAA16, and about 25 and 30% using PfAA16A and ApAA16 (Figure 45). This result shows that the three enzymes are able to modify the PASC substrate in a way that renders it more amenable for hydrolysis by CBHI, possibly by generating new chain ends to attack. The better performance of AjAA16 seems to be in line with the higher quantity of oxidized products it is able to release.

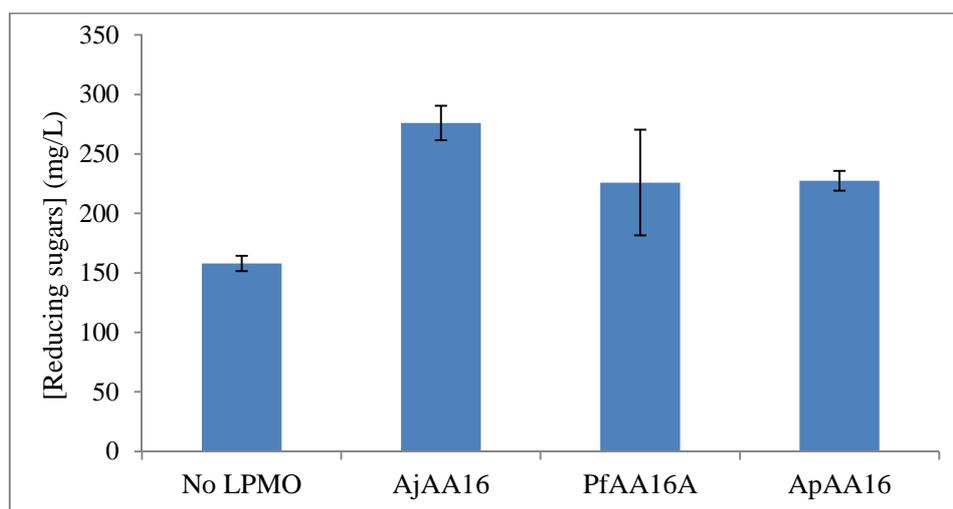


Figure 45 : Improved cellulose degradation using TrCBHI after a pretreatment with AA16 enzymes.

Reducing sugars released from 1% PASC by *T. reesei* CBHI (1 mg/g of substrate), with or without prior treatment by AA16 enzymes (10 mg/g), were quantified by DNS assay and expressed in cellobiose equivalents. Error bars represent the standard deviation calculated on three replicates.

3.5 Spectroscopic analysis of the active site

The configuration of the active site of AA16 enzymes was analyzed by EPR spectroscopy. The first tested enzyme was ApAA16, which displayed a typical mononuclear copper center spectrum (Figure 46, black spectrum) confirming the presence of Cu(II) ion in the active site. The determined spin-Hamiltonian parameters ($g_x = 2.058$, $g_y = 2.045$, $g_z = 2.25$; $|A_x| = 108$ MHz, $|A_y| = 45$ MHz, $|A_z| = 495$ MHz) reveal an axial symmetry with a very slight orthorhombic distortion. The g_z and $|A_z|$ values are in the range of those

obtained for all the other LPMO families (Hemsworth et al., 2018), and place the copper center within a Type 2 Peisach–Blumberg classification (Peisach and Blumberg, 1974), typical of LPMOs.

To investigate redox properties, ApAA16 was incubated with a 100-fold excess of ascorbate as a reducing agent. No change in the EPR spectrum was visible 10 minutes after addition, and even after 20 hours the signal was only slightly depleted (Figure 46, red spectrum), showing that the Cu(II) was only partially reduced to Cu(I). Comparatively, in the work describing the characterization of AA15 LPMOs, the Cu(II) signal disappeared totally after 10 minutes of incubation (Sabbadin et al., 2018). The fact that ascorbate is not able to fully reduce the copper in the active site could be linked to the low amount of oxidized products observed in the activity tests using ApAA16 and other AA16 enzymes with various electron donors.

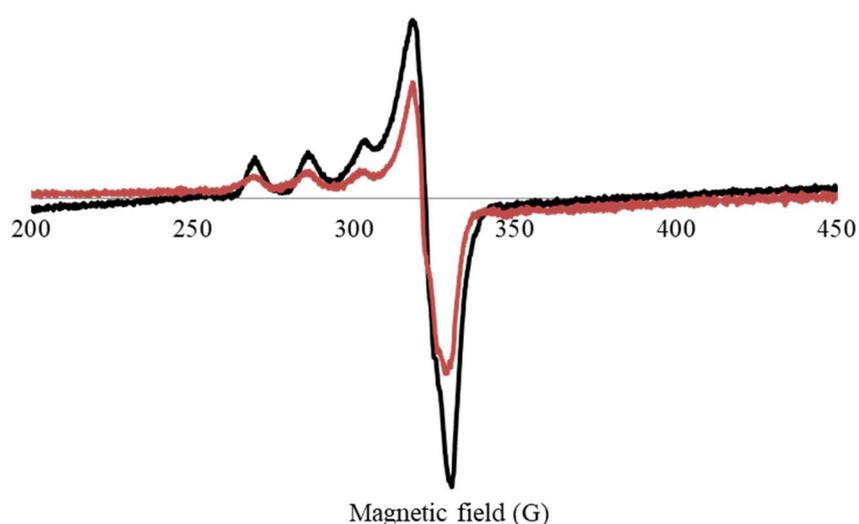


Figure 46 : Continuous wave X-band EPR spectra of ApAA16 in water, recorded before (black) and 20 h after addition of 100 equivalents of ascorbate (red). Spectra were recorded at 100 and 110K, 9.25 GHz, 4 mW power and 5 G modulation amplitude.

It must be noted that the apparently difficult reduction of the copper center in the absence of substrate does not necessarily mean that the same phenomenon occurs in the presence of a substrate. Indeed, gallic acid was not able to reduce the copper center of TdAA15A in EPR experiments conducted without substrate, but it was able to activate the protein in activity assays (Sabbadin et al., 2018). The authors suggest that it may be beneficial for the enzymes to be reduced only in the presence of substrate to avoid a loss of stability. To explore these hypotheses, more EPR experiments on several AA16 enzymes, with and without substrate (such as cellobiohexaose) and in the presence of various reductive agents, should be conducted.

3.6 Structural data

Interpretation of the spectroscopic data would be facilitated if an AA16 3D structure was available. AaAA16 and AjAA16, which are devoid of CBM1 module, were submitted to crystallization trials. As the proteins produced in *P. pastoris* were highly glycosylated, they were first incubated with Endoglycosidase H, leading to nearly full deglycosylation. Several crystallization screening assays were performed, allowing to obtain and mount several crystals (Figure 47), which were tested at synchrotrons facilities. Unfortunately, so far no protein diffraction was observed. The crystallization conditions are still being optimized; one option would be to use shorter proteins, since the ones that were tested bear at their C-terminal a *c-myc* epitope and a polyhistidine tag causing the region to be slightly disordered, which could impair correct crystallization (Figure 48).

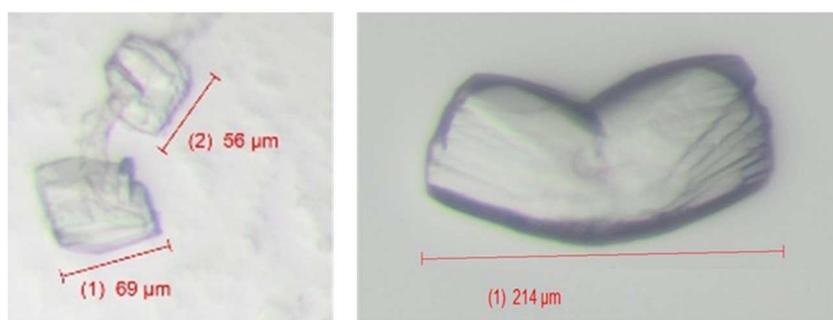


Figure 47 : Examples of crystals obtained with AaAA16 (left) and AjAA16 (right)

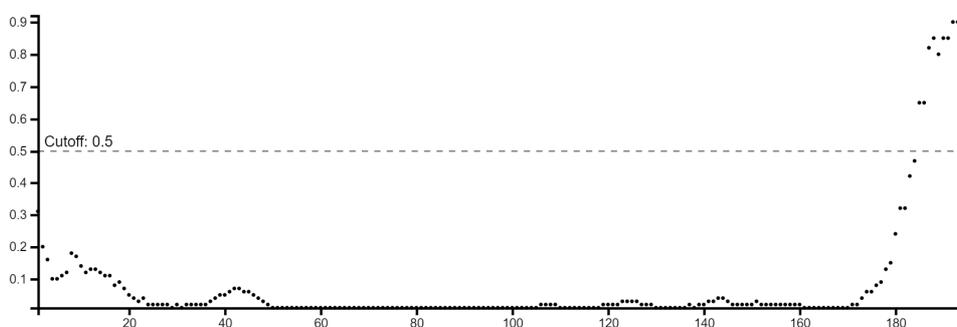


Figure 48 : Plot of the predicted disorder in the AaAA16 sequence

In the absence of a three-dimensional structure for AA16 enzymes, we built a model using Phyre2 and I-TASSER servers. PDB proteins that have the closest structural similarity to the AA16s are the AA15 from *Thermobia domestica* and several AA10 enzymes. The overall fold is a central immunoglobulin-like β -sandwich, typical of LPMOs. For the four proteins, the strictly conserved histidines at positions 1 and 90 (86 for PfAA16A) are found exposed at the surface, in a position where they could potentially coordinate a copper ion (Figure 49). A conserved tyrosine at position 164 (156 for PfAA16A) is also found in their close vicinity, as it is the case in most LPMO structures solved so far, with the notable exception of chitin active AA10

enzymes where this residue is usually a phenylalanine. The implication of two equatorial histidines and an axial tyrosine in the histidine brace is supported by the fact that the first ligand-binding site to be predicted by I-TASSER is a copper binding site involving these three residues. The active site tyrosine is part of a Q-T/N-Y-X-X-C motif, highly conserved in the AA16 family; a similar motif is found in AA9s, and it is known that mutations of glutamine to glutamate and tyrosine to phenylalanine are both detrimental to the AA9 activity (Harris et al., 2010; Vu et al., 2014a). Very close to the active site is found an alanine (position 88, or 85 in PfAA16A) that is strictly conserved across the clade of the phylogenetic tree containing the four studied proteins, and that is also found in a similar position in TdAA15A and AA10s, with an influence on the copper ion axial ligand position. Finally, it can be noticed that in AaAA16, AjAA16, and PfAA16A, a tyrosine residue in position 25 is exposed on the surface; in AA10s and AA9s, similarly positioned aromatic residues are involved in substrate binding (Forsberg et al., 2014a; Frandsen and Lo Leggio, 2016; Vaaje-Kolstad et al., 2017).

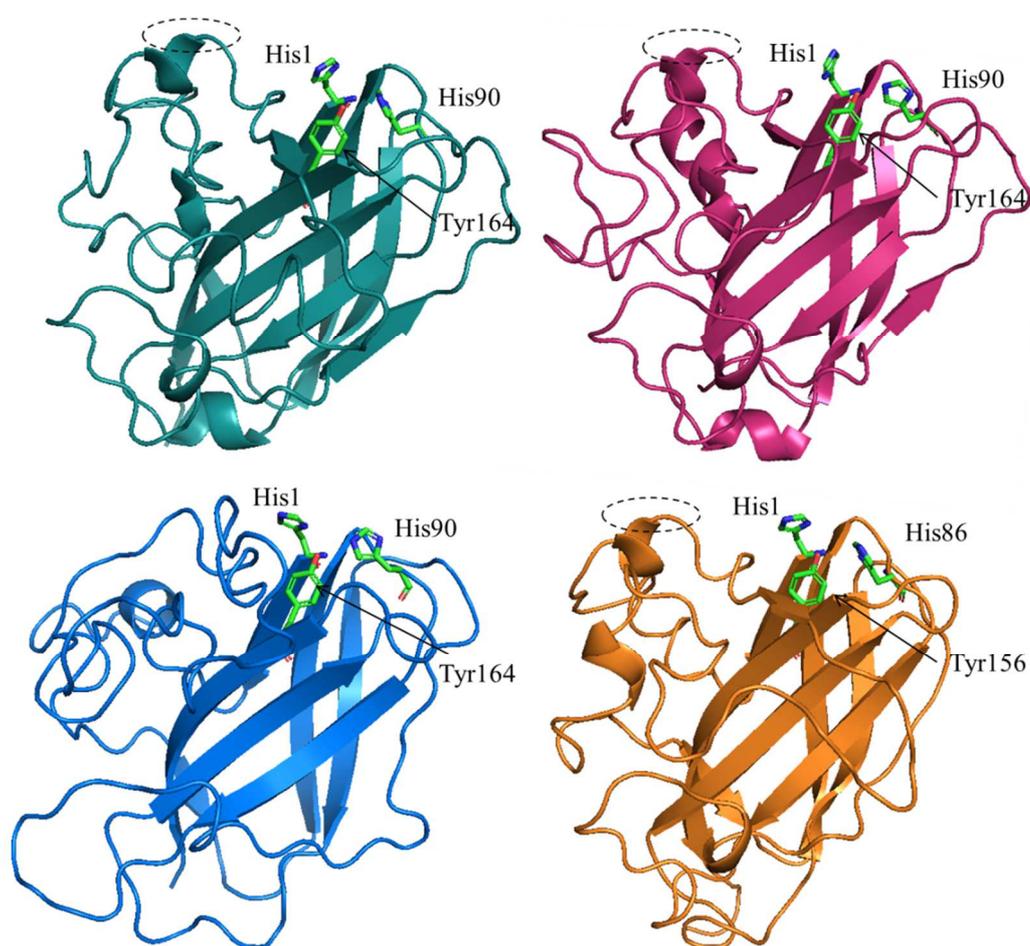


Figure 49 : 3D structures of AA16 enzymes modeled using Phyre2 server (Kelley et al., 2015).

AaAA16 is shown in green, AjAA16 in pink, ApAA16 in blue and PfAA16A in orange. The residues potentially involved in the binding of copper are shown as sticks; the position of the Tyr25 residue potentially involved in substrate interaction is circled by a dotted line.

4 Conclusions

Despite some recombinant production issues, the three novel AA16 enzymes produced in this work have similar properties compared to the founding member of the family, AaAA16, regarding degradation of cellulose and synergy with a CBH from *T. reesei*. Preliminary EPR experiments confirmed that the copper center of AA16 enzymes had a similar geometry as already observed in other LPMO families. Homology models revealed a 3D structure presenting some similarities with AA9, AA10 and AA15 families, especially concerning the residues potentially involved in copper coordination and substrate recognition.

Further characterization of AA16 enzymes is still required to fully understand the activity of this family, and especially the very low release of C1 oxidized products that was observed so far. In this view, more structural and spectroscopic investigations would be of great interest.

Chapitre V : Discussion générale et conclusions

1 Discussion générale

L'objectif initial de ce projet était d'identifier, sans a priori, une ou plusieurs protéines capables de compléter un cocktail de référence produit par *T. reesei* pour la saccharification de biomasse prétraitée. Il pouvait s'agir de protéines n'ayant pas d'équivalent dans le génome de *T. reesei*, ou dont l'activité était déjà présente mais peu exprimée. Les protéines d'intérêt pouvaient appartenir aux CAZymes ou à d'autres classes d'enzymes, avec une préférence pour des protéines pas encore caractérisées, voire de fonction inconnue. Bien qu'il s'agisse d'un projet exploratoire et très ouvert, la méthode à utiliser était en revanche bien définie. Elle s'appuyait sur l'exploration de sécrétomes de souches proches les unes des autres, mais générés dans des conditions permettant potentiellement la sécrétion de répertoires enzymatiques variés et distincts. Ces compositions variées devaient entraîner des performances variées pour la supplémentation du cocktail de référence, et donc permettre une étude comparative.

1.1 Exploration des sécrétomes et choix des cibles enzymatiques

1.1.1 Analyse comparative du contenu protéique des sécrétomes

Les premières différences entre les sécrétomes produits ont été remarquées dès la mesure de leurs concentrations en protéines (voir Table 5) : la concentration des sécrétomes produits sur Avicel est beaucoup plus faible que celle des sécrétomes produits sur fraction autoclavée de son de maïs (ASM ; MB en anglais) ou pulpe de betterave (SBP), quelle que soit la souche. Cela pourrait s'expliquer par le fait que la cellulose cristalline est un substrat à la composition moins complexe que celle des deux autres inducteurs, qui semble donc déclencher moins de voies de signalisation aboutissant à la sécrétion d'enzymes. Il s'agit aussi d'un substrat difficile à hydrolyser, sur lequel la croissance des souches fongiques est potentiellement plus lente ; cela aurait donc conduit à une production de biomasse fongique plus faible que sur les autres inducteurs, et donc à une production d'enzymes elle aussi diminuée. Suite à cette constatation, on pouvait s'attendre à ce que les sécrétomes produits sur ASM et SBP contiennent une plus grande diversité de protéines, et en particulier de CAZymes ; on peut voir sur la Figure 28 que c'est le cas pour toutes les souches, sauf la souche BRFM 1489, dont le sécrétome sur Avicel contient légèrement plus de CAZymes que les deux autres sécrétomes, et plus du double d'autres protéines.

La Figure 28 présentée dans le chapitre III permet de visualiser la composition globale des sécrétomes en termes de nombre de protéines différentes identifiées et de familles de protéines représentées, mais il est difficile de rendre compte de la totalité des données protéomiques récoltées, étant donné leur complexité. Pour aller plus loin dans l'analyse comparée des sécrétomes, il est possible d'utiliser d'autres outils simples, tels que les diagrammes de Venn. En classant les 501 protéines identifiées selon l'inducteur sur lequel elles ont

été produites (Figure 50A), on remarque que 195 d'entre elles, soit près de 40%, sont spécifiques d'un inducteur : 46 protéines sont présentes uniquement dans les sécrétomes produits sur pulpe de betterave, 72 dans ceux produits sur son de maïs, et 77 dans ceux produits sur Avicel. 176 protéines sont communes aux trois inducteurs, et 130 protéines sont communes à deux inducteurs seulement, ASM et SBP étant ceux qui ont le plus de protéines en commun. Le même type de diagramme peut être réalisé en prenant en compte non plus les inducteurs mais les souches utilisées (Figure 50B). Seules 37 protéines sont communes aux 5 souches, alors que 221 sont spécifiques d'une seule souche. La souche qui se démarque le plus des autres est logiquement la seule souche d'*Aspergillus niger* (BRFM 430), qui sécrète 73 protéines n'étant pas produites par les souches d'*Aspergillus japonicus*, alors que celles-ci produisent chacune entre 33 et 40 protéines de manière spécifique. On remarque également que 50 protéines sont communes uniquement aux quatre souches d'*Aspergillus japonicus*, ce qui est le plus grand nombre de protéines partagées par quatre souches. Les souches ayant le plus de protéines en commun sont BRFM 405 et BRFM 1490, avec 168 protéines communes au total, et 33 protéines retrouvées uniquement dans les sécrétomes produits par ces deux souches. Enfin, bien que la représentation graphique ne soit pas aisée, il est possible de réaliser le même type d'analyse pour les 15 sécrétomes (voir Annexes). Cela permet par exemple d'identifier que seules deux protéines sont présentes dans tous les sécrétomes ; il s'agit d'après leurs annotation d'une GH47 (alpha mannosidase putative) et d'une GH18 (chitinase putative). Plusieurs protéines sont au contraire présentes dans un seul sécrétome (157 au total), et tous les sécrétomes contiennent au moins une de ces protéines « spécifiques » : le sécrétome qui en contient le plus est celui produit par la souche BRFM 1489 sur Avicel, avec 20 protéines, et celui qui en contient le moins est celui produit par la même souche sur SBP, avec une seule protéine.

