UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA VETERINÁRIA

HIGH-THROUGHPUT PRODUCTION AND CHARACTERIZATION OF CARBOHYDRATE-ACTIVE ENZYMES FOR ANIMAL NUTRITION

VÂNIA ALEXANDRA DA SILVA CARDOSO LOPES

Orientadores: Professor Doutor Carlos Mendes Godinho de Andrade Fontes

Doutora Joana Luís Armada Brás

Tese especialmente elaborada para obtenção do grau de Doutor em Ciências Veterinárias na especialidade Produção Animal

2019

UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA VETERINÁRIA

HIGH-THROUGHPUT PRODUCTION AND CHARACTERIZATION OF CARBOHYDRATE-ACTIVE ENZYMES FOR ANIMAL NUTRITION

VÂNIA ALEXANDRA DA SILVA CARDOSO LOPES

Orientadores: Professor Doutor Carlos Mendes Godinho de Andrade Fontes

Doutora Joana Luís Armada Brás

Tese especialmente elaborada para obtenção do grau de Doutor em Ciências Veterinárias na especialidade Produção Animal

Júri
Presidente: Professor Doutor Luís Filipe Lopes da Costa
Vogais:
- Professor Doutor Luís Manuel dos Anjos Ferreira
- Professor Doutor Carlos Mendes Godinho de Andrade Fontes
- Professora Doutora Teresa Sacadura Santos Silva
- Professora Doutora Maria Isabel Ferraz de Oliveira

2019

DECLARAÇÃO RELATIVA ÀS CONDIÇÕES DE REPRODUÇÃO DA TESE

Nome: Vânia Alexandra da Silva Cardoso Lopes

Título da Tese ou Dissertação: High-Throughput production and characterization of Carbohydrate-Active enZYmes for animal nutrition

Ano de conclusão (indicar o da data da realização das provas públicas): 2020

Designação do curso de Mestrado ou de Doutoramento: Doutoramento em ciências veterinárias na especialidade produção animal

Área científica em que melhor se enquadra (assinale uma):

☐ Clínica ☒ Produção Animal e Segurança Alimentar

☐ Morfologia e Função ☐ Sanidade Animal

Declaro sobre compromisso de honra que a tese ou dissertação agora entregue corresponde à que foi aprovada pelo júri constituído pela Faculdade de Medicina Veterinária da ULISBOA.

Declaro que concedo à Faculdade de Medicina Veterinária e aos seus agentes uma licença não-exclusiva para arquivar e tornar acessível, nomeadamente através do seu repositório institucional, nas condições abaixo indicadas, a minha tese ou dissertação, no todo ou em parte, em suporte digital.

Declaro que autorizo a Faculdade de Medicina Veterinária a arquivar mais de uma cópia da tese ou dissertação e a, sem alterar o seu conteúdo, converter o documento entregue, para qualquer formato de ficheiro, meio ou suporte, para efeitos de preservação e acesso.

Retenho todos os direitos de autor relativos à tese ou dissertação, e o direito de a usar em trabalhos futuros (como artigos ou livros).

Concordo que a minha tese ou dissertação seja colocada no repositório da Faculdade de Medicina Veterinária com o seguinte estatuto (assinale um):

1. ☐ Disponibilização imediata do conjunto do trabalho para acesso mundial;
2. ☒ Disponibilização do conjunto do trabalho para acesso exclusivo na Faculdade de Medicina Veterinária durante o período de ☒ 6 meses, ☐ 12 meses, sendo que após o tempo assinalado autorizo o acesso mundial*;

* Indique o motivo do embargo (OBRIGATÓRIO)

Aguarda publicação de artigos científicos

Nos exemplares das dissertações de mestrado ou teses de doutoramento entregues para a prestação de provas na Universidade e dos quais é obrigatoriamente enviado um exemplar para depósito na Biblioteca da Faculdade de Medicina Veterinária da Universidade de Lisboa deve constar uma das seguintes declarações (incluir apenas uma das três):