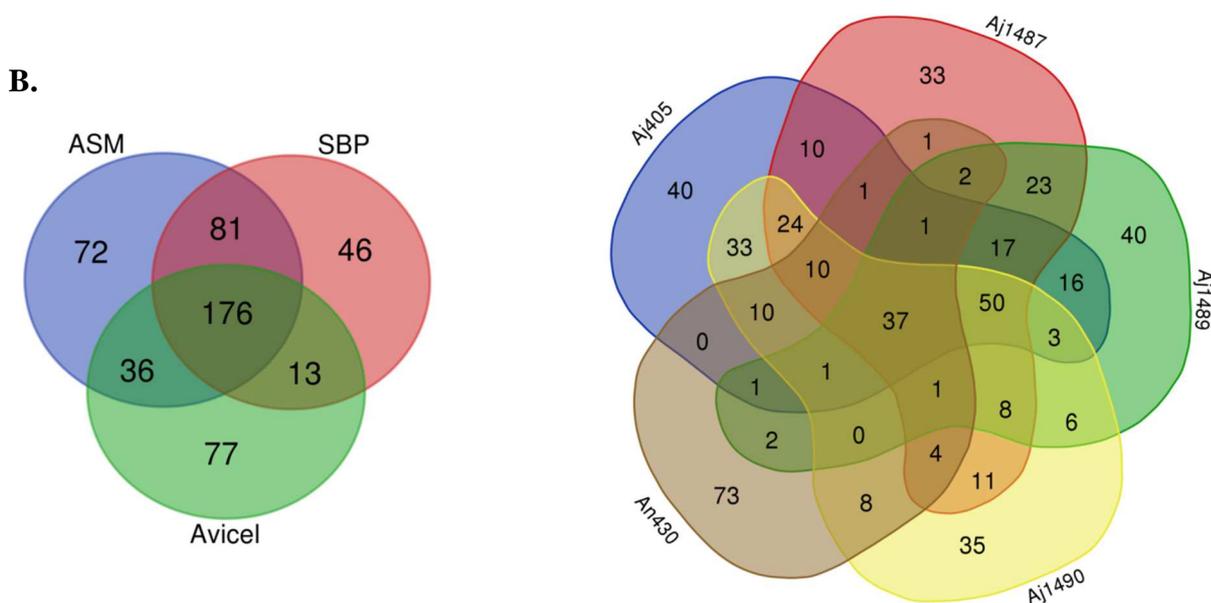


Figure 50 : Diagrammes de Venn représentant le nombre de protéines identifiées dans chaque type de sécrétome, classés selon **A.** l'inducteur sur lequel ils ont été produits ; **B.** les souches les ayant produits.

En plus de recenser les protéines présentes dans chaque sécrétome, les résultats de l'analyse LC-MS/MS permettent également d'avoir une indication de leur abondance. La quantification absolue des protéines n'est pas aisée, mais elle peut être approchée par exemple grâce au PAI (*Protein Abundance Index*), ou à l'emPAI (*exponentially modified PAI*) (Ishihama et al., 2005). Le PAI calcule pour chaque protéine identifiée le nombre de peptides détectés, normalisé par le nombre de peptides théoriquement observables après trypsinolyse ; cet indice permet de comparer l'abondance des différentes protéines dans un même sécrétome, ainsi que de comparer l'abondance relative d'une même protéine dans plusieurs sécrétomes, qui ont tous été analysés à quantités totales de protéines égales. L'emPAI, défini par la relation $emPAI = 10^{PAI} - 1$, est supposé être directement proportionnel à la concentration en protéines, et permet de calculer des fractions molaires ou massiques. Quelle que soit la méthode de quantification choisie, il est préférable d'effectuer l'analyse protéomique sur plusieurs réplicats. Cependant, dans notre cas, le nombre d'échantillons à analyser était déjà important, et la volonté d'avoir accès à un moyen de comparer rapidement les sécrétomes entre eux plutôt que d'entrer dans le détail de la composition de chacun était assumée. Afin de limiter le coût et le temps d'analyse, il a donc été décidé d'effectuer l'analyse sur un seul échantillon de chaque sécrétome.

La quantification a été utilisée notamment pour les CAZymes présentes dans les sécrétomes, qui appartiennent d'après l'annotation réalisée à 77 familles de modules catalytiques (GH, GT, CE, PL et AA), associées à 11 familles de CBM et à des protéines de type expansine. Ces données ont permis de réaliser des classifications ascendantes hiérarchiques, sur les familles représentées et sur les sécrétomes, aboutissant à la *heat-map* présentée en Figure 51. La classification des sécrétomes en 6 groupes fait apparaître des similitudes entre sécrétomes issus d'une même souche : les trois sécrétomes de la souche BRFM 1490 forment un groupe, et les sécrétomes produits sur ASM et SBP par les souches BRFM 405 et 430 en forment deux autres. D'autres regroupements se font plutôt sur la base de l'inducteur utilisé : les sécrétomes produits sur Avicel par les souches BRFM 405, 1487 et 1489 sont ainsi rassemblés. Enfin, les sécrétomes produits sur Avicel par la souche BRFM 430 et sur SBP par la souche BRFM 1490 se retrouvent dans un même groupe, probablement du fait de leur faible diversité de protéines identifiées, puisque ce sont les seuls à compter moins de 50 protéines différentes (voir Figure 28). Si le regroupement des familles de CAZymes est plus difficile à interpréter du fait de leur nombre très important, on peut tout de même s'intéresser au groupe contenant la famille AA16 et d'autres familles d'enzymes sécrétées en grandes quantités par la souche BRFM 1490 sur ASM et SBP. Ce groupe contient les familles GH10 et GH74, toutes les deux associées à des CBM1, ainsi que GH5_7, qui regroupent respectivement des xylanases, des xyloglucanases et des mannanases. On note aussi la présence des familles GH28, CE12 et PL4_1, actives sur les pectines. Enfin, les LPMO de la famille AA9 appartiennent également à ce groupe, de même que les enzymes de la famille AA3_2, qui contient plusieurs types d'enzymes pouvant agir comme donneurs d'électrons directs ou indirects pour les LPMO.

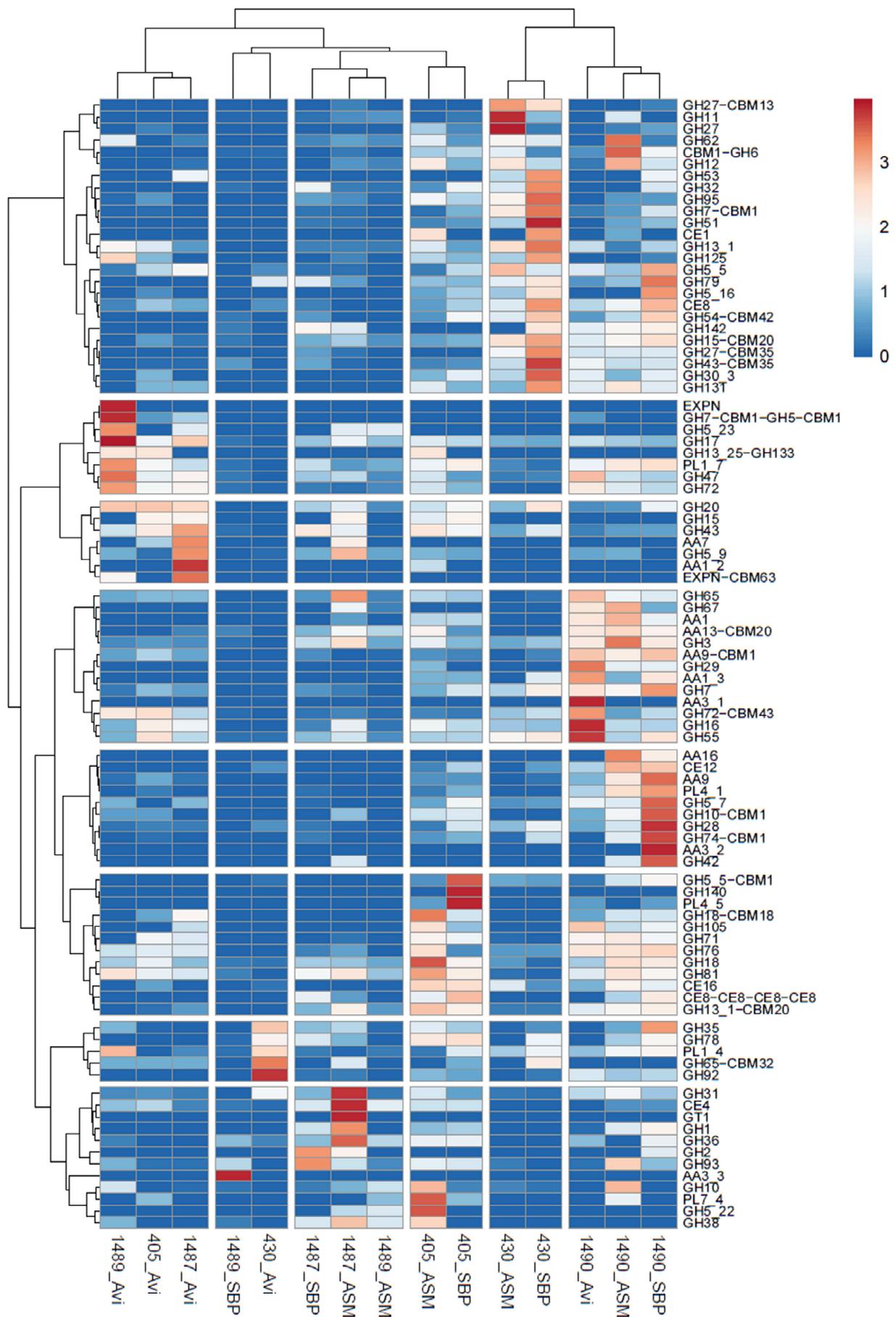


Figure 51 : Double classification hiérarchique des familles de CAZymes identifiées dans les sécrétomes étudiés, en fonction de leur abondance. L'échelle des couleurs est basée sur les valeurs de PAI après normalisation de la variance ; le bleu foncé correspondant aux protéines absentes, et le rouge aux protéines les plus abondantes.

D'autres analyses et représentations, basées sur la présence de certaines protéines ou leur quantification, seraient possibles. Cependant, les quelques analyses présentées permettent déjà de confirmer que le premier objectif du projet est rempli, puisque les cocktails enzymatiques produits présentent une importante diversité de compositions. L'étape suivante consistait à mettre au point un test permettant de comparer les performances des différents sécrétomes en déterminant lesquels ont le plus de potentiel pour l'amélioration des rendements d'hydrolyse enzymatique.

1.1.2 Tests de performances de saccharification des sécrétomes

Les conditions du test ont été soigneusement déterminées, afin de se rapprocher de celles envisagées dans les procédés industriels de bioconversion de la biomasse lignocellulosique, tout en facilitant l'analyse des données. Trois biomasses modèles, d'origines et de compositions différentes, ont été choisies pour cette étude : la paille de blé, le Miscanthus et le peuplier. Ces substrats, issus de déchets agricoles et de cultures énergétiques dédiées, sont fréquemment envisagés comme matières premières pour la production de bioéthanol en Europe. Ils ont été prétraités par explosion à la vapeur en présence d'acide sulfurique dilué ; après lavage et séchage, leur composition en sucres a été déterminée par CPG après hydrolyse acide (voir Tableau 7).

Tableau 7 : Conditions de prétraitement et composition des trois biomasses modèles utilisées

	Peuplier	Miscanthus	Paille de blé
Composition en % du poids sec			
Rhamnose	0,3	0,4	0,2
Arabinose	0,1	0,4	0,5
Xylose	0,6	2,1	1,6
Mannose	0,2	0	0,2
Glucose total	53,3	59,4	53,9
dont glucose cellulosique	50,8	57,0	51,3
Teneur en eau (%m/m)	6,3%	5,7%	7,0%

Le cocktail cellulolytique de *T. reesei* utilisé (noté K975) est issu de la souche CL847 cultivée en présence de lactose. Ce cocktail a une activité β -glucosidase assez faible, mais il s'agit d'un problème bien connu des cocktails de *T. reesei*, et des solutions techniques existent déjà pour y remédier (Ayrinhac et al., 2011). Afin de s'assurer que les potentielles améliorations de performances lors de la supplémentation ne soient pas dues à une simple augmentation de l'activité β -glucosidase, une préparation commerciale de β -glucosidases d'*A. niger* a été ajoutée en excès au cocktail K975. Ce cocktail de référence a été utilisé pour réaliser la saccharification des trois biomasses choisies, avec une concentration en biomasse relativement élevée (5% m/v) afin de se rapprocher de conditions réalistes du point de vue du procédé où les concentrations sont supérieures à 15%, tout en gardant des concentrations envisageables pour des tests de faibles volumes. Une dose de cocktail enzymatique assez faible a été choisie (5 mg/g de matière sèche), correspondant pour

l'industrie à des conditions permettant de réduire les coûts et pour lesquelles la marge d'amélioration des performances est élevée. La cinétique sur 5 jours de cette hydrolyse, présentée sur la Figure 26, a permis de déterminer les principaux objectifs d'amélioration pour chacune des biomasses. Dans les conditions choisies, le rendement final d'hydrolyse de la paille de blé est proche de 100%, ce qui signifie que la quasi-totalité de la cellulose a été transformée en monomères de glucose ; l'objectif dans ce cas sera donc d'accélérer la réaction au début de l'hydrolyse (à 24 et 48h). Le *Miscanthus* et le peuplier sont plus récalcitrants puisque dans les mêmes conditions ils ne sont hydrolysés qu'à 50 et 60% au maximum ; il sera donc possible de chercher à améliorer à la fois la cinétique initiale et le rendement après une période plus longue (96h par exemple).

Le cocktail de référence a ensuite été supplémenté par les différents secrétomes d'*Aspergillus*. Dans un souci de simplification expérimentale, à une quantité fixe de cocktail K975+SP188, des volumes égaux de chacun des secrétomes concentrés ont été ajoutés, comme cela avait été fait pour le criblage des secrétomes issus de la biodiversité fongique (Berrin et al., 2012). La quantité de protéines ajoutées n'est pas la même dans tous les échantillons, mais elle reflète les quantités initiales de protéines sécrétées par les souches fongiques sur chaque milieu de culture. Les améliorations observées seront donc plus qualitatives que quantitatives, mais cela s'intègre bien dans l'approche de criblage désirée.

Les essais ont initialement été réalisés en tubes, en triplicats, mais les biomasses utilisées étant hétérogènes, il est apparu que la variabilité des résultats était importante, et qu'il serait bénéfique de raisonner sur un nombre plus important de réplicats. Afin de pouvoir multiplier le nombre de réplicats sans trop augmenter le temps de manipulation et les quantités de secrétomes utilisées, les tests de saccharification ont donc été miniaturisés en microplaques et automatisés à l'aide d'un robot Tecan, en s'inspirant d'une méthode déjà développée au laboratoire BBF de l'INRA (Navarro et al., 2010). Il a fallu pour cela broyer finement la biomasse et réaliser une suspension suffisamment homogène pour pouvoir être pipetée en microplaques, puis utiliser le robot pour l'ajout du cocktail de référence et des différents secrétomes. Le volume réactionnel de chaque échantillon a ainsi été divisé par dix, et chaque test compte au minimum 7 répétitions d'une même condition. L'utilisation de suspension de biomasse et la diminution du volume réactionnel ont généralement eu pour effet d'augmenter les écart-types observés par rapport aux essais en tubes, mais cette technique a permis d'effectuer suffisamment de répétitions de chaque condition pour envisager un traitement statistique (40 à 60 réplicats pour les conditions références, 7 à 15 réplicats pour chaque condition de supplémentation). Les résultats présentés sur la Figure 27 ont ainsi été obtenus en calculant pour chaque condition la moyennes des pourcentages d'amélioration par rapport à la référence, et en soumettant les séries à une analyse statistique (test de Student) afin d'identifier les différences statistiquement significatives.

1.1.3 Stratégies pour le choix des cibles enzymatiques

Le traitement statistique utilisé permet, pour chaque biomasse et à chaque temps de réaction, d'identifier les sécrétomes capables d'apporter une amélioration de la saccharification. Ce sont ces sécrétomes qui peuvent être utilisés pour identifier des enzymes potentiellement capables de compléter le cocktail de *T. reesei*. L'analyse manuelle des protéines de fonctions inconnues dans certains de ces sécrétomes a mené à la découverte d'un membre d'une potentielle nouvelle famille de LPMO, nommée par la suite AA16, comme expliqué au chapitre III. Etant donné le fort impact potentiel des LPMO pour les biotechnologies et l'intérêt qui leur est porté depuis leur découverte, il a rapidement été décidé que ces enzymes feraient partie des cibles principales du projet, et au vu du temps et des efforts nécessaires à leur production et à leur caractérisation (voir discussion plus bas), les autres cibles potentielles ont été laissées de côté pour se consacrer totalement à l'étude des AA16. Cependant, d'autres cibles, identifiées par une analyse méthodique des données, pourraient faire l'objet d'une étude plus approfondie.

Une première approche consiste à lister les protéines plus abondantes dans les sécrétomes ayant de bonnes performances que dans les sécrétomes n'ayant pas apporté d'amélioration. Pour cela, une quantification approximative des protéines introduites lors des tests de saccharification a été réalisée en utilisant le PAI fourni par l'analyse LC-MS/MS pour chaque protéine et les concentrations en protéines totales déterminées pour chaque sécrétome. Les comparaisons ont ensuite été faites pour chaque biomasse et à chaque temps de réaction. Un seul sécrétome ayant été capable d'améliorer la saccharification du peuplier (405_ASM), il n'a pas été possible de réaliser de véritable étude comparative pour cette biomasse ; il y a plus d'une cinquantaine de protéines plus abondantes dans ce sécrétome que dans tous les autres, mais sans pouvoir les comparer à celles d'un autre sécrétome performant, il est impossible d'émettre des hypothèses sur les enzymes potentiellement intéressantes parmi elles. Pour la paille, trois sécrétomes en tout ont montré de bonnes performances, dont deux dès 24h de réaction (405_ASM et 1487_SBP) et un supplémentaire après 48h (430_Avi). Ces trois cocktails ont été produits par trois souches différentes, sur trois inducteurs différents, et ont seulement 23 protéines en commun ; aucune n'est présente dans ces trois cocktails en plus grandes quantités que dans les autres sécrétomes n'ayant pas amélioré l'hydrolyse. Il est donc probable que les protéines impliquées dans la supplémentation du cocktail de *T. reesei* ne soient pas les mêmes dans ces trois sécrétomes. En revanche, les deux sécrétomes actifs à 24h partagent plus de similarités ; quatre protéines, listées dans le Tableau 8, y sont particulièrement abondantes comparé aux sécrétomes moins performants. Parmi elles se trouve une enzyme de la famille GH95, dont les membres caractérisés sont des α -L-fucosidases ou des α -L-galactosidases.

Tableau 8 : Protéines plus abondantes dans les sécrétomes améliorant la saccharification de la paille à 24h (405_ASM et 1487_SBP) que dans les 13 autres. Les protéines sont repérées par le numéro d'accèsion NCBI de leur orthologue dans le génome d'*Aspergillus aculeatus* qui a servi à l'annotation.

N° accession NCBI	Annotation / Fonction prédite
XP_020057350.1	Uroporphyrinogen decarboxylase
XP_020056446.1	GH95
XP_020055920.1	Cystéine hydrolase
XP_020054543.1	DHDPS-like (Dihydrodipicolinate synthase)

La biomasse sur laquelle l'analyse comparative semble la plus prometteuse est le *Miscanthus* : huit sécrétomes, produits sur ASM et SBP, sont capables d'améliorer le rendement de saccharification dès 24h, dont cinq également après 96h ; deux autres, produits sur Avicel, ne sont actifs qu'à 96h. Comme pour le cas de la paille à 48h, les sept sécrétomes ayant de bonnes performances à 96h incluent des sécrétomes produits sur Avicel, et n'ont que très peu de protéines en commun (7 en tout). En revanche, la comparaison des sécrétomes actifs dès 24h révèle que 10 protéines sont présentes dans ces 8 sécrétomes en quantités plus importantes que dans les 5 sécrétomes n'apportant aucune amélioration (voir Tableau 9). A une exception près, ces protéines sont toutes des CAZymes, potentiellement actives sur la pectine, l'amidon et les β -1,3-glucanes (de la paroi végétale et/ou fongique). On remarque surtout la présence dans cette liste d'une CBH (GH7) et d'une EG (GH5) ; il est tout à fait probable que ce soit leur présence en début de saccharification, en plus des cellulases de *T. reesei*, qui ait eu un impact positif sur la libération de sucres.

Tableau 9 : Protéines plus abondantes dans les sécrétomes améliorant la saccharification du *Miscanthus* à 24h (405_ASM, 405_SBP, 1487_ASM, 1487_SBP, 1490_ASM, 1490_SBP, 430_ASM, 430_SBP) que dans ceux n'apportant aucune amélioration (1487_Avi, 1489_ASM, 1489_SBP, 1489_Avi, 1490_Avi). Les protéines sont repérées par le numéro NCBI de leur orthologue dans le génome d'*Aspergillus aculeatus* qui a servi à l'annotation.

N° accession NCBI	Annotation CAZy	Fonction prédite
XP_020053500.1	GH15-CBM20	Glucoamylase
XP_020055541.1	GH7-CBM1	Probable 1,4- β -D-glucan cellobiohydrolase
XP_020056433.1	GH17	Probable glucan endo- β -1,3-glucosidase
XP_020060749.1	GH55	Hypothetical glucan β -1,3-glucosidase
XP_020052817.1	GH72-CBM43	β -1,3-glucanosyltransferase
XP_020060420.1	GH13_1	Alpha-amylase
XP_020060878.1	GH16	Fungal CRH1 transglycosylase
XP_020059301.1		Lactonohydrolase
XP_020059366.1	GH5_5	Probable endo- β -1,4-glucanase B
XP_020058447.1	GH28	Exo-polygalacturonase

On peut voir dans ce résultat l'une des limitations de la méthode choisie consistant à rajouter des cocktails protéiques à une quantité constante d'enzymes de *T. reesei*, aboutissant à une charge enzymatique supérieure à celle de la référence au lieu de viser une charge totale d'enzymes (cocktail de *T. reesei* + sécrétome) égale à celle de la référence. Le risque était ainsi plus important d'identifier comme performants des sécrétomes contenant des enzymes déjà produites par *T. reesei*, notamment des cellulases. L'une des possibilités pour contrer cet effet est de tester les performances de saccharification des sécrétomes seuls (Figure 52). La comparaison des 8 sécrétomes précédemment cités pour l'hydrolyse de *Miscanthus* pendant 24h, confirme en effet que certains sont capables de dégrader par eux-mêmes une partie de la biomasse (405_SBP, 430_ASM, 430_SBP, 1490_SBP). Les autres sécrétomes, ayant de faibles capacités de dégradation lorsqu'ils sont utilisés seuls mais capable d'améliorer significativement le rendement de saccharification du cocktail de référence, présentent donc plus d'intérêt pour la recherche d'activités complémentaires à celle des cellulases de *T. reesei*. Ces résultats pourraient donc être combinés avec ceux des autres tests de supplémentation, afin d'établir une nouvelle liste des sécrétomes les plus performants, et de réaliser une nouvelle comparaison pouvant aboutir à une nouvelle liste de protéines cibles.

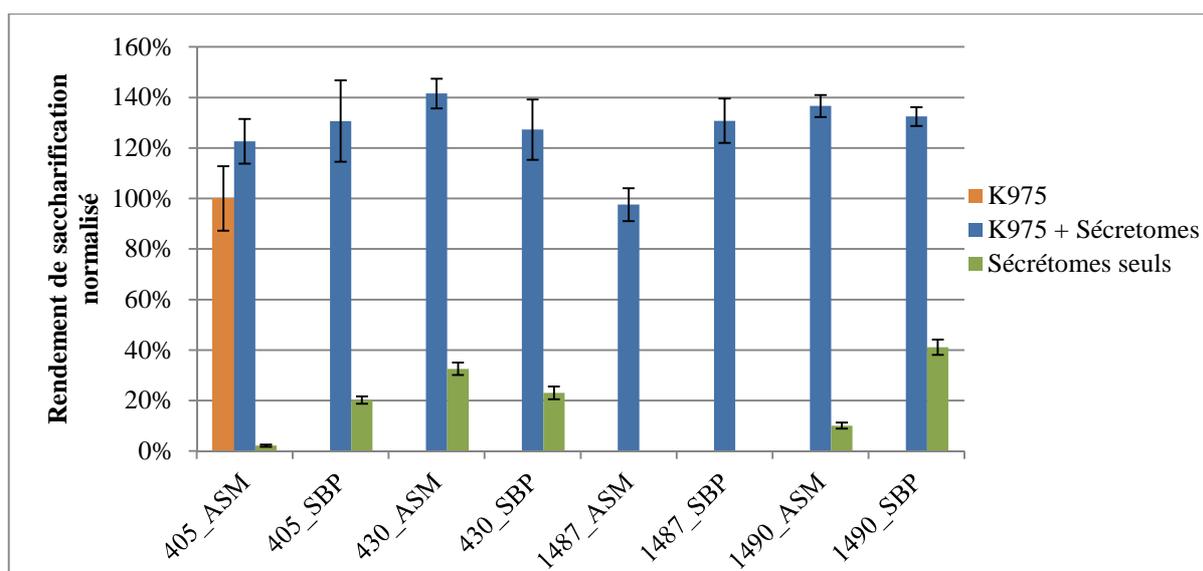


Figure 52 : Performances de saccharification de *Miscanthus* prétraité par des sécrétomes seuls et en supplémentation du cocktail de référence de *T. reesei*. La hauteur des barres représente les rendements de conversion de la cellulose obtenus après 24h, normalisés par rapport au rendement obtenu avec le cocktail de référence seul.

L'approche de comparaison « protéine par protéine » présentée plus haut, en plus d'être fastidieuse, fait ressortir seulement des protéines isolées dont la présence en grande quantité serait associée à de meilleures performances ; or il est possible que l'absence de certaines protéines et/ou l'association de plusieurs protéines puissent également avoir un impact positif sur la saccharification. Afin d'avoir une vision plus générale des possibles corrélations entre les profils protéiques et les performances des sécrétomes, il serait possible de réaliser une régression des moindres carrés (ou régression PLS, pour *partial least squares*), en s'inspirant de

l'approche choisie dans une étude de Lehmann et al. (2016) qui leur avait permis d'identifier dans des cocktails de *T. reesei*, produits dans différentes conditions, les protéines dont la concentration était corrélée avec les performances des cocktails sur canne de maïs prétraitée, indiquant qu'il pouvait s'agir d'enzymes capables d'améliorer ou de limiter l'hydrolyse de la biomasse étudiée.

Les nombreuses données générées lors de la première partie du projet pourraient donc faire l'objet d'une exploration et d'une analyse plus approfondie, et d'autres enzymes pourraient être étudiées par la suite. Pour ce projet de thèse cependant, comme expliqué plus haut, il a été décidé de se consacrer plus particulièrement à la production et à la caractérisation des enzymes de la famille AA16.

1.2 Production et purification des LPMO AA16

Pour pouvoir caractériser ces nouvelles enzymes, il a d'abord été nécessaire de les produire avec une pureté satisfaisante et en quantités suffisantes. Or l'obtention de protéines AA16 intactes et actives a été plus délicate que pour d'autres LPMO. La première protéine cible, AaAA16, a été exprimée de manière hétérologue dans *Pichia pastoris* en utilisant d'abord un protocole adapté de celui préconisé par Invitrogen, développé au laboratoire BBF et utilisé en routine pour la production de très nombreuses enzymes fongiques, notamment des LPMO des familles AA9 et AA14 (Bennati-Granier et al., 2015; Couturier et al., 2018; Haon et al., 2015). Après une première production en plaques Deepwell visant à sélectionner le clone permettant la meilleure expression, la protéine a été produite en fioles dans un volume de 200 mL, puis purifiée par chromatographie d'affinité sur colonne de nickel grâce à sa queue poly-histidine, et enfin chargée en cuivre. Cette 1^{ère} production a conduit à une enzyme ayant une activité de dégradation de la cellulose, qui a servi à réaliser quelques tests d'activité préliminaires, mais lorsque d'autres productions ont été réalisées selon le même protocole pour obtenir de plus grandes quantités de protéines, celles-ci n'avaient aucune activité sur la PASC. Le séquençage N-terminal de ces différentes protéines produites en fioles a révélé que l'histidine N-terminale était majoritairement clivée ou modifiée (voir paragraphe 3.4 du chapitre III) ; il a été prouvé que le site catalytiques des LPMO est particulièrement sensible aux dommages oxydatifs, notamment en conditions réductrices et en l'absence de leur substrat, et au moins une étude a déjà montré que de tels dommages pouvaient se produire durant leur expression hétérologue (Kadowaki et al., 2018). En parallèle des cultures en fiole, un test de production de AaAA16 en bioréacteur a été initié. Le surnageant issu de cette culture en bioréacteur était actif sur PASC, mais une grande partie de l'activité a été perdue après que la protéine a été purifiée en utilisant le protocole standard de chromatographie d'affinité sur ions métalliques immobilisés (IMAC), comme montré sur la Figure 32. Une autre technique de purification, la chromatographie par échange d'anion, a donc été testée, conduisant à une protéine purifiée active, sans avoir besoin d'être chargée en cuivre ; le cuivre contenu dans le milieu de culture a en effet été suffisant pour charger tous les sites actifs des protéines, comme confirmé

par spectrométrie de masse à plasma à couplage inductif (ICP-MS). Par ailleurs, le séquençage N-terminal de la protéine purifiée a montré que le résidu N-terminal n'était clivé que dans 10% de l'échantillon, et que lorsqu'il était présent il s'agissait bien d'une histidine intacte.

L'association d'une production en bioréacteur et d'une purification par chromatographie d'échange d'anion a donc permis l'obtention d'une protéine intacte en quantités importantes, correctement chargée en cuivre et active sur cellulose. Cependant, plusieurs phénomènes restent en partie incompris, notamment le fait que la purification par IMAC ait eu un tel impact sur l'activité de la protéine. Cette méthode n'est pas la plus utilisée pour la purification des LPMO, principalement parce que la présence d'une étiquette poly-histidine est susceptible par la suite de perturber la liaison de l'enzyme à son substrat, et de compliquer les tests d'activité en raison de son affinité pour les ions métalliques, dont le cuivre (Bjølgerud; Eijsink et al., 2019) ; il faut à ce sujet garder à l'esprit que toutes les LPMO produites lors de ce projet de thèse, même lorsqu'elles ont été purifiées par échange d'anions, restent porteuses de cette étiquette poly-histidine C-terminale. La perte d'activité constatée ici n'est cependant pas liée à la présence que cette étiquette. Elle n'est pas non plus liée au changement de pH réalisé avant la purification, ni au contact avec de l'imidazole concentrée lors de l'élution, comme confirmé par des tests d'activité reproduisant ces conditions ; cette perte d'activité est donc bien due au passage sur la colonne d'agarose greffée nickel-NTA (acide nitrilotriacétique). Il serait intéressant de voir si d'autres LPMO, d'autres familles ou produites par d'autres groupes, subissent le même effet. Le fait que la production en bioréacteur ait été plus favorable à la conservation de l'histidine N-terminale que la production en fioles est également difficile à expliquer, étant donné le nombre de paramètres qui varient entre ces deux types de production : la fermentation en bioréacteur permet d'atteindre des densités de cellules élevées, et d'augmenter ainsi la concentration de protéines produites ; les paramètres tels que le pH, la température et l'apport en oxygène sont par ailleurs contrôlés précisément, ce qui n'est pas le cas pour les productions en fioles.

L'importance du contrôle des conditions de fermentation est soulignée par les résultats obtenus lors de la production en bioréacteurs des trois autres AA16. Cette production a été réalisée avec des équipements différents de celui utilisé pour la production de AaAA16, et les paramètres utilisés pour le premier bioréacteur n'ont pas pu être transposés tels quels. Notamment, l'agitation et/ou le flux d'air ayant été insuffisants en fin de fermentation, la concentration de dioxygène dissout dans le milieu est descendue en dessous des 20% généralement préconisés ; la métabolisation du méthanol par les levures a ainsi pu être ralentie, conduisant notamment à des concentrations en protéines plus faibles qu'espérées. Par ailleurs, les trois protéines exprimées ont subi des clivages de leurs résidus N-terminaux, ainsi que des probables modifications de l'histidine N-terminale lorsqu'elle était présente (voir le paragraphe 3.1 du chapitre IV). Contrairement aux séquençages réalisés sur les AaAA16 produites en fiole, où l'acide aminé modifié a pu être identifié comme étant probablement un résidu aspartate, les modifications subies par l'histidine n'ont cette fois pas pu être

identifiées précisément, soit parce que le pic correspondant était confondu avec le pic d'injection, soit parce qu'il est sorti au-delà de la fenêtre d'analyse. On pouvait s'attendre à ce que les sites catalytiques comportant une histidine modifiée ne soient pas capables de coordonner des ions cuivre, et à ce que les protéines soient inactives. S'il est vrai que les deux protéines ApAA16 et Pf AA16 produisent très peu de sucres oxydés, on remarque tout de même que AjAA16 a une activité sur PASC supérieure à celle de AaAA16, alors que cette dernière a un site catalytique majoritairement intact. On peut donc se demander quelle est la véritable nature de ces modifications et quel est leur véritable impact sur le site actif. L'analyse par spectroscopie RPE de la protéine ApAA16 semble en tout cas montrer que le centre actif est bien capable de fixer un ion cuivre (II), avec une géométrie similaire à celle retrouvée dans le site actif des autres LPMO étudiées jusqu'à présent.

1.3 Particularités de l'activité des AA16

Comme cela a été mis en évidence par de nombreux articles portant sur l'étude de diverses LPMO, et souligné par une revue récente (Eijsink et al., 2019), la caractérisation de ces enzymes peut se révéler particulièrement difficile, notamment lorsqu'il s'agit d'évaluer leur activité sur des polysaccharides complexes. L'une des particularités des essais enzymatiques en présence des LPMO, au-delà du choix du substrat et des considérations habituelles (température, pH...), est la prise en compte de leur état de réduction et de son influence sur leur activité et leur stabilité. En réalisant des tests d'activité des AA16 sur cellulose dans différents tampons, en présence ou non de réducteurs, il est rapidement apparu que l'activité observée était identique en l'absence ou en présence de donneurs d'électrons tels que l'ascorbate ou la cystéine. Lorsque les mêmes tests ont été réalisés dans l'eau, cet effet n'a pas été observé : très peu de produits solubles étaient détectés en l'absence de réducteur, et l'ajout de celui-ci conduisait à l'activation de l'enzyme, visible grâce à l'apparition de certains produits de dégradation. Une étude plus élargie, menée dans 6 tampons différents et sur une large gamme de pH allant de 3,5 à 8,5 (16 conditions différentes en tout), a révélé que ce phénomène se produisait dans tous les cas, sauf lorsque la réaction était réalisée dans l'eau (Figure 53). Tout se passe donc comme si ces différents tampons, bien qu'ils ne comportent pas tous de composés réducteurs, étaient tous capables de servir de donneur d'électrons et que l'ajout d'un autre réducteur était alors superflu.

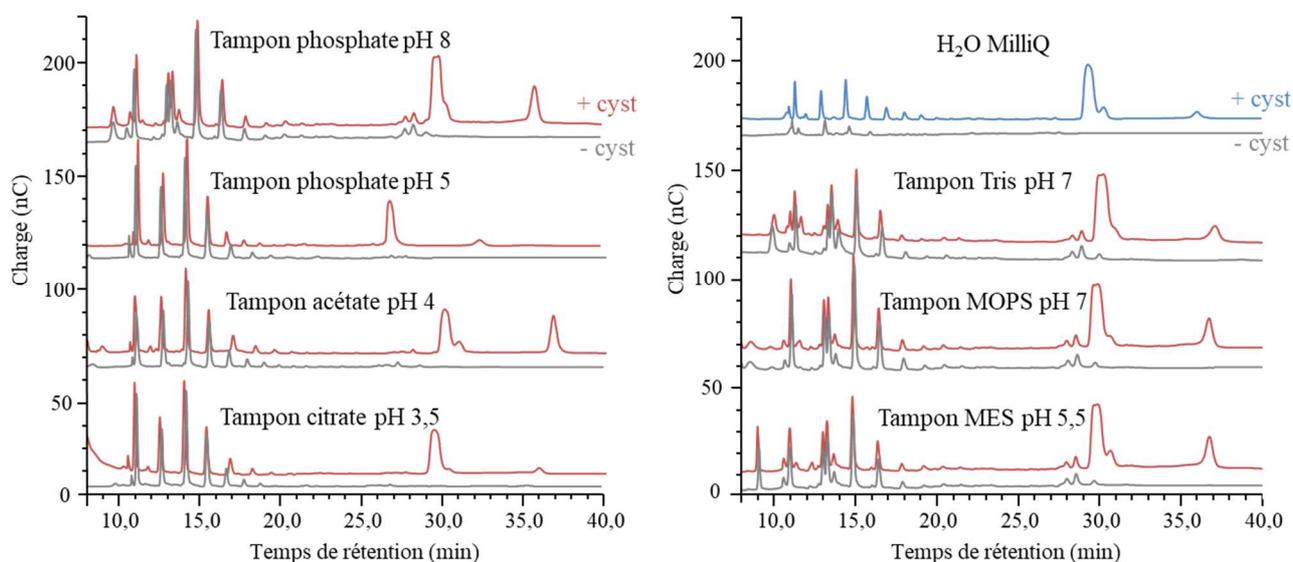


Figure 53 : Effet de différents tampons et pH sur l'activité de la LPMO AaAA16 (2 μ M), en présence ou non de cystéine.

Les chromatogrammes HPAEC-PAD montrent les produits solubles issus de la dégradation de PASC après 15 minutes de réaction ; pour chaque condition (eau ou tampons citrate de sodium, acétate de sodium, phosphate de potassium, Tris, MOPS, MES), le chromatogramme gris correspond à la condition sans cystéine, et le chromatogramme rouge ou bleu à la condition en présence de cystéine (1 mM).

Afin d'éviter que cet effet ne perturbe les réactions, tous les tests ont été réalisés dans l'eau MilliQ ; les enzymes et les substrats ont été préparés et conservés dans l'eau. Ces précautions ont permis d'observer les véritables effets de l'ajout de réducteurs, et notamment de différencier leurs capacités à activer les AA16. Plusieurs des petites molécules testées (acide tannique, pyrogallol, ménadione), ainsi que la cellobiose déshydrogénase de *Podospora anserina*, n'ont pas semblé pouvoir servir de donneurs d'électrons pour AaAA16. Les composés qui en sont capables semblent tous avoir la même efficacité puisque les profils de sucres produits sont quasiment identiques, et aucun ne provoque la production de produits C1-oxydés en grande quantité. Les premières expériences de spectroscopie RPE ayant révélé que l'ascorbate n'avait pas pu réduire entièrement le cuivre du site actif d'ApAA16 même après plusieurs heures d'incubation, il est probable que le même effet soit observé avec les autres réducteurs testés. Les spectres RPE de la protéine placée dans l'eau et dans un tampon acétate de sodium (pH 5.2) ont été comparés afin d'essayer d'expliquer le phénomène décrit plus haut ; la présence du tampon n'a pas fait décroître le signal du Cu(II), mais une modification de la valeur du paramètre $|A_z|$, passant de 495 à 480 MHz, a été observée. Cela pourrait être interprété comme une modification de l'environnement du cuivre, mais il faudrait vérifier que la différence observée est significative en répétant les expériences. La spectroscopie RPE est un bon moyen d'accéder à des informations sur l'état de réduction du site actif des LPMO ; afin de mieux comprendre l'effet des tampons et réducteurs sur ce site actif, il serait intéressant de faire de nouvelles expériences, impliquant différents réducteurs (potentiellement plus forts que ceux déjà testés) et différentes enzymes, chargées en cuivre ou non (apoenzyme vs. holoenzyme). De plus, la production de protéines AA16 ne portant pas d'étiquette N-

terminale serait utile pour vérifier qu'une partie du signal observé ne soit pas dû à des ions cuivre fixés sur la queue histidine des protéines au lieu du site actif.