1. É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE/TRABALHO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Faculdade de Medicina Veterinária da Universidade de Lisboa, 16 de janeiro de 2020

Assinatura: Vânia Cardoso

À minha filha e ao meu marido

AGRADECIMENTOS

Esta tese não é apenas o resultado de muitos anos trabalho, mas também o resultado da colaboração de muitas pessoas. Nesta longa viagem, de quatro anos, as pessoas que me acompanharam foram fundamentais desde o primeiro até ao último dia desta jornada. Por isso quero agradecer:

À Fundação para a Ciência e Tecnologia pelo financiamento de parte da minha bolsa de doutoramento;

À NZYTech por me ter acolhido e permitido fazer o doutoramento em empresa. Por me ter financiado a segunda parte da minha bolsa. Pelos meios físicos e materiais que me disponibilizou para a realização de todo o trabalho;

À Faculdade de Medicina Veterinária e ao CIISA por me terem aceite como estudante de doutoramento e por terem disponibilizado meios físicos e materiais para a realização de todo o trabalho;

Ao Professor Doutor Carlos Fontes, por ter aceite ser meu orientador. Foi um privilégio ser sua aluna, não só por ser um investigador excecional e profissional dedicado, mas também por ter sido o melhor orientador que podia ter tido. Um grande e sincero obrigada;

À Doutora Joana Brás por me ter orientado sempre que precisei. Pelo apoio e ensinamentos em todas as tarefas laboratoriais e científicas, principalmente no que tocas às técnicas HTP. Obrigada;

Ao Professor Doutor Luís Ferreira, pela sua orientação e pelas palavras sábias no momento certo;

Ao Professor Doutor José Prates, pelos ensinamentos que me transmitiu pela sua disponibilidade e simpatia constantes;

À Teresa Ribeiro por ter sido o suporte mais importante no início da minha jornada pela investigação;

Ao Pedro Bule pelo apoio nas tarefas laboratoriais, pelos ensinamentos e pela troca de ideias;

À Maria Prata por todo o companheirismo e amizade que temos partilhado. Por ter estado sempre a apoiar-me e a dar-me força ao longo desta jornada;

À Vânia Fernandes pela amizade, companheirismo, entreajuda e suporte, em especial nas tarefas laboratoriais;

À Andreia Peixoto, João Belchior, Mara Ramires e Rosa Marques pela vossa amizade e companhia, mas a cima de tudo, pelo sentido de humor e as risadas que me provocaram;

À Virgínia Pires e Immacolata Venditto pela amizade e companheirismo;

À Patrícia Ponte, Filipa Sequeira e Nádia Rodrigues por todas as partilhas de conhecimentos e orientação no funcionamento de uma empresa;

Ao Diogo Comprido, Pedro Rio e José Albuquerque pela vossa capacidade de trabalho e pela boa disposição e simpatia;

Ao Instituto Superior de Agronomia na cedência de instalações para os ensaios de frangos e a toda a equipa envolvida nos mesmos;

À minha família, em especial aos meus pais, João e Fátima Cardoso, pelo apoio constante, motivação, amor e carinho. Sem eles não teria conseguido chegar onde cheguei. Agradeço-vos muito por tudo o que já fizeram por mim;

Ao meu marido Pedro Lopes pela motivação, apoio incondicional e toda a paciência. Por toda a assistência informática e bioinformática que me deu, desde o desenvolvimento de ferramentas bioinformáticas (p.e. Gene alignment algorithm) à resolução das pequenas dificuldades encontradas no dia-a-dia. Por nunca ter duvidado das minhas capacidades. Principalmente por me aturar;

À minha filha Leonor Lopes por me ter feito tanta companhia durante o processo de escrita e por involuntariamente me ter dado tanta motivação e inspiração para finalizar esta etapa da minha vida.