L'absence de réduction correcte du cuivre dans le site actif serait une hypothèse possible pour expliquer la faible quantité de sucres oxydés observés ; cependant, elle n'explique pas l'apparition massive de produits de dégradation non-oxydés. Il est connu que les cello-oligosaccharides oxydés en C4 sont instables en conditions alcalines ; l'unité glucose oxydée peut en effet être clivée en partie lors du passage dans les colonnes de chromatographie, ce qui conduit à l'observation de pics important de produits non-oxydés (Westereng et al., 2016). Il serait donc possible que les AA16 étudiées aient une activité d'oxydation en C4 expliquant leur profil inhabituel. Cependant, aucune trace de pics dans les régions caractéristiques des produits C4 ou C1/C4 n'a été observée sur les chromatogrammes HPAEC-PAD, même lorsque la méthode d'arrêt de la réaction a été modifiée pour passer de l'ébullition à l'ajout de soude et à la filtration, afin d'éviter la dégradation d'éventuels produits oxydés en C4, qui sont également sensibles aux hautes températures. L'analyse par spectrométrie de masse n'a pas non plus détecté la présence de produits oxydés autres qu'en C1. Ces deux méthodes, qui sont les plus utilisées et celles qui donnent le plus d'informations sur la nature des produits, ont cependant l'inconvénient de n'identifier que les sucres solubles issus de l'activité des LPMO ; les produits ayant un degré de polymérisation supérieur à 8 ne sont donc pas visualisés. Afin de compléter la vision que nous avons de l'activité des AA16, et de détecter la présence de produits oxydés restés en phase solide ou ayant été réadsorbés, il serait donc intéressant d'utiliser des techniques permettant de caractériser la partie insoluble du substrat. Plusieurs méthodes ont été développées dans ce sens ces dernières années (Eibinger et al., 2014; Vuong et al., 2017; Wang et al., 2018b).

La production importante de sucres natifs pourrait aussi être simplement due à une activité hydrolase, qui expliquerait également la présence de traces de produits natifs en l'absence de réducteur (voir Figure 36 et Figure 38 par exemple) ; l'ajout d'un réducteur permettrait alors de déclencher une faible activité LPMO, qui serait tout de même suffisante pour agir en synergie avec l'hydrolase et augmenter de façon spectaculaire la quantité de sucres natifs libérés, tandis que les sucres oxydés resteraient minoritaires. Cette hypothèse permettrait également d'expliquer pourquoi les AA16 sont capables de dégrader le celohexaose, sans produire de sucres oxydés détectables (voir Figure 35). Eijsink et al. (2019) ont suggéré que les préparations de LPMO pourraient fréquemment être contaminées par la présence de glycosyl-hydrolases tels que les cellulases, qui changeraient le profil des produits observés. C'est peut-être le cas pour les protéines exprimées chez *Trichoderma reesei*, *Aspergillus niger*, ou d'autres hôtes fongiques connus pour exprimer des cellulases ; cependant, dans le cas d'une production par *Pichia pastoris*, très peu de protéines de l'hôte sont sécrétées dans le surnageant de culture – 20 en tout en présence de glucose, dont quelques endo-1,3- β -glucanases mais pas de cellulases (Mattanovich et al., 2009). Par ailleurs, le système d'expression hétérologue de protéines chez *P. pastoris* a déjà été largement utilisé au laboratoire BBF pour produire des LPMO et autres CAZymes

recombinantes. Ainsi, si des cellulases étaient secrétées par *P. pastoris* dans les conditions utilisées, cette activité indésirable aurait probablement contaminé les autres productions réalisées ; le fait que certaines protéines produites au laboratoire ne montrent aucune activité sur certains substrats cellulotiques, comme par exemple les AA14 (Couturier et al., 2018), démontre que ce n'est pas le cas. De plus, pour que cette activité contaminante se retrouve dans toutes les fractions d'AA16 produites et testées, il faudrait que la protéine correspondante ait un comportement physico-chimique strictement identique à celui des LPMO ; plusieurs techniques de purification différentes ayant été testées, cela paraît très peu probable. Afin de pouvoir vérifier les hypothèses émises quant à l'origine de cette importante production de sucres natifs, il serait possible d'effectuer des tests en conditions anaérobies : la véritable activité LPMO sera inhibée par l'absence d'oxygène, et rétablie en présence de peroxyde d'hydrogène, tandis qu'une potentielle activité hydrolase sera maintenue malgré l'absence d'oxygène.

Bien que de nombreuses questions se posent encore sur l'activité des AA16, il n'en reste pas moins que leur activité de production de sucres C1-oxidés à partir de cellulose a été établie. La présence de modules de liaison à la cellulose dans un des sous-groupes de la famille laisse penser qu'il s'agit de l'un des substrats naturels de ces enzymes. Cependant, elles pourraient potentiellement avoir une spécificité de substrat beaucoup plus large, comme c'est le cas pour les AA9. D'autres substrats ont été testés, sans qu'aucune activité ne soit détectée : β -1,3/1,4-glucane d'orge, glucomannane de konjac, xylohexaose ; beaucoup n'ont pas été testés, comme la chitine, l'amidon, le xylane ou le xyloglucane. Des substrats plus complexes associant plusieurs types de polysaccharides pourraient également être envisagés, puisque certaines LPMO sont par exemple actives sur le xylane uniquement lorsqu'il est associé à la cellulose (Couturier et al., 2018; Frommhagen et al., 2015).

L'activité sur la cellulose a également été confirmée par le fait que les AA16 testées sont capables de modifier la cellulose, amorphe ou cristalline, de manière à améliorer sa dégradation par la CBHI de *T. reesei* (Figure 37 et Figure 45). Ces tests ont été réalisés à de fortes doses de LPMO, celle-ci ayant un rapport 10:1 avec la CBHI. Dans l'optique d'une utilisation dans des procédés biotechnologiques, il faudrait donc réussir à diminuer largement cette dose ; on pourrait imaginer y parvenir si des conditions permettant aux AA16 de produire plus de produits oxydés étaient déterminées. Par ailleurs, pour s'approcher de conditions d'utilisations réelles, il faudrait idéalement que cette synergie soit effective lorsque les deux enzymes sont utilisées au sein d'un même cocktail. Lors de tests préliminaires, l'ajout d'AaAA16 simultanément à la CBHI, toujours dans un rapport 10 :1, a mené à une diminution de la quantité de cellobiose libéré. Cet effet pourrait être dû à un excès de LPMO : celles-ci restant immobiles une fois fixées au substrat, pendant plusieurs secondes voire minutes (Eibinger et al., 2017), elles pourraient gêner l'accès des cellobiohydrolases si elles sont présentes en trop grandes quantités. Une diminution de leur concentration serait donc encore une fois souhaitable. Plusieurs études ont montré l'effet bénéfique de LPMO C1-spécifiques sur l'action des CBHI de *Trichoderma reesei* et *Trichoderma longibrachiatum* (Harris et al., 2010; Langston et al., 2011; Selig et al.,

2015; Song et al., 2018), mais il a également été démontré que dans certaines conditions leur présence pouvait inhiber les CBH1, et au contraire améliorer l'activité des CBHII (Zhou et al., 2019). Nous avons voulu tester cette hypothèse, mais dans les conditions utilisées pour nos tests (ajout séquentiel d'AA16 pendant 24h puis CBHI sur cellulose amorphe, dans un ratio 10:1), aucune des AA16 n'a été capable d'améliorer l'activité de la CBHII de *T. reesei*, ou d'un cocktail complet. D'autres tests seront donc nécessaires pour déterminer s'il existe des conditions dans lesquelles l'ajout d'AA16 a un impact positif sur la dégradation de la cellulose. En effet, de manière générale, l'effet des LPMO sur l'activité des CBH semble largement dépendre des conditions du test utilisé (ajout séquentiel ou simultané, temps d'incubation, ratio des enzymes), ainsi que des caractéristiques du substrat choisi et des CBH étudiées, puisque les CBHI sont plus actives que les CBHII sur les parties cristallines de la cellulose, et qu'elles sont strictement processives alors que les CBHII se déplacent aussi par des mouvements de diffusion (Eibinger et al., 2014; Zhou et al., 2019).

2 Conclusions et perspectives

Ce projet de thèse se proposait d'explorer les secrétomes de plusieurs souches du genre *Aspergillus* et d'y identifier des enzymes d'intérêt pour la saccharification de biomasse lignocellulosique. L'utilisation d'une analyse protéomique complète et d'une méthode de criblage des performances de saccharification a permis l'obtention de nombreuses données, qui n'ont pas toutes été exploitées et pourront faire l'objet d'une étude plus approfondie dans le futur.

Suite à l'identification de protéines appartenant à une nouvelle famille de LPMO, la suite des travaux a été focalisée sur la production et la caractérisation de ces cibles, permettant ainsi de démontrer leur activité sur les substrats cellulosiques, et leur intérêt pour l'amélioration des performances des cellulases de *T. reesei*. Cependant, toutes les protéines produites appartiennent à un même sous-groupe de la famille. Il serait donc intéressant de caractériser des enzymes appartenant à d'autres sous-groupes, ayant potentiellement d'autres spécificités de substrats ; les protéines portant un système d'ancrage GPI les maintenant liées à la paroi fongique pourraient notamment avoir un rôle biologique différent. Par ailleurs, bien que les structures tridimensionnelles de certaines AA16 aient été modélisées et révèlent certains points communs avec d'autres familles de LPMO actives sur la cellulose, il est impératif d'obtenir des informations structurales expérimentales pour mieux comprendre les particularités de cette nouvelle famille.

Afin de poursuivre les travaux initiés dans ce projet, il serait également bénéfique d'envisager la production recombinante des AA16 dans un autre hôte que *Pichia pastoris*. Il est possible que la méthylation de l'histidine N-terminale, probablement impliquée dans sa protection contre les dommages oxydatifs, puisse avoir encore plus d'importance pour les AA16 que pour d'autres familles ; leur production chez *P. pastoris*, qui est incapable de faire cette modification post-traductionnelle, pourrait alors être délétère pour leur activité. Une

production dans un hôte capable d'assurer la méthylation de l'histidine, comme par exemple *Aspergillus niger*, pourrait être mieux adaptée. On peut aussi envisager une expression dans *Trichoderma reesei*. Si le bénéfice d'une ou plusieurs enzymes pour l'amélioration de la saccharification est avéré, l'objectif sera en effet d'intégrer le gène correspondant dans le génome de *T. reesei*, afin de produire l'enzyme conjointement avec le cocktail natif et de diminuer ainsi les coûts.

Si les tests réalisés ont permis d'observer une synergie des AA16 avec l'une des CBH de *Trichoderma reesei* sur cellulose pure, aucune synergie n'a pour l'instant été observée avec le cocktail entier, ni sur des substrats plus complexes. Il serait donc intéressant de voir si en améliorant les conditions de production des AA16, et après avoir déterminé les bonnes conditions de fonctionnement de ces enzymes (choix du donneur d'électron, ratio entre les LPMO et les cellulases etc.), il est possible d'observer un apport des AA16 pour la saccharification de biomasse prétraitée telle que le Miscanthus. Cela permettrait d'envisager des applications pour la production industrielle de bioéthanol ou d'autres produits issus de biomasse lignocellulosique.

Annexes

Annexe 1 : Séquences utilisées pour la production de protéines recombinantes

Organisme d'origine de la séquence	Identifiant Genbank	Séquence du peptide signal	Séquence de la protéine (seules les parties entre crochets ont été exprimées)
<i>Aspergillus aculeatus</i> ATCC 16872	XP_020060743.1	MKQTGSILAL AGLVSMANA	[HGFVTSPQPRMPGSAMEKACGQQVYNNQEAD NYGNIQGELQIASGQSDYDAEACDIWLCKGYKY ADNTANVQSYKPGEVIDFTVDIRAPHTGTANVS VVDATNTMLSQPLIYWSVYASTATGVTANETS FSVTMPDLGDKCNEAGACVLQWWWDARSIDQ TYESCVDFTLSG]SSSSSSSSSSSSSSSSSSSSSSSSSS AAAAVATSAASSSTSSAAETTTAASAAVTSPAA ANNIAPVSSSGPTTLATSVKQTATAAAVSSATST SVSLPTDGTAEEQLTWVASVFKALLNYAN
<i>Aspergillus japonicus</i> BRFM 405	Génome non publié	MKQAGSILAL AGLVSLAQG	[HGFVTSPQARMPGSAMEKACGQQVYNNQEAD NYGNVQGELQIASGQSDYDADACDIWLCKGYK YADNKDNVYSYKAGETVDFTVDIRAPHTGTAN VSVVDASNTVIGSPLIYWSVYASTSTGVTANET SFSVTIPDDLGDKCTEGGACVLQWWWNAASID QTYESCIDFTVGGSGSSSSSS]SSSSSSSSSSAAASQQ TTTTAAAAAVTTPSTANNVAPASSVPTTLATSV KQSATVAPVAASSAAAAASSSGVSIPTDGTAEEQ LTWVASLLKALLKYAN
<i>Aureobasidium pullulans</i> EXF-150	KEQ78697.1	MKYTTTALAV AGMAATVRG	[HGFLSSPKARMPGTTFQAACGQQVYYNQAGDN YGNIQGELQVAKTQTDYKAAECDVWLCKGFKY EDNKDLVQTYTPGQKVPITFDIRAPHTGTANVSI VDTTNTIIGAPLISWDVYASTASSIPANETSFDIT IPEDLGSKCSTAGACVIQHWWSREVDQTYESCI DFTVSGSGS GAAPASSKAASSKPASSAVASSSAA PVVSSSAAPVTTNVAKVTSSSSVVAKAPVTSTKA AASSVKPTTMATVVKSSSAAPASSGSASGTVA YGCQGGKSYTGATQCASGYTCKVQNDYYSQCI AGSAKV]
<i>Pestalotiopsis fici</i> W106-1	XP_007832240.1	MKYAITFGAA VAPMVAA	[HGMIKTPTPRSAGTAMSSACGSQIYNMSSDSY GNIQGETQLIGSDFTDECNLWQCKGMQYSDNTA NVQSYSTGDVVDITYDIRAPHTGTANVSIVDTAS NTIIGDVLAYWSVFASNSQASAANETSFSITIPDL GGKCTTGGEQVIQHYWDSQSAGQTYESCIDFTV GGSSSGSSGSASTTTAAATSAAAATSQVATTSQA ATSVVAVSSTKSVVSAATSSAAAETGDDDED SCDADDDDEAEDAGDDEDDDDSCDADDEDDEPV ASSAPASSAAPTTLVTRTSTAAAQATSAASSGSS SGSVALYGQCGGINYSGSTTCASGTCTVMNDYY SQC]

Nb de secrétomes dans le grp	Nom des secrétomes	Nb de prot. dans l'intersection	Nb de secrétomes dans le grp	Nom des secrétomes	Nb de prot. dans l'intersection
10	1487_ASM 1487_Avi 1487_SBP 1489_ASM 1489_Avi 1489_SBP 1490_SBP 405_ASM 405_Avi 405_SBP	1	8	1487_ASM 1489_Avi 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_Avi 405_SBP	1
10	1487_ASM 1487_Avi 1487_SBP 1489_ASM 1489_Avi 1490_ASM 1490_SBP 405_ASM 405_Avi 405_SBP	1	8	1487_Avi 1489_Avi 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_Avi 405_SBP	8
8	1487_Avi 1489_Avi 1490_ASM 1490_Avi 405_ASM 405_Avi 405_SBP 430_Avi	1	6	1489_Avi 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_Avi	1
8	1487_Avi 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_Avi 405_SBP 430_SBP	1	6	1487_ASM 1487_Avi 1487_SBP 1489_ASM 1490_ASM 405_ASM	1
8	1489_Avi 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_Avi 405_SBP 430_ASM	1	6	1487_ASM 1487_SBP 1489_ASM 1489_Avi 1489_SBP 405_ASM	1
8	1490_ASM 1490_Avi 1490_SBP 405_ASM 405_Avi 405_SBP 430_ASM 430_SBP	1	6	1490_ASM 1490_Avi 1490_SBP 405_SBP 430_ASM 430_SBP	1
8	1487_ASM 1487_SBP 1489_ASM 1489_SBP 1490_Avi 1490_SBP 405_ASM 405_SBP	1	6	1487_Avi 1487_SBP 1489_Avi 1490_Avi 405_Avi 430_SBP	1
8	1487_ASM 1487_SBP 1489_ASM 1489_Avi 1490_ASM 1490_SBP 405_ASM 405_SBP	1	6	1487_ASM 1487_Avi 1487_SBP 1489_ASM 1489_Avi 1489_SBP	1
8	1487_ASM 1487_SBP 1489_ASM 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_SBP	2	6	1487_ASM 1487_Avi 1487_SBP 1489_ASM 1489_Avi 1490_SBP	1
8	1487_ASM 1489_ASM 1490_ASM 1490_SBP 405_ASM 405_SBP 430_ASM 430_SBP	1	6	1487_SBP 1490_ASM 1490_Avi 1490_SBP 430_ASM 430_SBP	1
8	1487_ASM 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_SBP 430_ASM 430_SBP	1	5	1487_ASM 1487_Avi 405_ASM 405_Avi 405_SBP	1
8	1487_ASM 1487_Avi 1487_SBP 1489_ASM 1489_Avi 1490_ASM 405_ASM 405_Avi	1	5	1487_Avi 1490_ASM 405_ASM 405_Avi 405_SBP	1
8	1487_ASM 1487_Avi 1489_Avi 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_Avi	1	5	1490_ASM 1490_SBP 405_ASM 405_Avi 405_SBP	3
8	1487_Avi 1489_Avi 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_Avi 430_ASM	1	5	1490_ASM 1490_Avi 405_ASM 405_Avi 405_SBP	1
8	1487_Avi 1487_SBP 1489_ASM 1489_Avi 1489_SBP 1490_ASM 1490_Avi 405_ASM	1	5	1487_ASM 1487_SBP 1489_ASM 405_ASM 405_SBP	1
7	1487_ASM 1487_Avi 1489_Avi 1490_ASM 405_ASM 405_Avi 405_SBP	1	5	1487_ASM 1487_SBP 1490_SBP 405_ASM 405_SBP	1
7	1487_ASM 1487_Avi 1490_ASM 1490_SBP 405_ASM 405_Avi 405_SBP	1	5	1487_ASM 1489_ASM 1490_ASM 405_ASM 405_SBP	1
7	1487_ASM 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_Avi 405_SBP	1	5	1487_ASM 1490_ASM 1490_SBP 405_ASM 405_SBP	1
7	1487_SBP 1490_ASM 1490_SBP 405_ASM 405_Avi 405_SBP 430_SBP	1	5	1487_ASM 1490_SBP 405_ASM 405_SBP 430_SBP	1
7	1487_SBP 1490_Avi 1490_SBP 405_ASM 405_Avi 405_SBP 430_ASM	1	5	1487_SBP 1490_ASM 1490_SBP 405_ASM 405_SBP	1
7	1487_Avi 1489_ASM 1489_Avi 1490_ASM 405_ASM 405_Avi 405_SBP	1	5	1487_Avi 1490_ASM 1490_SBP 405_ASM 405_SBP	1
7	1487_Avi 1489_Avi 1490_ASM 1490_SBP 405_ASM 405_Avi 405_SBP	1	5	1490_ASM 1490_Avi 1490_SBP 405_ASM 405_SBP	5
7	1487_Avi 1489_Avi 1490_Avi 1490_SBP 405_ASM 405_Avi 405_SBP	1	5	1490_SBP 405_ASM 405_SBP 430_ASM 430_SBP	1
7	1490_ASM 1490_SBP 405_ASM 405_Avi 405_SBP 430_ASM 430_SBP	1	5	1487_ASM 1489_Avi 1490_ASM 405_ASM 405_Avi	1
7	1490_ASM 1490_Avi 1490_SBP 405_ASM 405_SBP 430_ASM 430_SBP	1	5	1487_ASM 1490_ASM 1490_SBP 405_ASM 405_Avi	1
7	1487_Avi 1489_ASM 1489_Avi 1490_ASM 1490_Avi 405_ASM 405_Avi	1	5	1487_Avi 1489_ASM 1489_Avi 405_ASM 405_Avi	1
7	1490_ASM 1490_Avi 1490_SBP 405_Avi 405_SBP 430_ASM 430_SBP	1	5	1487_Avi 1489_Avi 1490_Avi 405_ASM 405_Avi	1
7	1487_ASM 1487_SBP 1489_SBP 1490_ASM 1490_Avi 1490_SBP 430_SBP	1	5	1487_ASM 1487_Avi 1487_SBP 1489_Avi 405_ASM	1
6	1487_ASM 1487_SBP 1490_ASM 405_ASM 405_Avi 405_SBP	1	5	1487_ASM 1490_ASM 1490_Avi 1490_SBP 405_ASM	1
6	1487_ASM 1487_SBP 1490_Avi 405_ASM 405_Avi 405_SBP	1	5	1490_ASM 1490_SBP 405_SBP 430_ASM 430_SBP	1
6	1487_ASM 1487_Avi 1489_Avi 405_ASM 405_Avi 405_SBP	1	5	1487_ASM 1487_SBP 1489_Avi 1490_SBP 405_Avi	1
6	1487_SBP 1490_ASM 1490_SBP 405_ASM 405_Avi 405_SBP	1	5	1487_ASM 1487_Avi 1489_ASM 1489_Avi 405_Avi	1
6	1487_Avi 1489_Avi 1490_ASM 405_ASM 405_Avi 405_SBP	2	5	1487_ASM 405_Avi 430_ASM 430_Avi 430_SBP	1
6	1489_Avi 1490_ASM 1490_SBP 405_ASM 405_Avi 405_SBP	1	5	1487_ASM 1487_Avi 1487_SBP 1489_Avi 1490_ASM	1
6	1489_Avi 405_ASM 405_Avi 405_SBP 430_ASM 430_SBP	1	5	1487_ASM 1487_SBP 1490_SBP 430_ASM 430_SBP	1
6	1490_ASM 1490_Avi 1490_SBP 405_ASM 405_Avi 405_SBP	2	5	1487_ASM 1487_Avi 1489_Avi 1490_Avi 1490_SBP	1
6	1487_ASM 1487_SBP 1489_ASM 1490_ASM 405_ASM 405_SBP	1	5	1487_ASM 1489_ASM 1490_ASM 1490_Avi 1490_SBP	1
6	1487_ASM 1487_SBP 1490_ASM 1490_SBP 405_ASM 405_SBP	2	5	1487_Avi 1487_SBP 1489_Avi 1490_ASM 1490_Avi	1

Nb de sécrétomes dans le grp	Nom des sécrétomes	Nb de prot. dans l'intersection	Nb de sécrétomes dans le grp	Nom des sécrétomes	Nb de prot. dans l'intersection
6	1487_ASM 1487_SBP 1490_SBP 405_ASM 405_SBP 430_SBP	1	5	1487_Avi 1489_Avi 1490_ASM 1490_Avi 1490_SBP	1
6	1487_ASM 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_SBP	1	5	1487_Avi 1490_ASM 1490_Avi 1490_SBP 430_ASM	1
6	1487_SBP 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_SBP	2	5	1490_ASM 1490_SBP 430_ASM 430_Avi 430_SBP	2
6	1487_Avi 1490_ASM 1490_Avi 1490_SBP 405_ASM 405_SBP	1	4	1489_Avi 405_ASM 405_Avi 405_SBP	2
6	1487_Avi 1487_SBP 1489_ASM 1489_Avi 405_ASM 405_Avi	1	4	1490_ASM 405_ASM 405_Avi 405_SBP	1
4	1490_SBP 405_ASM 405_Avi 405_SBP	1	3	1490_ASM 1490_Avi 1490_SBP	5
4	1487_ASM 1487_SBP 405_ASM 405_SBP	1	3	1490_ASM 1490_SBP 430_ASM	1
4	1487_SBP 1490_Avi 405_ASM 405_SBP	1	3	1490_ASM 430_ASM 430_SBP	1
4	1490_ASM 1490_SBP 405_ASM 405_SBP	3	3	430_ASM 430_Avi 430_SBP	7
4	1487_ASM 1489_Avi 405_ASM 405_Avi	1	2	405_ASM 405_SBP	2
4	1487_Avi 1489_Avi 405_ASM 405_Avi	2	2	1487_ASM 405_ASM	3
4	1487_ASM 1487_SBP 1489_ASM 405_ASM	1	2	1487_SBP 405_ASM	1
4	1487_ASM 1489_ASM 1489_Avi 405_ASM	1	2	1489_Avi 405_ASM	5
4	1487_SBP 1489_SBP 1490_ASM 405_ASM	1	2	1490_ASM 405_ASM	5
4	1489_ASM 1489_Avi 1490_ASM 405_ASM	1	2	405_Avi 405_SBP	2
4	1490_ASM 1490_Avi 1490_SBP 405_ASM	1	2	1487_SBP 405_SBP	1
4	1487_Avi 1487_SBP 1489_Avi 405_SBP	1	2	1487_ASM 405_Avi	1
4	1487_SBP 1490_ASM 1490_SBP 405_SBP	1	2	1489_Avi 405_Avi	3
4	1487_SBP 1490_ASM 405_SBP 430_ASM	1	2	1490_SBP 405_Avi	2
4	1490_ASM 1490_Avi 1490_SBP 405_SBP	2	2	1487_ASM 1487_SBP	3
4	1490_ASM 1490_Avi 1490_SBP 405_Avi	1	2	1487_ASM 1487_Avi	3
4	1490_SBP 405_Avi 430_ASM 430_SBP	1	2	1487_ASM 1489_ASM	1
4	1487_ASM 1487_SBP 1490_Avi 430_SBP	1	2	1487_ASM 1490_ASM	1
4	1487_ASM 1489_ASM 1489_Avi 1489_SBP	1	2	1487_ASM 1490_SBP	2
4	1487_ASM 1489_ASM 430_ASM 430_SBP	1	2	1487_ASM 430_SBP	1
4	1487_ASM 1490_ASM 1490_Avi 1490_SBP	2	2	1487_SBP 1489_SBP	2
4	1487_SBP 1489_SBP 430_ASM 430_SBP	1	2	1487_Avi 1489_Avi	13
4	1487_SBP 1490_ASM 1490_Avi 1490_SBP	1	2	1487_Avi 1490_SBP	1
4	1489_Avi 1490_ASM 1490_Avi 1490_SBP	1	2	1489_ASM 1490_Avi	1
4	1490_SBP 430_ASM 430_Avi 430_SBP	1	2	1489_Avi 1490_ASM	1
3	405_ASM 405_Avi 405_SBP	2	2	1489_Avi 1490_Avi	1
3	1487_ASM 405_ASM 405_SBP	1	2	1489_Avi 430_SBP	1
3	1490_ASM 405_ASM 405_SBP	2	2	1489_Avi 430_Avi	1
3	1490_SBP 405_ASM 405_SBP	1	2	1490_ASM 1490_SBP	3
3	1489_Avi 405_ASM 405_Avi	4	2	1490_ASM 1490_Avi	2
3	1490_ASM 405_ASM 405_Avi	1	2	1490_ASM 430_ASM	1
3	1487_ASM 1487_Avi 405_ASM	1	2	1490_Avi 1490_SBP	2
3	1487_ASM 1489_ASM 405_ASM	2	2	1490_SBP 430_ASM	1
3	1487_ASM 1489_Avi 405_ASM	1	2	1490_SBP 430_SBP	1
3	1487_Avi 1490_ASM 405_ASM	1	2	430_ASM 430_SBP	31
3	1489_ASM 1489_Avi 405_ASM	2	2	430_ASM 430_Avi	1
3	1490_ASM 1490_SBP 405_SBP	1	1	405_ASM	20
3	1490_Avi 1490_SBP 405_SBP	1	1	405_SBP	6
3	1487_Avi 1489_Avi 405_Avi	1	1	405_Avi	8
3	1487_Avi 1490_Avi 405_Avi	1	1	1487_ASM	18
3	1487_ASM 1487_Avi 1487_SBP	1	1	1487_SBP	3
3	1487_ASM 1487_SBP 1489_ASM	1	1	1487_Avi	5
3	1487_ASM 1487_SBP 1489_Avi	2	1	1489_ASM	5
3	1487_ASM 1487_SBP 1490_SBP	1	1	1489_SBP	1
3	1487_ASM 1487_Avi 1489_Avi	1	1	1489_Avi	34
3	1487_ASM 1490_ASM 1490_SBP	1	1	1490_ASM	4
3	1487_Avi 1487_SBP 1489_Avi	1	1	1490_SBP	14
3	1487_SBP 1489_Avi 1490_Avi	1	1	1490_Avi	5
3	1487_SBP 1490_ASM 1490_SBP	1	1	430_ASM	12
3	1487_Avi 1489_Avi 1490_Avi	1	1	430_SBP	18
3	1487_Avi 1490_ASM 1490_Avi	1	1	430_Avi	4
3	1489_ASM 1490_ASM 1490_SBP	2			

Références

- Abarca, M. L., Accensi, F., Cano, J. and Cabañes, F. J. (2004) Taxonomy and significance of black aspergilli. *Antonie van Leeuwenhoek* **86**, 33–49.
- Adav, S. S., Chao, L. T. and Sze, S. K. (2012) Quantitative Secretomic Analysis of *Trichoderma reesei* Strains Reveals Enzymatic Composition for Lignocellulosic Biomass Degradation. *Molecular and Cellular Proteomics* **11**.
- Adler, E. (1977) Lignin chemistry—past, present and future. *Wood Science and Technology* **11**, 169–218.
- Agostoni, M., Hangasky, J. A. and Marletta, M. A. (2017) Physiological and Molecular Understanding of Bacterial Polysaccharide Monooxygenases. *Microbiology and molecular biology reviews MMBR* **81**.
- Arantes, V., Jellison, J. and Goodell, B. (2012) Peculiarities of brown-rot fungi and biochemical Fenton reaction with regard to their potential as a model for bioprocessing biomass. *Applied microbiology and biotechnology* **94**, 323–338.
- Arantes, V. and Saddler, J. N. (2011) Cellulose accessibility limits the effectiveness of minimum cellulase loading on the efficient hydrolysis of pretreated lignocellulosic substrates. *Biotechnology for biofuels* **4**, 3.
- Arfi, Y., Chevret, D., Henrissat, B., Berrin, J.-G., Levasseur, A. and Record, E. (2013) Characterization of salt-adapted secreted lignocellulolytic enzymes from the mangrove fungus *Pestalotiopsis* sp. *Nature communications* **4**, 1810.
- Aro, N., Pakula, T. and Penttilä, M. (2005) Transcriptional regulation of plant cell wall degradation by filamentous fungi. *FEMS Microbiology Reviews* **29**, 719–739.
- Ayrinhac, C., Margeot, A., Ferreira, N. L., Ben Chaabane, F., Monot, F., Ravot, G., Sonet, J.-M. and Fourage, L. (2011) Improved Saccharification of Wheat Straw for Biofuel Production Using an Engineered Secretome of *Trichoderma reesei*. *Organic Process Research and Development* **15**, 275–278.
- Badino, S. F., Christensen, S. J., Kari, J., Windahl, M. S., Hvidt, S., Borch, K. and Westh, P. (2017) Exo-exo synergy between Cel6A and Cel7A from *Hypocrea jecorina*: Role of carbohydrate binding module and the endo-lytic character of the enzymes. *Biotechnology and Bioengineering* **114**, 1639–1647.
- Batista-García, R. A., Sutton, T., Jackson, S. A., Tovar-Herrera, O. E., Balcázar-López, E., Sánchez-Carbente, M. D. R., Sánchez-Reyes, A., Dobson, A. D. W. and Folch-Mallol, J. L. (2017) Characterization of lignocellulolytic activities from fungi isolated from the deep-sea sponge *Stelletta normani*. *PLoS ONE* **12**.
- Beeson, W. T., Vu, V. V., Span, E. A., Phillips, C. M. and Marletta, M. A. (2015) Cellulose Degradation by Polysaccharide Monooxygenases. *Annual review of biochemistry* **84**, 923–946.
- Bennati-Granier, C., Garajova, S., Champion, C., Grisel, S., Haon, M., Zhou, S., Fanuel, M., Ropartz, D., Rogniaux, H., Gimbert, I., Record, E. and Berrin, J.-G. (2015) Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by *Podospora anserina*. *Biotechnology for biofuels* **8**, 90.
- Benoit, I., Culleton, H., Zhou, M., DiFalco, M., Aguilar-Osorio, G., Battaglia, E., Bouzid, O., Brouwer, C. P. J. M., El-Bushari, H. B. O., Coutinho, P. M., Gruben, B. S., Hildén, K. S., Houbraeken, J., Barboza, L. A. J., Levasseur, A., Majoor, E., Mäkelä, M. R., Narang, H.-M., Trejo-Aguilar, B., van den Brink, J., vanKuyk, P. A., Wiebenga, A., McKie, V., McCleary, B., Tsang, A., Henrissat, B. and Vries, R. P. de (2015) Closely related fungi employ diverse enzymatic strategies to degrade plant biomass. *Biotechnology for biofuels* **8**, 107.
- Berlin, A., Maximenko, V., Gilkes, N. and Saddler, J. (2007) Optimization of enzyme complexes for lignocellulose hydrolysis. *Biotechnology and Bioengineering* **97**, 287–296.
- Berrin, J.-G., Navarro, D., Couturier, M., Olivé, C., Grisel, S., Haon, M., Taussac, S., Lechat, C., Courtecuisse, R., Favel, A., Coutinho, P. M. and Lesage-Meessen, L. (2012) Exploring the Natural Fungal Biodiversity of Tropical and Temperate Forests toward Improvement of Biomass Conversion. *Applied and environmental microbiology* **78**, 6483–6490.
- Berrin, J.-G., Navarro, D., Lopes Ferreira, N., Margeot, A., Coutinho, P.M. and Henrissat, B. (2013). *Préparation multi-enzymatique contenant le génome d'une souche d'Aspergillus japonicus* (WO2014037925A1).
- Berrin, J.-G., Navarro, D., Lopes-Ferreira, N., Margeot, A., Coutinho, P. and Henrissat, B. (2015). *Multi-enzymatic preparation containing the secretome of an Aspergillus japonicus strain* (WO2014037925A1).
- Berrin, J.-G., Rosso, M.-N. and Abou Hachem, M. (2017) Fungal secretomics to probe the biological functions of lytic polysaccharide monooxygenases. *Carbohydrate research* **448**, 155–160.
- Bischof, R. H., Ramoni, J. and Seiboth, B. (2016) Cellulases and beyond: The first 70 years of the enzyme producer *Trichoderma reesei*. *Microbial Cell Factories* **15**, 106.
- Bissaro, B., Forsberg, Z., Ni, Y., Hollmann, F., Vaaje-Kolstad, G. and Eijsink, V. G. H. (2016) Fueling biomass-degrading oxidative enzymes by light-driven water oxidation. *Green Chemistry* **18**, 5357–5366.
- Bissaro, B., Rohr, A. K., Muller, G., Chylenski, P., Skaugen, M., Forsberg, Z., Horn, S. J., Vaaje-Kolstad, G. and Eijsink, V. G. H. (2017) Oxidative cleavage of polysaccharides by monocopper enzymes depends on H₂O₂. *Nature Chemical Biology* **13**, 1123–1128.

- Bissaro, B., Várnai, A., Røhr, Å. K. and Eijsink, V. G. H. (2018) Oxidoreductases and Reactive Oxygen Species in Conversion of Lignocellulosic Biomass. *Microbiology and molecular biology reviews MMBR* **82**.
- Bjølgerud, H. *Molecular evolution of the substrate specificity of bacterial lytic polysaccharide monoxygenases (LPMOs)*: Norwegian University of Life Sciences, Ås.
- Blanchette, R. A. (2000) A review of microbial deterioration found in archaeological wood from different environments. *International Biodeterioration & Biodegradation* **46**, 189–204.
- Borin, G. P., Sanchez, C. C., Santana, E. S. de, Zanini, G. K., dos Santos, R. A. C., Oliveira Pontes, A. de, Souza, A. T. de, Dal'Mas, R. M. M. T. S., Riaño-Pachón, D. M., Goldman, G. H. and Oliveira, J. V. d. C. (2017) Comparative transcriptome analysis reveals different strategies for degradation of steam-exploded sugarcane bagasse by *Aspergillus niger* and *Trichoderma reesei*. *BMC Genomics* **18**, 501.
- Bouws, H., Wattenberg, A. and Zorn, H. (2008) Fungal secretomes—nature's toolbox for white biotechnology. *Applied microbiology and biotechnology* **80**, 381.
- Brar, K. K., Agrawal, D., Chadha, B. S. and Lee, H. (2018) Evaluating novel fungal secretomes for efficient saccharification and fermentation of composite sugars derived from hydrolysate and molasses into ethanol. *Bioresource Technology* **273**, 114–121.
- Brown, K., Harris, P., Zaretsky, E., Re, E., Vlasenko, E., McFarland, K. and Leon, A.L.d. (2008). *Polypeptide from a cellulolytic fungus having cellulolytic enhancing activity* (WO2005074647A2).
- Brown, N. A., Ries, L. N. A. and Goldman, G. H. (2014) How nutritional status signalling coordinates metabolism and lignocellulolytic enzyme secretion. *Fungal Genetics and Biology* **72**, 48–63.
- Brown, R. M. (2004) Cellulose structure and biosynthesis: What is in store for the 21st century? *Journal of Polymer Science Part A: Polymer Chemistry* **42**, 487–495.
- Brumer, H., Spencer, W. J., Svensson, B., Henrissat, B., Davies, G. J., Gilbert, H. J., Clarke, A. J., Wakarchuk, W. W., Wade, A. D., Vocadlo, D. J., Ficko-Blean, E., Boraston, A., Planas, A. and Fushinobu, S. (2018) Ten years of CAZyedia: A living encyclopedia of carbohydrate-active enzymes. *Glycobiology* **28**, 3–8.
- Butler, M. I., Gray, J., Goodwin, T. J. D. and Poulter, R. T. M. (2006) The distribution and evolutionary history of the PRP8 intein. *BMC evolutionary biology* **6**, 42.
- Caffall, K. H. and Mohnen, D. (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate research* **344**, 1879–1900.
- Cannella, D., Hsieh, C.-W. C., Felby, C. and Jørgensen, H. (2012) Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content. *Biotechnology for biofuels* **5**, 26.
- Cannella, D., Möllers, K. B., Frigaard, N.-U., Jensen, P. E., Bjerrum, M. J., Johansen, K. S. and Felby, C. (2016) Light-driven oxidation of polysaccharides by photosynthetic pigments and a metalloenzyme. *Nature communications* **7**, 11134.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic acids research* **37**, D233-D238.
- Chalak, A., Villares, A., Moreau, C., Haon, M., Grisel, S., d'Orlando, A., Herpoël-Gimbert, I., Labourel, A., Cathala, B. and Berrin, J.-G. (2019) Influence of the carbohydrate-binding module on the activity of a fungal AA9 lytic polysaccharide monoxygenase on cellulosic substrates. *Biotechnology for biofuels* **12**, 1–10.
- Cherubini, F. (2010) The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. *Energy Conversion and Management* **51**, 1412–1421.
- Ciano, L., Davies, G. J., Tolman, W. B. and Walton, P. H. (2018) Bracing copper for the catalytic oxidation of C–H bonds. *Nature Catalysis*, 571–577.
- Corrêa, T. L. R., Júnior, A. T., Wolf, L. D., Buckeridge, M. S., dos Santos, L. V. and Murakami, M. T. (2019) An actinobacteria lytic polysaccharide monoxygenase acts on both cellulose and xylan to boost biomass saccharification. *Biotechnology for biofuels* **12**, 219.
- Couturier, M., Bennati-Granier, C., Barbian Urio, M., Pereira Ramos, L. and Berrin, J.-G. (2016) Fungal Enzymatic Degradation of Cellulose. In: *Green Fuels Technology*.
- Couturier, M. and Berrin, J.-G. (2013) The Saccharification Step: The Main Enzymatic Components. In: *Lignocellulose Conversion*, pp. 93–110: Springer, Berlin, Heidelberg.
- Couturier, M., Haon, M., Coutinho, P. M., Henrissat, B., Lesage-Meessen, L. and Berrin, J.-G. (2011) *Podospira anserina* Hemicellulases Potentiate the *Trichoderma reesei* Secretome for Saccharification of Lignocellulosic Biomass. *Applied and environmental microbiology* **77**, 237–246.
- Couturier, M., Ladevèze, S., Sulzenbacher, G., Ciano, L., Fanuel, M., Moreau, C., Villares, A., Cathala, B., Chaspoul, F., Frandsen, K. E., Labourel, A., Herpoël-Gimbert, I., Grisel, S., Haon, M., Lenfant, N., Rogniaux, H., Ropartz, D., Davies, G. J., Rosso, M.-N., Walton, P. H., Henrissat, B. and Berrin, J.-G. (2018) Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. *Nature Chemical Biology*.