This work was co-funded by Fundação para a Ciência e a Tecnologia, grant SFRH/BDE/101026/2014 from Ministério da Ciência, Tecnologia e Ensino Superior, and by NZYTech, Lda, genes & enzymes

RESUMO

Na natureza, a biodegradação dos hidratos de carbono da parede celular vegetal é realizada por enzimas microbianas, geralmente conhecidas como CAZymes. Os animais monogástricos produzem um reportório limitado de enzimas para degradação destes hidratos de carbono, não conseguindo usar eficientemente alguns ingredientes da dieta, que muitas vezes manifestam propriedades anti nutritivas. Assim, é sabido que a suplementação com CAZymes exógenas melhora o valor nutritivo das dietas e aumenta o desempenho produtivo dos animais. Este trabalho revelou que as enzimas mais apropriadas para suplementar dietas ricas em 1,3-1,4-β-glucanos são as 1,3-1,4-β-glucanases e não as celulases. Além disso, verificou-se que a suplementação com β-xilanases melhorou o valor nutritivo de dietas que continham variedades de trigo com maior viscosidade e menor atividade endógena de endo-1,4-β-xilanase. Em oposição, quando o lote de trigo apresentou menor viscosidade e maiores níveis de atividade endógena de endo-1,4-β-xilanase, a resposta dos animais à adição das enzimas foi menor. Este trabalho mostra, igualmente, que os xilo-oligossacarídeos, resultantes da degradação de arabinoxilanos por xilanases exógenas, possuem uma ação pré-biótica na alimentação de frangos, promovendo a melhoria do desempenho zootécnico. Contudo, apesar de estarem descritas uma grande diversidade CAZymes, poucas são as estudadas com potencial para serem usadas em alimentação animal. Portanto, este trabalho pretendeu, também, desenvolver metodologias para isolar e caracterizar enzimas potencialmente importantes em larga escala. Foram selecionadas, produzidas e expressas na forma recombinante 1476 CAZymes. Os dados revelaram que 79% das proteínas recombinantes foram produzidas na forma solúvel em *Escherichia coli*. Verificou-se, ainda, que fatores como o organismo de origem, a estratégia de produção, a fusão com marcadores de solubilidade, o peso molecular da proteína e composição de aminoácidos das sequências primárias, parecem justificar os resultados da solubilidade. Estes ensinamentos foram utilizados para produzir enzimas, tais como ferulolil esterases (FAEs) e glucuronil esterases (GEs), que removem as cadeias laterais e quebram as ligações cruzadas entre hidratos de carbono hemicelulósicos e a lenhina. Assim sendo, foram selecionadas 480 FAEs e 20 GEs para produção recombinante e caracterização bioquímica. Cerca de 372 FAEs e 11 GEs foram produzidos em forma solúvel em *E. coli* e aproximadamente 50% das enzimas produzidas mantiveram níveis significativos de atividade e estabilidade. Com isto, foi possível identificar e produzir um número significativo de FAEs e GEs com potencial para alimentação animal, em especial as que libertam celulose e hemicelulose da lenhina.

Palavras-chave: CAZymes, pré-bióticos, HTP, feruloil esterases e glucuronoil esterases