- Couturier, M., Navarro, D., Olivé, C., Chevret, D., Haon, M., Favel, A., Lesage-Meessen, L., Henrissat, B., Coutinho, P. M. and Berrin, J.-G. (2012) Post-genomic analyses of fungal lignocellulosic biomass degradation reveal the unexpected potential of the plant pathogen *Ustilago maydis*. *BMC Genomics* **13**, 57.
- Crooks, G. E., Hon, G., Chandonia, J.-M. and Brenner, S. E. (2004) WebLogo: A Sequence Logo Generator. *Genome Research* **14**, 1188–1190.
- Crouch, L. I., Labourel, A., Walton, P. H., Davies, G. J. and Gilbert, H. J. (2016) The Contribution of Non-catalytic Carbohydrate Binding Modules to the Activity of Lytic Polysaccharide Monooxygenases. *Journal of Biological Chemistry* **291**, 7439–7449.
- Culleton, H., McKie, V. and Vries, R. P. de (2013) Physiological and molecular aspects of degradation of plant polysaccharides by fungi: What have we learned from *Aspergillus*? *Biotechnology Journal* **8**, 884–894.
- Dashtban, M., Schraft, H. and Qin, W. (2009) Fungal Bioconversion of Lignocellulosic Residues; Opportunities & Perspectives. *Interbational Journal of Biological Sciences* **5**, 578–595.
- Davies, G. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859.
- Davin, L. B., Patten, A. M., Jourdes, M. and Lewis, N. G. (2008) Lignins: A Twenty-First Century Challenge. In: *Biomass Recalcitrance Deconstructing the Plant Cell Wall for Bioenergy* (Michael E. Himmel, ed), pp. 213–305: Wiley-Blackwell.
- de Bhowmick, G., Sarmah, A. K. and Sen, R. (2018) Lignocellulosic biorefinery as a model for sustainable development of biofuels and value added products. *Bioresource Technology* **247**, 1144–1154.
- de Vries, R. P., Riley, R., Wiebenga, A., Aguilar-Osorio, G., Amillis, S., Uchima, C. A., Anderluh, G., Asadollahi, M., Askin, M., Barry, K., Battaglia, E., Bayram, Ö., Benocci, T., Braus-Stromeyer, S. A., Caldana, C., Cánovas, D., Cerqueira, G. C., Chen, F., Chen, W., Choi, C., Clum, A., Santos, R. A. C. d., Damásio, A. R. d. L., Diallinas, G., Emri, T., Fekete, E., Flipphi, M., Freyberg, S., Gallo, A., Gournas, C., Habgood, R., Hainaut, M., Harispe, M. L., Henrissat, B., Hildén, K. S., Hope, R., Hossain, A., Karabika, E., Karaffa, L., Karányi, Z., Kraševc, N., Kuo, A., Kusch, H., LaButti, K., Lagendijk, E. L., Lapidus, A., Levasseur, A., Lindquist, E., Lipzen, A., Logrieco, A. F., MacCabe, A., Mäkelä, M. R., Malavazi, I., Melin, P., Meyer, V., Mielnichuk, N., Miskei, M., Molnár, Á. P., Mulé, G., Ngan, C. Y., Orejas, M., Orosz, E., Ouedraogo, J. P., Overkamp, K. M., Park, H.-S., Perrone, G., Piumi, F., Punt, P. J., Ram, A. F. J., Ramón, A., Rauscher, S., Record, E., Riaño-Pachón, D. M., Robert, V., Röhrig, J., Ruller, R., Salamov, A., Salih, N. S., Samson, R. A., Sándor, E., Sanguinetti, M., Schütze, T., Sepčić, K., Shelest, E., Sherlock, G., Sophianopoulou, V., Squina, F. M., Sun, H., Susca, A., Todd, R. B., Tsang, A., Unkles, S. E., van de Wiele, N., van Rossen-Uffink, D., Oliveira, J. V. d. C., Vesth, T. C., Visser, J., Yu, J.-H., Zhou, M., Andersen, M. R., Archer, D. B., Baker, S. E., Benoit, I., Brakhage, A. A., Braus, G. H., Fischer, R., Frisvad, J. C., Goldman, G. H., Houbraken, J., Oakley, B., Pócsi, I., Scazzocchio, C., Seiboth, B., vanKuyk, P. A., Wortman, J., Dyer, P. S. and Grigoriev, I. V. (2017) Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biology* **18**, 28.
- de Vries, R. P. and Visser, J. (2001) *Aspergillus* Enzymes Involved in Degradation of Plant Cell Wall Polysaccharides. *Microbiology and molecular biology reviews MMBR* **65**, 497–522.
- Derntl, C., Kiesenhofer, D. P., Mach, R. L. and Mach-Aigner, A. R. (2015) Novel Strategies for Genomic Manipulation of *Trichoderma reesei* with the Purpose of Strain Engineering. *Applied and environmental microbiology* **81**, 6314–6323.
- Di Cologna, N. d. M., Gómez-Mendoza, D. P., Zanoelo, F. F., Giannesi, G. C., Guimarães, N. C. d. A., Moreira, L. R. d. S., Filho, E. X. F. and Ricart, C. A. O. (2017) Exploring *Trichoderma* and *Aspergillus* secretomes: Proteomics approaches for the identification of enzymes of biotechnological interest. *Enzyme and Microbial Technology*.
- Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K., Teeri, T. T. and Jones, T. A. (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **265**, 524–528.
- dos Santos Castro, L., Pedersoli, W. R., Antoniêto, A. C. C., Steindorff, A. S., Silva-Rocha, R., Martinez-Rossi, N. M., Rossi, A., Brown, N. A., Goldman, G. H., Faça, V. M., Persinoti, G. F. and Silva, R. N. (2014) Comparative metabolism of cellulose, sophorose and glucose in *Trichoderma reesei* using high-throughput genomic and proteomic analyses. *Biotechnology for biofuels* **7**, 41.
- Dotson, W., Greenier, J. and Ding, H. (2005). *Polypeptides having cellulolytic enhancing activity and polynucleotides encoding same* (WO2005074656A2).
- Du, J., Cao, Y., Liu, G., Zhao, J., Li, X. and Qu, Y. (2017) Identifying and overcoming the effect of mass transfer limitation on decreased yield in enzymatic hydrolysis of lignocellulose at high solid concentrations. *Bioresource Technology* **229**, 88–95.
- Durand, H., Clanet, M. and Tiraby, G. (1988) Genetic improvement of *Trichoderma reesei* for large scale cellulase production. *Enzyme and Microbial Technology* **10**, 341–346.
- Edgar, R. C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797.

- Eibinger, M., Ganner, T., Bubner, P., Rošker, S., Kracher, D., Haltrich, D., Ludwig, R., Plank, H. and Nidetzky, B. (2014) Cellulose Surface Degradation by a Lytic Polysaccharide Monooxygenase and Its Effect on Cellulase Hydrolytic Efficiency. *The Journal of biological chemistry* **289**, 35929–35938.
- Eibinger, M., Sattelkow, J., Ganner, T., Plank, H. and Nidetzky, B. (2017) Single-molecule study of oxidative enzymatic deconstruction of cellulose. *Nature communications* **8**, 894.
- Eijsink, V. G. H., Petrovic, D., Forsberg, Z., Mekasha, S., Røhr, Å. K., Várnai, A., Bissaro, B. and Vaaje-Kolstad, G. (2019) On the functional characterization of lytic polysaccharide monooxygenases (LPMOs). *Biotechnology for biofuels* **12**, 58.
- Eriksson, K.-E., Pettersson, B. and Westermark, U. (1974) Oxidation: An important enzyme reaction in fungal degradation of cellulose. *FEBS Letters* **49**, 282–285.
- Eveleigh, D. E. and Montenecourt, B. S. (1979) Increasing yields of extracellular enzymes. *Advances in applied microbiology* **25**, 57–74.
- Fanuel, M., Garajova, S., Ropartz, D., McGregor, N., Brumer, H., Rogniaux, H. and Berrin, J.-G. (2017) The *Podospira anserina* lytic polysaccharide monooxygenase PaLPMO9H catalyzes oxidative cleavage of diverse plant cell wall matrix glycans. *Biotechnology for biofuels* **10**, 63.
- Filiatrault-Chastel, C., Navarro, D., Haon, M., Grisel, S., Herpoël-Gimbert, I., Chevret, D., Fanuel, M., Henrissat, B., Heiss-Blanquet, S., Margeot, A. and Berrin, J.-G. (2019) AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes. *Biotechnology for biofuels* **12**, 55.
- Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R. A., Henrissat, B., Martínez, A. T., Otilar, R., Spatafora, J. W., Yadav, J. S., Aerts, A., Benoit, I., Boyd, A., Carlson, A., Copeland, A., Coutinho, P. M., Vries, R. P. de, Ferreira, P., Findley, K., Foster, B., Gaskell, J., Glotzer, D., Górecki, P., Heitman, J., Hesse, C., Hori, C., Igarashi, K., Jurgens, J. A., Kallen, N., Kersten, P., Kohler, A., Kües, U., Kumar, T. K. A., Kuo, A., LaButti, K., Larrondo, L. F., Lindquist, E., Ling, A., Lombard, V., Lucas, S., Lundell, T., Martin, R., McLaughlin, D. J., Morgenstern, I., Morin, E., Murat, C., Nagy, L. G., Nolan, M., Ohm, R. A., Patyshakuliyeva, A., Rokas, A., Ruiz-Dueñas, F. J., Sabat, G., Salamov, A., Samejima, M., Schmutz, J., Slot, J. C., St John, F., Stenlid, J., Sun, H., Sun, S., Syed, K., Tsang, A., Wiebenga, A., Young, D., Pisabarro, A., Eastwood, D. C., Martin, F., Cullen, D., Grigoriev, I. V. and Hibbett, D. S. (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**, 1715–1719.
- Forsberg, Z., Bissaro, B., Gullesen, J., Dalhus, B., Vaaje-Kolstad, G. and Eijsink, V. G. H. (2018) Structural determinants of bacterial lytic polysaccharide monooxygenase functionality. *The Journal of biological chemistry* **293**, 1397–1412.
- Forsberg, Z., Mackenzie, A. K., Sørli, M., Røhr, Å. K., Helland, R., Arvai, A. S., Vaaje-Kolstad, G. and Eijsink, V. G. H. (2014a) Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 8446–8451.
- Forsberg, Z., Røhr, A. K., Mekasha, S., Andersson, K. K., Eijsink, V. G. H., Vaaje-Kolstad, G. and Sørli, M. (2014b) Comparative study of two chitin-active and two cellulose-active AA10-type lytic polysaccharide monooxygenases. *Biochemistry* **53**, 1647–1656.
- Forsberg, Z., Vaaje-Kolstad, G., Westereng, B., Bunæs, A. C., Stenstrøm, Y., MacKenzie, A., Sørli, M., Horn, S. J. and Eijsink, V. G. H. (2011) Cleavage of cellulose by a CBM33 protein. *Protein science a publication of the Protein Society* **20**, 1479–1483.
- Frandsen, K. E. H. and Lo Leggio, L. (2016) Lytic polysaccharide monooxygenases: A crystallographer's view on a new class of biomass-degrading enzymes. *IUCrJ* **3**, 448–467.
- Frandsen, K. E. H., Simmons, T. J., Dupree, P., Poulsen, J.-C. N., Hemsworth, G. R., Ciano, L., Johnston, E. M., Tovborg, M., Johansen, K. S., Freiesleben, P. von, Marmuse, L., Fort, S., Cottaz, S., Driguez, H., Henrissat, B., Lenfant, N., Tuna, F., Baldansuren, A., Davies, G. J., Lo Leggio, L. and Walton, P. H. (2016) The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nature Chemical Biology* **12**, 298–303.
- Frommhagen, M., Koetsier, M. J., Westphal, A. H., Visser, J., Hinz, S. W. A., Vincken, J.-P., van Berkel, W. J. H., Kabel, M. A. and Gruppen, H. (2016) Lytic polysaccharide monooxygenases from *Myceliophthora thermophila* C1 differ in substrate preference and reducing agent specificity. *Biotechnology for biofuels* **9**, 186.
- Frommhagen, M., Sforza, S., Westphal, A. H., Visser, J., Hinz, S. W. A., Koetsier, M. J., van Berkel, W. J. H., Gruppen, H. and Kabel, M. A. (2015) Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase. *Biotechnology for biofuels* **8**, 101.
- Frommhagen, M., Westphal, A. H., van Berkel, H. W. J. and Kabel, M. A. (2018) Distinct Substrate Specificities and Electron-Donating Systems of Fungal Lytic Polysaccharide Monooxygenases. *Frontiers in microbiology* **9**.
- Fry, S. C., York, W. S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J.-P., Kato, Y., Lorences, E. P., Maclachlan, G. A., McNeil, M., Mort, A. J., Reid, J. S. G., Seitz, H. U., Selvendran, R. R., Voragen, A. G. J. and White, A. R. (1993) An unambiguous nomenclature for xyloglucan-derived oligosaccharides. *Physiologia Plantarum* **89**, 1–3.

- Garajova, S., Mathieu, Y., Beccia, M. R., Bennati-Granier, C., Biaso, F., Fanuel, M., Ropartz, D., Guigliarelli, B., Record, E., Rogniaux, H., Henrissat, B. and Berrin, J.-G. (2016) Single-domain flavoenzymes trigger lytic polysaccharide monoxygenases for oxidative degradation of cellulose. *Scientific reports* **6**, 28276.
- García, A., González Alriols, M. and Labidi, J. (2014) Evaluation of different lignocellulosic raw materials as potential alternative feedstocks in biorefinery processes. *Industrial Crops and Products* **53**, 102–110.
- Gascuel, O. (1997) BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Molecular biology and evolution* **14**, 685–695.
- Gilbert, H. J., Knox, J. P. and Boraston, A. B. (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Current opinion in structural biology* **23**, 669–677.
- Gottschalk, L. M. F., Oliveira, R. A. and Bon, E. P. d. S. (2010) Cellulases, xylanases, β -glucosidase and ferulic acid esterase produced by *Trichoderma* and *Aspergillus* act synergistically in the hydrolysis of sugarcane bagasse. *Biochemical Engineering Journal* **51**, 72–78.
- Gupta, A. and Verma, J. P. (2015) Sustainable bio-ethanol production from agro-residues: A review. *Renewable and Sustainable Energy Reviews* **41**, 550–567.
- Gupta, V. K., Steindorff, A. S., Paula, R. G. de, Silva-Rocha, R., Mach-Aigner, A. R., Mach, R. L. and Silva, R. N. (2016) The Post-genomic Era of *Trichoderma reesei*: What's Next? *Trends in Biotechnology* **34**, 970–982.
- Habibi, Y., Lucia, L. A. and Rojas, O. J. (2010) Cellulose Nanocrystals: Chemistry, Self-Assembly, and Applications. *Chemical Reviews* **110**, 3479–3500.
- Häkkinen, M., Arvas, M., Oja, M., Aro, N., Penttilä, M., Saloheimo, M. and Pakula, T. M. (2012) Re-annotation of the CAZy genes of *Trichoderma reesei* and transcription in the presence of lignocellulosic substrates. *Microbial Cell Factories* **11**, 134.
- Hamby, S. E. and Hirst, J. D. (2008) Prediction of glycosylation sites using random forests. *BMC Bioinformatics* **9**, 500.
- Hamed, S. A. M. (2013) In-vitro studies on wood degradation in soil by soft-rot fungi: *Aspergillus niger* and *Penicillium chrysogenum*. *International Biodeterioration & Biodegradation* **78**, 98–102.
- Haon, M., Grisel, S., Navarro, D., Gruet, A., Berrin, J.-G. and Bignon, C. (2015) Recombinant protein production facility for fungal biomass-degrading enzymes using the yeast *Pichia pastoris*. *Frontiers in microbiology* **6**.
- Harholt, J., Suttangkakul, A. and Scheller, H. V. (2010) Biosynthesis of Pectin. *Plant Physiology* **153**, 384–395.
- Harris, P. V., Welner, D., McFarland, K. C., Re, E., Navarro Poulsen, J.-C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S. and Lo Leggio, L. (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: Structure and function of a large, enigmatic family. *Biochemistry* **49**, 3305–3316.
- Hasunuma, T., Okazaki, F., Okai, N., Hara, K. Y., Ishii, J. and Kondo, A. (2013) A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. *Bioresource Technology* **135**, 513–522.
- He, J., Wu, A.-m., Chen, D., Yu, B., Mao, X., Zheng, P., Yu, J. and Tian, G. (2014) Cost-effective lignocellulolytic enzyme production by *Trichoderma reesei* on a cane molasses medium. *Biotechnology for biofuels* **7**, 43.
- Heinzelman, P., Komor, R., Kanaan, A., Romero, P., Yu, X., Mohler, S., Snow, C. and Arnold, F. (2010) Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination. *Protein Engineering, Design and Selection* **23**, 871–880.
- Heinzelman, P., Snow, C. D., Smith, M. A., Yu, X., Kannan, A., Boulware, K., Villalobos, A., Govindarajan, S., Minshull, J. and Arnold, F. H. (2009a) SCHEMA recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability. *The Journal of biological chemistry* **284**, 26229–26233.
- Heinzelman, P., Snow, C. D., Wu, I., Nguyen, C., Villalobos, A., Govindarajan, S., Minshull, J. and Arnold, F. H. (2009b) A family of thermostable fungal cellulases created by structure-guided recombination. *PNAS* **106**, 5610–5615.
- Hemsworth, G. R., Ciano, L., Davies, G. J. and Walton, P. H. (2018) Production and spectroscopic characterization of lytic polysaccharide monoxygenases. *Methods in Enzymology* **613**, 63–90.
- Hemsworth, G. R., Henrissat, B., Davies, G. J. and Walton, P. H. (2014) Discovery and characterization of a new family of lytic polysaccharide monoxygenases. *Nature Chemical Biology* **10**, 122.
- Hemsworth, G. R., Johnston, E. M., Davies, G. J. and Walton, P. H. (2015) Lytic Polysaccharide Monoxygenases in Biomass Conversion. *Trends in Biotechnology* **33**, 747–761.
- Henikoff, S. and Henikoff, J. G. (1992) Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 10915–10919.
- Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* **280**, 309–316.
- Herpoël-Gimbert, I., Margeot, A., Dolla, A., Jan, G., Mollé, D., Lignon, S., Mathis, H., Sigoillot, J.-C., Monot, F. and Asther, M. (2008) Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. *Biotechnology for biofuels* **1**, 18.

- Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., Huhndorf, S., James, T., Kirk, P. M., Lücking, R., Thorsten Lumbsch, H., Lutzoni, F., Matheny, P. B., McLaughlin, D. J., Powell, M. J., Redhead, S., Schoch, C. L., Spatafora, J. W., Stalpers, J. A., Vilgalys, R., Aime, M. C., Aptroot, A., Bauer, R., Begerow, D., Benny, G. L., Castlebury, L. A., Crous, P. W., Dai, Y.-C., Gams, W., Geiser, D. M., Griffith, G. W., Gueidan, C., Hawksworth, D. L., Hestmark, G., Hosaka, K., Humber, R. A., Hyde, K. D., Ironside, J. E., Kõljalg, U., Kurtzman, C. P., Larsson, K.-H., Lichtwardt, R., Longcore, J., Miadlikowska, J., Miller, A., Moncalvo, J.-M., Mozley-Standridge, S., Oberwinkler, F., Parmasto, E., Reeb, V., Rogers, J. D., Roux, C., Ryvarden, L., Sampaio, J. P., Schüssler, A., Sugiyama, J., Thorn, R. G., Tibell, L., Untereiner, W. A., Walker, C., Wang, Z., Weir, A., Weiss, M., White, M. M., Winka, K., Yao, Y.-J. and Zhang, N. (2007) A higher-level phylogenetic classification of the Fungi. *Mycological research* **111**, 509–547.
- Himmel, M. E., Xu, Q., Luo, Y., Ding, S.-Y., Lamed, R. and Bayer, E. A. (2010) Microbial enzyme systems for biomass conversion: Emerging paradigms. *Biofuels* **1**, 323–341.
- Ho, D. P., Ngo, H. H. and Guo, W. (2014) A mini review on renewable sources for biofuel. *Bioresource Technology* **169**, 742–749.
- Horn, S. J., Vaaje-Kolstad, G., Westereng, B. and Eijsink, V. (2012) Novel enzymes for the degradation of cellulose. *Biotechnology for biofuels* **5**, 45.
- Hu, J., Arantes, V., Pribowo, A. and Saddler, J. N. (2013) The synergistic action of accessory enzymes enhances the hydrolytic potential of a “cellulase mixture” but is highly substrate specific. *Biotechnology for biofuels* **6**, 112.
- Huson, D. H. and Scornavacca, C. (2012) Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic biology* **61**, 1061–1067.
- Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttilä, M., Ando, T. and Samejima, M. (2011) Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* **333**, 1279–1282.
- Iiyama, K., Lam, T. and Stone, B. A. (1994) Covalent Cross-Links in the Cell Wall. *Plant Physiology* **104**, 315–320.
- Ilmén, M., Thrane, C. and Penttilä, M. (1996) The glucose repressor gene cre1 of *Trichoderma*: Isolation and expression of a full-length and a truncated mutant form. *Molecular & general genetics* **251**, 451–460.
- Isaksen, T., Westereng, B., Aachmann, F. L., Agger, J. W., Kracher, D., Kittl, R., Ludwig, R., Haltrich, D., Eijsink, V. G. H. and Horn, S. J. (2014) A C4-oxidizing Lytic Polysaccharide Monooxygenase Cleaving Both Cellulose and Cello-oligosaccharides. *The Journal of biological chemistry* **289**, 2632–2642.
- Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J. and Mann, M. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular and Cellular Proteomics* **4**, 1265–1272.
- Johansen, K. S. (2016) Discovery and industrial applications of lytic polysaccharide mono-oxygenases. *Biochemical Society transactions* **44**, 143–149.
- Jørgensen, H., Kristensen, J. B. and Felby, C. (2007a) Enzymatic conversion of lignocellulose into fermentable sugars: Challenges and opportunities. *Biofuels, Bioproducts and Biorefining* **1**, 119–134.
- Jørgensen, H., Vibe-Pedersen, J., Larsen, J. and Felby, C. (2007b) Liquefaction of lignocellulose at high-solids concentrations. *Biotechnology and Bioengineering* **96**, 862–870.
- Jun, H., Guangye, H. and Daiwen, C. (2013) Insights into enzyme secretion by filamentous fungi: Comparative proteome analysis of *Trichoderma reesei* grown on different carbon sources. *Journal of Proteomics* **89**, 191–201.
- Kadowaki, M. A. S., Várnai, A., Jameson, J.-K., T Leite, A. E., Costa-Filho, A. J., Kumagai, P. S., Prade, R. A., Polikarpov, I. and Eijsink, V. G. H. (2018) Functional characterization of a lytic polysaccharide monooxygenase from the thermophilic fungus *Myceliophthora thermophila*. *PLoS ONE* **13**.
- Kämper, J., Kahmann, R., Bölker, M., Ma, L.-J., Brefort, T., Saville, B. J., Banuett, F., Kronstad, J. W., Gold, S. E., Müller, O., Perlín, M. H., Wösten, H. A. B., Vries, R. de, Ruiz-Herrera, J., Reynaga-Peña, C. G., Snetselaar, K., McCann, M., Pérez-Martín, J., Feldbrügge, M., Basse, C. W., Steinberg, G., Ibeas, J. I., Holloman, W., Guzman, P., Farman, M., Stajich, J. E., Sentandreu, R., González-Prieto, J. M., Kennell, J. C., Molina, L., Schirawski, J., Mendoza-Mendoza, A., Greilinger, D., Münch, K., Rössel, N., Scherer, M., Vraneš, M., Ladendorf, O., Vincon, V., Fuchs, U., Sandrock, B., Meng, S., Ho, E. C. H., Cahill, M. J., Boyce, K. J., Klose, J., Klosterman, S. J., Deelstra, H. J., Ortiz-Castellanos, L., Li, W., Sanchez-Alonso, P., Schreier, P. H., Häuser-Hahn, I., Vaupel, M., Koopmann, E., Friedrich, G., Voss, H., Schlüter, T., Margolis, J., Platt, D., Swimmer, C., Gnirke, A., Chen, F., Vysotskaia, V., Mannhaupt, G., Güldener, U., Münsterkötter, M., Haase, D., Oesterheld, M., Mewes, H.-W., Mauceli, E. W., DeCaprio, D., Wade, C. M., Butler, J., Young, S., Jaffe, D. B., Calvo, S., Nusbaum, C., Galagan, J. and Birren, B. W. (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**, 97–101.
- Karkehabadi, S., Hansson, H., Kim, S., Piens, K., Mitchinson, C. and Sandgren, M. (2008) The First Structure of a Glycoside Hydrolase Family 61 Member, Cel61B from *Hypocrea jecorina*, at 1.6 Å Resolution. *Journal of Molecular Biology* **383**, 144–154.

- Karnaouri, A., Muraleedharan, M. N., Dimarogona, M., Topakas, E., Rova, U., Sandgren, M. and Christakopoulos, P. (2017) Recombinant expression of thermostable processive MtEG5 endoglucanase and its synergism with MtLPMO from *Myceliophthora thermophila* during the hydrolysis of lignocellulosic substrates. *Biotechnology for biofuels* **10**, 126.
- Katsimpouras, C., Dedes, G., Thomaidis, N. S. and Topakas, E. (2019) A novel fungal GH30 xylanase with xylobiohydrolase auxiliary activity. *Biotechnology for biofuels* **12**, 120.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845.
- Kim, K., Rhee, S. G. and Stadtman, E. R. (1985) Nonenzymatic cleavage of proteins by reactive oxygen species generated by dithiothreitol and iron. *The Journal of biological chemistry* **260**, 15394–15397.
- Kim, S., Ståhlberg, J., Sandgren, M., Paton, R. S. and Beckham, G. T. (2014) Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxy, oxygen-rebound mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 149–154.
- Kittl, R., Kracher, D., Burgstaller, D., Haltrich, D. and Ludwig, R. (2012) Production of four *Neurospora crassa* lytic polysaccharide monooxygenases in *Pichia pastoris* monitored by a fluorimetric assay. *Biotechnology for biofuels* **5**, 79.
- Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A. and Blanch, H. W. (2012) The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and Bioengineering* **109**, 1083–1087.
- Komor, R. S., Romero, P. A., Xie, C. B. and Arnold, F. H. (2012) Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods. *Protein Engineering, Design and Selection* **25**, 827–833.
- Kracher, D., Oros, D., Yao, W., Preims, M., Rezić, I., Haltrich, D., Rezić, T. and Ludwig, R. (2014) Fungal secretomes enhance sugar beet pulp hydrolysis. *Biotechnology Journal* **9**, 483–492.
- Kracher, D., Scheiblbrandner, S., Felice, A. K. G., Breslmayr, E., Preims, M., Ludwicka, K., Haltrich, D., Eijsink, V. G. H. and Ludwig, R. (2016) Extracellular electron transfer systems fuel cellulose oxidative degradation. *Science*, 1098-1101.
- Kristensen, J. B., Felby, C. and Jørgensen, H. (2009) Yield-determining factors in high-solids enzymatic hydrolysis of lignocellulose. *Biotechnology for biofuels* **2**, 11.
- Kubicek, C. P. (2013) Systems biological approaches towards understanding cellulase production by *Trichoderma reesei*. *Journal of Biotechnology* **163**, 133–142.
- Kubicek, C. P., Mikus, M., Schuster, A., Schmoll, M. and Seiboth, B. (2009) Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. *Biotechnology for biofuels* **2**, 1–14.
- Langston, J. A., Shaghasi, T., Abbate, E., Xu, F., Vlasenko, E. and Sweeney, M. D. (2011) Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Applied and environmental microbiology* **77**, 7007–7015.
- Lantz, S. E., Goedegebuur, F., Hommes, R., Kaper, T., Kelemen, B. R., Mitchinson, C., Wallace, L., Ståhlberg, J. and Larenas, E. A. (2010) *Hypocrea jecorina* CEL6A protein engineering. *Biotechnology for biofuels* **3**, 20.
- Lavoine, N., Desloges, I., Dufresne, A. and Bras, J. (2012) Microfibrillated cellulose – Its barrier properties and applications in cellulosic materials: A review. *Carbohydrate Polymers* **90**, 735–764.
- Le Crom, S., Schackwitz, W., Pennacchio, L., Magnuson, J. K., Culley, D. E., Collett, J. R., Martin, J., Druzhinina, I. S., Mathis, H., Monot, F., Seiboth, B., Cherry, B., Rey, M., Berka, R., Kubicek, C. P., Baker, S. E. and Margeot, A. (2009) Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei* using massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 16151–16156.
- Lehmann, L., Rønneest, N. P., Jørgensen, C. I., Olsson, L., Stocks, S. M., Jørgensen, H. S. and Hobley, T. (2016) Linking hydrolysis performance to *Trichoderma reesei* cellulolytic enzyme profile. *Biotechnology and Bioengineering* **113**, 1001–1010.
- Lenfant, N., Hainaut, M., Terrapon, N., Drula, E., Lombard, V. and Henrissat, B. (2017) A bioinformatics analysis of 3400 lytic polysaccharide oxidases from family AA9. *Carbohydrate research* **448**, 166–174.
- Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M. and Henrissat, B. (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for biofuels* **6**, 41.
- Liu, R., Chen, L., Jiang, Y., Zhou, Z. and Zou, G. (2015) Efficient genome editing in filamentous fungus *Trichoderma reesei* using the CRISPR/Cas9 system. *Cell Discovery* **1**.
- Lo Leggio, L., Simmons, T. J., Poulsen, J.-C. N., Frandsen, K. E. H., Hemsworth, G. R., Stringer, M. A., Freiesleben, P. von, Tovborg, M., Johansen, K. S., Maria, L. D., Harris, P. V., Soong, C.-L., Dupree, P., Tryfona, T., Lenfant, N., Henrissat, B., Davies, G. J. and Walton, P. H. (2015) Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. *Nature communications* **6**, 5961.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* **42**, D490-D495.

- Loose, J. S. M., Arntzen, M. Ø., Bissaro, B., Ludwig, R., Eijsink, V. G. H. and Vaaje-Kolstad, G. (2018) Multipoint Precision Binding of Substrate Protects Lytic Polysaccharide Monooxygenases from Self-Destructive Off-Pathway Processes. *Biochemistry* **57**, 4114–4124.
- Lynd, L. R., Weimer, P. J., van Zyl, W. H. and Pretorius, I. S. (2002) Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiology and Molecular Biology Reviews* **66**, 506–577.
- Margeot, A., Hahn-Hagerdal, B., Edlund, M., Slade, R. and Monot, F. (2009) New improvements for lignocellulosic ethanol. *Current Opinion in Biotechnology* **20**, 372–380.
- Martin, F., Kohler, A., Murat, C., Veneault-Fourrey, C. and Hibbett, D. S. (2016) Unearthing the roots of ectomycorrhizal symbioses. *Nature Reviews Microbiology* **14**, 760–773.
- Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., Chapman, J., Chertkov, O., Coutinho, P. M., Cullen, D., Danchin, E. G. J., Grigoriev, I. V., Harris, P., Jackson, M., Kubicek, C. P., Han, C. S., Ho, I., Larrondo, L. F., Leon, A. L. d., Magnuson, J. K., Merino, S., Misra, M., Nelson, B., Putnam, N., Robbertse, B., Salamov, A. A., Schmoll, M., Terry, A., Thayer, N., Westerholm-Parvinen, A., Schoch, C. L., Yao, J., Barabote, R., Nelson, M. A., Detter, C., Bruce, D., Kuske, C. R., Xie, G., Richardson, P., Rokhsar, D. S., Lucas, S. M., Rubin, E. M., Dunn-Coleman, N., Ward, M. and Brettin, T. S. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnology* **26**, 553.
- Mattanovich, D., Graf, A., Stadlmann, J., Dragosits, M., Redl, A., Maurer, M., Kleinheinz, M., Sauer, M., Altmann, F. and Gasser, B. (2009) Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. *Microbial Cell Factories* **8**, 1–13.
- McMahon, H. E., Mangé, A., Nishida, N., Créminon, C., Casanova, D. and Lehmann, S. (2001) Cleavage of the amino terminus of the prion protein by reactive oxygen species. *The Journal of biological chemistry* **276**, 2286–2291.
- Menon, V. and Rao, M. (2012) Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept. *Progress in Energy and Combustion Science* **38**, 522–550.
- Metsalu, T. and Vilo, J. (2015) ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic acids research* **43**, W566–W570.
- Morgenstern, I., Powlowski, J. and Tsang, A. (2014) Fungal cellulose degradation by oxidative enzymes: From dysfunctional GH61 family to powerful lytic polysaccharide monooxygenase family. *Briefings in Functional Genomics* **13**, 471–481.
- Mukarakate, C., Mittal, A., Ciesielski, P. N., Budhi, S., Thompson, L., Iisa, K., Nimlos, M. R. and Donohoe, B. S. (2016) Influence of Crystal Allomorph and Crystallinity on the Products and Behavior of Cellulose during Fast Pyrolysis. *ACS Sustainable Chemistry & Engineering* **4**, 4662–4674.
- Müller, G., Chylenski, P., Bissaro, B., Eijsink, V. G. H. and Horn, S. J. (2018) The impact of hydrogen peroxide supply on LPMO activity and overall saccharification efficiency of a commercial cellulase cocktail. *Biotechnology for biofuels* **11**, 209.
- Müller, G., Várnai, A., Johansen, K. S., Eijsink, V. G. H. and Horn, S. J. (2015) Harnessing the potential of LPMO-containing cellulase cocktails poses new demands on processing conditions. *Biotechnology for biofuels* **8**, 187.
- Nakamichi, Y., Fouquet, T., Ito, S., Watanabe, M., Matsushika, A. and Inoue, H. (2019) Structural and functional characterization of a bifunctional GH30-7 xylanase B from the filamentous fungus *Talaromyces cellulolyticus*. *The Journal of biological chemistry* **294**, 4065–4078.
- Navarro, D., Couturier, M., Damasceno da Silva, G. G., Berrin, J.-G., Rouau, X., Asther, M. and Bignon, C. (2010) Automated assay for screening the enzymatic release of reducing sugars from micronized biomass. *Microbial Cell Factories* **9**, 58.
- Navarro, D., Rosso, M.-N., Haon, M., Olivé, C., Bonnin, E., Lesage-Meessen, L., Chevret, D., Coutinho, P. M., Henrissat, B. and Berrin, J.-G. (2014) Fast solubilization of recalcitrant cellulosic biomass by the basidiomycete fungus *Laetisaria arvalis* involves successive secretion of oxidative and hydrolytic enzymes. *Biotechnology for biofuels* **7**, 143.
- Nekiunaite, L., Arntzen, M. Ø., Svensson, B., Vaaje-Kolstad, G. and Abou Hachem, M. (2016) Lytic polysaccharide monooxygenases and other oxidative enzymes are abundantly secreted by *Aspergillus nidulans* grown on different starches. *Biotechnology for biofuels* **9**, 187.
- NREL (2014). *NREL 2012 Achievement of Ethanol Cost Targets: Biochemical Ethanol Fermentation via Dilute-Acid Pretreatment and Enzymatic Hydrolysis of Corn Stover*.
- Olsen, S. N., Borch, K., Cruys-Bagger, N. and Westh, P. (2014) The role of product inhibition as a yield-determining factor in enzymatic high-solid hydrolysis of pretreated corn stover. *Applied Biochemistry and Biotechnology* **174**, 146–155.
- O'Sullivan, A. C. (1997) Cellulose: the structure slowly unravels. *Cellulose* **4**, 173–207.
- Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J. and Beckham, G. T. (2015) Fungal Cellulases. *Chemical Reviews* **115**, 1308–1448.
- Peciulyte, A., Pisano, M., Vries, R. P. de and Olsson, L. (2017) Hydrolytic potential of five fungal supernatants to enhance a commercial enzyme cocktail. *Biotechnol Letters* **39**, 1403–1411.

- Peisach, J. and Blumberg, W. E. (1974) Structural implications derived from the analysis of electron paramagnetic resonance spectra of natural and artificial copper proteins. *Archives of Biochemistry and Biophysics* **165**, 691–708.
- Peterson, R. and Nevalainen, H. (2012) *Trichoderma reesei* RUT-C30 – thirty years of strain improvement. *Microbiology* **158**, 58–68.
- Petrovic, D. M., Bissaro, B., Chylenski, P., Skaugen, M., Sørli, M., Jensen, M. S., Aachmann, F. L., Courtade, G., Várnai, A. and Eijsink, V. G. H. (2018) Methylation of the N-terminal histidine protects a lytic polysaccharide monoxygenase from auto-oxidative inactivation. *Protein science a publication of the Protein Society*.
- Phillips, C. M., Beeson, W. T., Cate, J. H. and Marletta, M. A. (2011) Cellobiose Dehydrogenase and a Copper-Dependent Polysaccharide Monoxygenase Potentiate Cellulose Degradation by *Neurospora crassa*. *ACS chemical biology* **6**, 1399–1406.
- Poggi-Parodi, D., Bidard, F., Pirayre, A., Portnoy, T., Blugeon, C., Seiboth, B., Kubicek, C. P., Le Crom, S. and Margeot, A. (2014) Kinetic transcriptome analysis reveals an essentially intact induction system in a cellulase hyper-producer *Trichoderma reesei* strain. *Biotechnology for biofuels* **7**, 173.
- Poidevin, L., Berrin, J.-G., Bennati-Granier, C., Levasseur, A., Herpoël-Gimbert, I., Chevret, D., Coutinho, P. M., Henrissat, B., Heiss-Blanquet, S. and Record, E. (2014) Comparative analyses of *Podospora anserina* secretomes reveal a large array of lignocellulose-active enzymes. *Applied microbiology and biotechnology* **98**, 7457–7469.
- Portnoy, T., Margeot, A., Linke, R., Atanasova, L., Fekete, E., Sándor, E., Hartl, L., Karaffa, L., Druzhinina, I. S., Seiboth, B., Le Crom, S. and Kubicek, C. P. (2011a) The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: A master regulator of carbon assimilation. *BMC Genomics* **12**, 269.
- Portnoy, T., Margeot, A., Seidl-Seiboth, V., Le Crom, S., Ben Chaabane, F., Linke, R., Seiboth, B. and Kubicek, C. P. (2011b) Differential Regulation of the Cellulase Transcription Factors XYR1, ACE2, and ACE1 in *Trichoderma reesei* Strains Producing High and Low Levels of Cellulase. *Eukaryotic Cell* **10**, 262–271.
- Qing, Q. and Wyman, C. E. (2011) Supplementation with xylanase and β -xylosidase to reduce xylo-oligomer and xylan inhibition of enzymatic hydrolysis of cellulose and pretreated corn stover. *Biotechnology for biofuels* **4**, 18.
- Quinlan, R. J., Sweeney, M. D., Lo Leggio, L., Otten, H., Poulsen, J.-C. N., Johansen, K. S., Krogh, K. B. R. M., Jørgensen, C. I., Tovborg, M., Anthonsen, A., Tryfona, T., Walter, C. P., Dupree, P., Xu, F., Davies, G. J. and Walton, P. H. (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 15079–15084.
- Quiroz-Castañeda, R. E. and Folch-Mallol, J. L. (2014) Hydrolysis of Biomass Mediated by Cellulases for the Production of Sugars. In: *Sustainable degradation of lignocellulosic biomass: Techniques, applications and commercialization* (Chandel, A.K. and Silva, S.S.d., eds). Rijeka, Croatia: InTech.
- Rahman, Z., Shida, Y., Furukawa, T., Suzuki, Y., Okada, H., Ogasawara, W. and Morikawa, Y. (2009) Application of *Trichoderma reesei* Cellulase and Xylanase Promoters through Homologous Recombination for Enhanced Production of Extracellular β -Glucosidase I. *Bioscience, Biotechnology, and Biochemistry* **73**, 1083–1089.
- Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P. F., Marita, J. M., Hatfield, R. D., Ralph, S. A., Christensen, J. H. and Boerjan, W. (2004) Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl- propanoids. *Phytochemistry Reviews* **3**, 29–60.
- Rassinger, A., Gacek-Matthews, A., Strauss, J., Mach, R. L. and Mach-Aigner, A. R. (2018) Truncation of the transcriptional repressor protein Cre1 in *Trichoderma reesei* Rut-C30 turns it into an activator. *Fungal Biology and Biotechnology* **5**, 15.
- Ravalason, H., Grisel, S., Chevret, D., Favel, A., Berrin, J.-G., Sigoillot, J.-C. and Herpoël-Gimbert, I. (2012) *Fusarium verticillioides* secretome as a source of auxiliary enzymes to enhance saccharification of wheat straw. *Bioresource Technology* **114**, 589–596.
- Riley, R., Salamov, A. A., Brown, D. W., Nagy, L. G., Floudas, D., Held, B. W., Levasseur, A., Lombard, V., Morin, E., Otilar, R., Lindquist, E. A., Sun, H., LaButti, K. M., Schmutz, J., Jabbour, D., Luo, H., Baker, S. E., Pisabarro, A. G., Walton, J. D., Blanchette, R. A., Henrissat, B., Martin, F., Cullen, D., Hibbett, D. S. and Grigoriev, I. V. (2014) Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9923–9928.
- Rojas-Jiménez, K. and Hernández, M. (2015) Isolation of Fungi and Bacteria Associated with the Guts of Tropical Wood-Feeding Coleoptera and Determination of Their Lignocellulolytic Activities. *International journal of microbiology* **2015**.
- Rosgaard, L., Pedersen, S., Cherry, J. R., Harris, P. and Meyer, A. S. (2006) Efficiency of New Fungal Cellulase Systems in Boosting Enzymatic Degradation of Barley Straw Lignocellulose. *Biotechnology Progress* **22**, 493–498.
- Rytioja, J., Hildén, K., Yuzon, J., Hatakka, A., Vries, R. P. de and Mäkelä, M. R. (2014) Plant-Polysaccharide-Degrading Enzymes from Basidiomycetes. *Microbiology and Molecular Biology Reviews* **78**, 614–649.
- Sabbadin, F., Hemsworth, G. R., Ciano, L., Henrissat, B., Dupree, P., Tryfona, T., Marques, R. D. S., Sweeney, S. T., Besser, K., Elias, L., Pesante, G., Li, Y., Dowle, A. A., Bates, R., Gomez, L. D., Simister, R., Davies, G. J., Walton, P. H., Bruce, N. C.