ABSTRACT

The biodegradation of plant cell wall (PCW) carbohydrates is performed by microbial enzymes that are generally referred to as CAZymes. In animal nutrition, it is now well established that the monogastric animals produce a limited repertoire of CAZymes and as such cannot use efficiently some dietary ingredients that sometimes display antinutritional properties. The dietary supplementation with exogenous CAZymes improves the nutritive value of diets and increases animal's performance. In particular, this study demonstrated that 1,3-1,4-β-glucanases and not cellulases improve the nutritive value of β-glucan-containing diets for monogastric animals. In addition, it was revealed that exogenous enzyme supplementation with β-xylanases improved the nutritive value of diets incorporating wheat lots with high viscosity and low endogenous endo-1,4-β-xylanase activity. In contrast, when the wheat lot showed lower viscosity and higher levels of endogenous endo-1,4-β-xylanase activity, broiler response was clearly diminished. Moreover, the data revealed that xylo-oligosaccharides released by xylanases acting on cereal arabinoxylans display a pre-biotic and positive effect in broiler chicks. However, although we observe an exponential accumulation of genomic and metagenomic information, knowledge on CAZYmes with potential to be used in animal nutrition is limited. This work also aimed to develop high-throughput (HTP) methodologies to isolate and characterize potentially important enzymes for animal nutrition. Thus, 1476 recombinant enzymes were selected and produced recombinantly. The data revealed that 79% of recombinant proteins were produced in the soluble form in *Escherichia coli*. Factors, such as, organism of origin, gene production strategy, fusion with solubility tags, protein molecular weight and amino acids composition of primary sequences may be used to justify and predict levels of solubility. The establishment of a high-throughput pipeline for recombinant enzyme production was used to obtain a library of feruloyl esterases (FAEs) and glucuronoyl esterases (GEs), enzymes which remove the side chains and break crosslinks between hemicellulosic carbohydrates and lignin. Thus 480 putative FAEs and 20 GEs were produced and biochemically characterized. Following gene isolation, 372 FAEs and 11 GEs were produced in a soluble form in *E. coli*. Activity results showed that 50% of the enzymes produced retained significant levels of activity and stability. The library of innovative FAEs and GEs produced during this project will be used to develop a novel generation of enzymes for animal nutrition, in particular to exploit the release of cellulose and hemicellulose from lignin.

Key-words: CAZymes, prebiotics, HTP, feruloyl esterases and glucuronoyl esterases

This thesis was based on the following manuscripts:

V. Cardoso, T. Ribeiro, J.L.A. Brás, V.O. Fernandes, H. Santos, M.M. Lordelo, L.M.A. Ferreira and C.M.G.A. Fontes. Novel evidence supporting the role of 1,3-1,4-β-glucanases and not 1,4-β-glucanases to improve the nutritive value of β-glucan-containing diets for monogastric animals. (manuscript submitted for publication)

V. Cardoso*, E. A. Fernandes*, H. M. M. Santos, B. Maçãs, M. M. Lordelo, L.T. Gama, L. M. A. Ferreira, C. M. G. A. Fontes and T. Ribeiro. 2018. Variation in levels of non-starch polysaccharides and endogenous endo-1,4-β-xylanases affects the nutritive value of wheat for poultry. *British Poultry Science* 59(2):218-226. (DOI: 10.1080/00071668.2018.1423674).

T. Ribeiro*, **V. Cardoso***, L.M.A. Ferreira, M.M. Lordelo, E. Coelho, A.S.P Moreira, M.R.M. Domingues, M.A. Coimbra, M.R. Bedford, and C.M.G.A. Fontes. 2018. Xylo-oligosaccharides display a prebiotic activity when used to supplement wheat or corn-based diets for broilers. *Poultry Science* 97(12):4330-4341. (DOI: 10.3382/ps/pey336)

V. Cardoso, J.L.A. Brás, L.M.A. Ferreira, L.T. Gama, R. Vincentelli, B. Henrissat and C.M.G.A. Fontes. Generation of a library of Carbohydrate-Active Enzymes for plant biomass deconstruction. (manuscript in preparation)

V. Cardoso, J.L.A. Brás, L.M.A. Ferreira, V.D. Alves, C.M.G.A. Fontes and P. Bule. Structural determinants of substrate specificity in family 15 Carbohydrate Esterases. (manuscript in preparation)

V. Cardoso, J.L.A. Brás, A.L Carvalho, V.D. Alves, R.P. de Vries, V. Puchart, P. Biely, L.M.A. Ferreira, C.M.G.A. Fontes and P. Bule. Subtle changes at the catalytic site of feruloyl esterases modulate enzyme specificity to hydroxycinnamic acids. (manuscript submitted for publication)