- and McQueen-Mason, S. J. (2018) An ancient family of lytic polysaccharide monoxygenases with roles in arthropod development and biomass digestion. *Nature communications* **9**, 756.
- Saloheimo, M., Paloheimo, M., Hakola, S., Pere, J., Swanson, B., Nyysönen, E., Bhatia, A., Ward, M. and Penttilä, M. (2002) Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. *European journal of biochemistry* **269**, 4202–4211.
- Scheller, H. V. and Ulvskov, P. (2010) Hemicelluloses. *Annual Review of Plant Biology* **61**, 263–289.
- Seidl, V. and Seiboth, B. (2010) *Trichoderma reesei*: Genetic approaches to improving strain efficiency. *Biofuels* **1**, 343–354.
- Selig, M. J., Vuong, T. V., Gudmundsson, M., Forsberg, Z., Westereng, B., Felby, C. and Master, E. R. (2015) Modified cellobiohydrolase–cellulose interactions following treatment with lytic polysaccharide monoxygenase CelS2 (ScLPMO10C) observed by QCM-D. *Cellulose* **22**, 2263–2270.
- Shoemaker, S., Schweickart, V., Ladner, M., Gelfand, D., Kwok, S., Myambo, K. and Innis, M. (1983) Molecular Cloning of Exo–Cellobiohydrolase I Derived from *Trichoderma Reesei* Strain L27. *Nature Biotechnology* **1**, 691–696.
- Shrestha, P., Ibáñez, A. B., Bauer, S., Glassman, S. I., Szaro, T. M., Bruns, T. D. and Taylor, J. W. (2015) Fungi isolated from *Miscanthus* and sugarcane: Biomass conversion, fungal enzymes, and hydrolysis of plant cell wall polymers. *Biotechnology for biofuels* **8**, 38.
- Silar, P. (2016). *Protistes Eucaryotes: Origine, évolution et biologie des microbes eucaryotes*. California: Creative Commons.
- Sills, D. L. and Gossett, J. M. (2011) Assessment of commercial hemicellulases for saccharification of alkaline pretreated perennial biomass. *Bioresource Technology* **102**, 1389–1398.
- Silva-Rocha, R., Castro, L. d. S., Antoniêto, A. C. C., Guazzaroni, M.-E., Persinoti, G. F. and Silva, R. N. (2014) Deciphering the cis-regulatory elements for XYR1 and CRE1 regulators in *Trichoderma reesei*. *PLoS ONE* **9**.
- Simmons, T. J., Frandsen, K. E. H., Ciano, L., Tryfona, T., Lenfant, N., Poulsen, J. C., Wilson, L. F. L., Tandrup, T., Tovborg, M., Schnorr, K., Johansen, K. S., Henrissat, B., Walton, P. H., Lo Leggio, L. and Dupree, P. (2017) Structural and electronic determinants of lytic polysaccharide monoxygenase reactivity on polysaccharide substrates. *Nature communications* **8**, 1064.
- Sista Kameshwar, A. K. and Qin, W. (2018) Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi. *Mycology* **9**, 93–105.
- Song, B., Li, B., Wang, X., Shen, W., Park, S., Collings, C., Feng, A., Smith, S. J., Walton, J. D. and Ding, S.-Y. (2018) Real-time imaging reveals that lytic polysaccharide monoxygenase promotes cellulase activity by increasing cellulose accessibility. *Biotechnology for biofuels* **11**, 41.
- Sreedharan, S., Prakasan, P., Sasidharan, S. and Sailas, B. (2016) An Overview on Fungal Cellulases with an Industrial Perspective. *Journal of Nutrition & Food Sciences* **6**, 1–13.
- Stadtman, E. R. (1993) Oxidation of free amino acids and amino acid residues in proteins by radiolysis and by metal-catalyzed reactions. *Annual review of biochemistry* **62**, 797–821.
- Ståhlberg, J., Johansson, G. and Pettersson, G. (1993) *Trichoderma reesei* has no true exo-cellulase: All intact and truncated cellulases produce new reducing end groups on cellulose. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1157**, 107–113.
- Sticklen, M. B. (2008) Plant genetic engineering for biofuel production: Towards affordable cellulosic ethanol. *Nature Reviews Genetics* **9**, 433.
- Stoll, S. and Schweiger, A. (2006) EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *Journal of magnetic resonance* **178**, 42–55.
- Sun, Y. and Cheng, J. (2002) Hydrolysis of lignocellulosic materials for ethanol production: A review. *Bioresource Technology* **83**, 1–11.
- Świątek, K., Lewandowska, M., Świątek, M., Bednarski, W. and Brzozowski, B. (2014) The improvement of enzymatic hydrolysis efficiency of rape straw and *Miscanthus giganteus* polysaccharides. *Bioresource Technology* **151**, 323–331.
- Tabka, M. G., Herpoël-Gimbert, I., Monod, F., Asther, M. and Sigoillot, J. C. (2006) Enzymatic saccharification of wheat straw for bioethanol production by a combined cellulase xylanase and feruloyl esterase treatment. *Enzyme and Microbial Technology* **39**, 897–902.
- Talebnia, F., Karakashev, D. and Angelidaki, I. (2010) Production of bioethanol from wheat straw: An overview on pretreatment, hydrolysis and fermentation. *Bioresource Technology* **101**, 4744–4753.
- Tandrup, T., Frandsen, K. E. H., Johansen, K. S., Berrin, J.-G. and Lo Leggio, L. (2018) Recent insights into lytic polysaccharide monoxygenases (LPMOs). *Biochemical Society transactions*.
- Tangthirasunun, N., Navarro, D., Garajova, S., Chevret, D., Tong, L. C. H., Gautier, V., Hyde, K. D., Silar, P. and Berrin, J.-G. (2017) Inactivation of Cellobiose Dehydrogenases Modifies the Cellulose Degradation Mechanism of *Podospira anserina*. *Applied and environmental microbiology* **83**.
- Tengborg, C., Galbe, M. and Zacchi, G. (2001) Influence of enzyme loading and physical parameters on the enzymatic hydrolysis of steam-pretreated softwood. *Biotechnology Progress* **17**, 110–117.

- Thareja, A., Puniya, A. K., Goel, G., Nagpal, R., Sehgal, J. P., Singh, P. K. and Singh, K. (2006) In vitro degradation of wheat straw by anaerobic fungi from small ruminants. *Archives of animal nutrition* **60**, 412–417.
- Uchida, K. and Kawakishi, S. (1989) Ascorbate-mediated specific oxidation of the imidazole ring in a histidine derivative. *Bioorganic Chemistry* **17**, 330–343.
- Urresti, S., Cartmell, A., Liu, F., Walton, P. H. and Davies, G. J. (2018) Structural studies of the unusual metal-ion site of the GH124 endoglucanase from *Ruminiclostridium thermocellum*. *Acta crystallographica. Section F, Structural biology communications* **74**, 496–505.
- Vaaje-Kolstad, G., Bøhle, L. A., Gåseidnes, S., Dalhus, B., Bjørås, M., Mathiesen, G. and Eijsink, V. G. H. (2012) Characterization of the Chitinolytic Machinery of *Enterococcus faecalis* V583 and High-Resolution Structure of Its Oxidative CBM33 Enzyme. *Journal of Molecular Biology* **416**, 239–254.
- Vaaje-Kolstad, G., Forsberg, Z., Loose, J. S., Bissaro, B. and Eijsink, V. G. (2017) Structural diversity of lytic polysaccharide monoxygenases. *Current opinion in structural biology* **44**, 67–76.
- Vaaje-Kolstad, G., Horn, S. J., van Aalten, D. M. F., Synstad, B. and Eijsink, V. G. H. (2005) The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *The Journal of biological chemistry* **280**, 28492–28497.
- Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sørli, M. and Eijsink, V. G. H. (2010) An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides. *Science* **330**, 219–222.
- van den Brink, J., Maitan-Alfenas, G. P., Zou, G., Wang, C., Zhou, Z., Guimarães, V. M. and Vries, R. P. de (2014) Synergistic effect of *Aspergillus niger* and *Trichoderma reesei* enzyme sets on the saccharification of wheat straw and sugarcane bagasse. *Biotechnology Journal* **9**, 1329–1338.
- van den Brink, J. and Vries, R. P. (2011) Fungal enzyme sets for plant polysaccharide degradation. *Applied Biochemistry and Biotechnology* **91**, 1477.
- van Dyk, J. S. and Pletschke, B. I. (2012) A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes—Factors affecting enzymes, conversion and synergy. *Biotechnology Advances* **30**, 1458–1480.
- Villares, A., Moreau, C., Bennati-Granier, C., Garajova, S., Foucat, L., Falourd, X., Saake, B., Berrin, J.-G. and Cathala, B. (2017) Lytic polysaccharide monoxygenases disrupt the cellulose fibers structure. *Scientific reports* **7**, 40262.
- Villegas, J. and Gnansounou, E. (2008) Techno-economic and environmental evaluation of lignocellulosic biochemical refineries: Need for a modular platform for integrated assessment (MPIA). *Journal of Scientific and Industrial Research* **67**.
- Vitikainen, M., Arvas, M., Pakula, T., Oja, M., Penttilä, M. and Saloheimo, M. (2010) Array comparative genomic hybridization analysis of *Trichoderma reesei* strains with enhanced cellulase production properties. *BMC Genomics* **11**, 441.
- Vlasenko, E., Schüle, M., Cherry, J. and Xu, F. (2010) Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases. *Bioresource Technology* **101**, 2405–2411.
- Voshol, G. P., Vijgenboom, E. and Punt, P. J. (2017) The discovery of novel LPMO families with a new Hidden Markov model. *BMC Research Notes* **10**, 105.
- Vu, V. V., Beeson, W. T., Phillips, C. M., Cate, J. H. D. and Marletta, M. A. (2014a) Determinants of regioselective hydroxylation in the fungal polysaccharide monoxygenases. *Journal of the American Chemical Society* **136**, 562–565.
- Vu, V. V., Beeson, W. T., Span, E. A., Farquhar, E. R. and Marletta, M. A. (2014b) A family of starch-active polysaccharide monoxygenases. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13822–13827.
- Vu, V. V., Hangasky, J. A., Detomasi, T. C., Henry, S. J. W., Ngo, S. T., Span, E. A. and Marletta, M. A. (2019) Substrate selectivity in starch polysaccharide monoxygenases. *The Journal of biological chemistry*.
- Vu, V. V. and Marletta, M. A. (2016) Starch-degrading polysaccharide monoxygenases. *Cellular and Molecular Life Sciences* **73**, 2809–2819.
- Vuong, T. V., Liu, B., Sandgren, M. and Master, E. R. (2017) Microplate-Based Detection of Lytic Polysaccharide Monoxygenase Activity by Fluorescence-Labeling of Insoluble Oxidized Products.
- Wang, D., Li, J., Salazar-Alvarez, G., McKee, L. S., Srivastava, V., Sellberg, J. A., Bulone, V. and Hsieh, Y. S. Y. (2018a) Production of functionalised chitins assisted by fungal lytic polysaccharide monoxygenase. *Green Chemistry* **20**, 2091–2100.
- Wang, D., Li, J., Wong, A. C. Y., Aachmann, F. L. and Hsieh, Y. S. Y. (2018b) A colorimetric assay to rapidly determine the activities of lytic polysaccharide monoxygenases. *Biotechnology for biofuels* **11**, 215.
- Wang, T., Liu, X., Yu, Q., Zhang, X., Qu, Y., Gao, P. and Wang, T. (2005) Directed evolution for engineering pH profile of endoglucanase III from *Trichoderma reesei*. *Biomolecular engineering* **22**, 89–94.
- Westereng, B., Agger, J. W., Horn, S. J., Vaaje-Kolstad, G., Aachmann, F. L., Stenstrøm, Y. H. and Eijsink, V. G. H. (2013) Efficient separation of oxidized cello-oligosaccharides generated by cellulose degrading lytic polysaccharide monoxygenases. *Journal of Chromatography A* **1271**, 144–152.

- Westereng, B., Arntzen, M. Ø., Aachmann, F. L., Várnai, A., Eijsink, V. G. H. and Agger, J. W. (2016) Simultaneous analysis of C1 and C4 oxidized oligosaccharides, the products of lytic polysaccharide monoxygenases acting on cellulose. *Journal of Chromatography A* **1445**, 46–54.
- Westereng, B., Cannella, D., Agger, J. W., Jørgensen, H., Andersen, M. L., Eijsink, V. G.H. and Felby, C. (2015) Enzymatic cellulose oxidation is linked to lignin by long-range electron transfer. *Scientific reports* **5**, 18561.
- Witzler, M., Alzagameem, A., Bergs, M., Khaldi-Hansen, B. E., Klein, S. E., Hielscher, D., Kamm, B., Kreyenschmidt, J., Tobiasch, E. and Schulze, M. (2018) Lignin-Derived Biomaterials for Drug Release and Tissue Engineering. *Molecules* **23**.
- Wood, T. M. (1988) Preparation of crystalline, amorphous, and dyed cellulase substrates. *Methods in Enzymology* **160**, 19–25.
- Yang, J. and Zhang, Y. (2015) I-TASSER server: New development for protein structure and function predictions. *Nucleic acids research* **43**, W174-81.
- Zhang, F., Anasontzis, G. E., Labourel, A., Champion, C., Haon, M., Kemppainen, M., Commun, C., Deveau, A., Pardo, A., Veneault-Fourrey, C., Kohler, A., Rosso, M.-N., Henrissat, B., Berrin, J.-G. and Martin, F. (2018) The ectomycorrhizal basidiomycete *Laccaria bicolor* releases a secreted beta-1,4 endoglucanase that plays a key role in symbiosis development. *The New phytologist* **220**, 1309–1321.
- Zhang, P. Y.-H., Himmel, M. E. and Mielenz, J. R. (2006) Outlook for cellulase improvement: Screening and selection strategies. *Biotechnology Advances* **24**, 452–481.
- Zhou, H., Li, T., Yu, Z., Ju, J., Zhang, H., Tan, H., Li, K. and Yin, H. (2019) A lytic polysaccharide monoxygenase from *Myceliophthora thermophila* and its synergism with cellobiohydrolases in cellulose hydrolysis. *International Journal of Biological Macromolecules* **139**, 570–576.

Title : Exploration of *Aspergillus* spp. secretomes for the complementation of *Trichoderma reesei* cellulolytic cocktail

Keywords : Bioethanol ; Lignocellulosic biomass ; Filamentous fungi ; Secretomes ; LPMOs

Abstract

Lignocellulosic biomass is considered as a promising alternative to fossil resources. This abundant and renewable feedstock, derived from plant cell wall, is composed of polysaccharides and aromatic polymers linked together to form a highly resistant structure. Its transformation into bioethanol or other products requires the degradation of polysaccharides into fermentable sugars, using enzymes such as cellulases, hemicellulases and lytic polysaccharide monoxygenases (LPMOs). These enzymes are secreted by biomass-degrading microorganisms, such as bacteria and filamentous fungi. The most commonly used is *Trichoderma reesei*, a fungal species that is known for its efficient secretion of cellulases, but has a low diversity of other enzymes. Enzymatic hydrolysis is still a bottleneck for the cost-effective utilization of lignocellulosic biomass, and needs to be optimized. Providing activities that are missing in *T. reesei* using enzymes produced by other fungal organisms is a promising approach, which has been implemented during this PhD project.

Using a proteomic approach, different *Aspergillus* strains were studied in order to find enzymatic activities that could improve the hydrolysis of pretreated biomass. For this purpose, 5 strains were grown on 3 inducers, and the protein content of these different secretomes was determined by mass spectrometry. The secretomes were then tested for their ability to improve hydrolysis yields by supplementing a *T. reesei* cocktail. Analysis of the protein content of some of the best-performing secretomes revealed the presence of a protein belonging to a putative family of LPMO. Several proteins of this family were recombinantly produced and tested on different substrates in the presence of redox partners. Oxidative cleavage of cellulose was demonstrated, leading to the creation of a new LPMO family, called AA16. The pretreatment of cellulosic substrates by AA16 LPMOs resulted in a significant improvement of their hydrolysis by a cellulase from *T. reesei*, demonstrating their potential interest for biotechnological applications in biorefineries.

Titre : Exploration de sécrétomes d'*Aspergillus* spp. en vue de la complémentation du cocktail cellulolytique de *Trichoderma reesei*

Mots-clés : Bioéthanol ; Biomasse lignocellulosique ; Champignons filamenteux ; Sécrétomes ; LPMO

Résumé

L'une des alternatives envisagées pour remplacer les ressources fossiles est l'utilisation de biomasse lignocellulosique. Cette ressource abondante et renouvelable, issue de la paroi des végétaux, est composée de polysaccharides et de polymères aromatiques liés entre eux qui lui confèrent une grande résistance. Aujourd'hui, la biomasse lignocellulosique est transformée en bioéthanol via un procédé dont l'une des étapes clé est la dégradation des polysaccharides en sucres fermentescibles. Pour cela, des enzymes, notamment des cellulases, hémicellulases et « lytic polysaccharide monoxygenases » (LPMO) sont utilisées. Ces enzymes sont sécrétées par des organismes dégradeurs de la biomasse tels que l'ascomycète *Trichoderma reesei*, un champignon filamenteux capable de sécréter de très grandes quantités de cellulases, mais qui possède une faible diversité enzymatique. L'hydrolyse enzymatique restant à ce jour une étape très coûteuse, il est nécessaire d'en améliorer encore les rendements. L'apport d'activités enzymatiques manquantes chez *T. reesei* à partir d'autres organismes fongiques est une approche prometteuse, qui a été développée dans le cadre de ce projet de thèse.

Différentes souches du genre *Aspergillus* ont été étudiées afin d'y trouver des activités enzymatiques permettant d'améliorer l'hydrolyse de biomasses prétraitées. Pour cela, 5 souches ont été cultivées en présence de 3 inducteurs, et le contenu protéique de ces différents sécrétomes a été déterminé par spectrométrie de masse. Les sécrétomes ont ensuite été testés pour leur capacité à compléter un cocktail de *T. reesei* en améliorant ses rendements d'hydrolyse. L'analyse des protéines contenues dans quelques sécrétomes performants a révélé la présence d'une protéine appartenant à une famille putative de LPMO. Plusieurs protéines de cette famille ont été produites de manière recombinante et testées sur différents substrats en présence de partenaires redox. Une activité de clivage oxydatif de la cellulose a été démontrée, permettant la définition d'une nouvelle famille de LPMO fongiques, nommée AA16. Le prétraitement de substrats cellulotiques par les LPMO AA16 entraîne une amélioration significative de leur hydrolyse par une des cellulases de *T. reesei*, ce qui démontre l'intérêt potentiel de cette nouvelle famille d'enzymes pour des applications biotechnologiques dans le domaine des bioraffineries.