* Authors contributed equally to the work

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

%	Percentage
[M+Na]$^+$	Sodium adduct ions
Å	Angstrom
A. cellulolyticus	*Acetivibrio cellulolyticus*
A550	Absorbance at 550 nanometers
AA	Auxiliary Activities
AAO	Aryl-alcohol oxidase
ABF	α-arabinofuranosidase
AcE	Acetyl esterases
AcXEs	Acetylxylan esterases
ADF	Acid Detergent Fibre
ADL	Acid Detergent Lignin
AGA	Apiogalacturonan
Ala	Alanine (A)
ANOVA	Analysis of variance
Araf	Arabinofuranose
Arg	Arginine (R)
Ash	Mineral matter present in feed
Asn	Asparagine (N)
Asp	Aspartic acid (D)
AXE	Acetyl xylan esterase
AXOS	Arabinoxylo-oligosaccharides
B. cellulosolvens	*Bacteroides cellulosolvens*
BL	Breeding lines
BL21(DE3)	*E. coli* expression strain containing the DE3 lysogen that carries the gene for T7 RNA polymerase under control of the *lac*UV5 promoter
BLAST	Basic Local Alignment Search Tool
BnzGlcA	Benzyl D-glucuronate
BSA	Bovine serum albumin
BW	Body Weight
BXL	1,4-β-xylosidase
C-	Non-supplemented diet
C. acetobutylicum	*Clostridium acetobutylicum*
C. cellulolyticum	*Clostridium cellulolyticum*
C. cellulolyticus	*Clostridium cellulolyticus*
C. cellulovorans	*Clostridium cellulovorans*
C. clariflavum	*Clostridium clariflavum*
C. japonicus	*Clostridium japonicus*
C. josui	*Clostridium josui*
C. papyrosolvens	*Clostridium papyrosolvens*
C. S. usitatus	*Candidatus Solibacter usitatus*
C. thermocellum	*Clostridium thermocellum*
Ca^{2+}	Calcium ion
$CaCl_2$	Calcium chloride
CAI	Codon Adaptation Index
CAZyme	Carbohydrate-Active enZyme
CBH	Cellobiohydrolases
CBM	Carbohydrate-Binding Module

CE	Carbohydrate Esterase
CE1	Carbohydrate Esterase family 1
CE15	Carbohydrate Esterase family 15
cm/kg	Centimetre per kilogram
coh	Cohesin
cP	Centipoise
CP	Crude Protein
CV	Cultivar lines
Cys	Cysteine (C)
Dha	2-keto-3-deoxy-D-lyxo heptulosaric acid
DHB	2,5-dihydroxybenzoic acid
DNA	Desoxyribonucleic Acid
DNSA	3,5-dinitrosalicylic acid
Doc	Dockerin
DyPs	Decolorizing peroxidases
E. coli	*Escherichia coli*
EC	Enzyme Commission Number
EG	Endoglucanase
ES	Spanish origin
eV	Electron volt
F ratio	Variation between sample means / variation within the samples
FA	Ferulic acid
FAE	Ferulic Acid Esterase
FCR	Feed Conversion Ratio
FOS	fructo-oligosaccharides
FR	French origin
Fuc	L-Fucose
G	Guaiacyl
g/kg	Gram per kilogram
Gal	D-Galactose
GalpA	Galacturonic acid
GAXs	Glucuronoarabinoxylans
GC	Gas chromatography
GC–MS	Gas chromatography–mass spectrometry
gDNA	Genomic desoxyribonucleic Acid
GE	Glucuronoyl esterase
GFP	Green-fluorescence protein
GH	Glycoside Hydrolase
GI	Gastrointestinal tract
GlcA	Glucuronic acid
Gln	Glutamine (Q)
GLOX	Glyoxal oxidase
Glu	Glutamic acid (E)
Gly	Glycine (G)
GOS	Galacto-oligosaccharides
GS	Gene synthesis
GST	Glutathione S-transferase
GT	Glycosyl Transferase
GT-A	Glycosyltransferase fold A

GT-B	Glycosyltransferase fold B
GUS	α-glucuronidase
h	Hour
H	Hydroxyphenyl
H_2O_2	Hydrogen peroxide
H2SO4	Sulfuric acid
HA	Enzymes with high activity
HALV	High endogenous endo-1,4-β-xylanases activity and low viscosity
HCl	Hydrogen chloride
Hepes	Hydroxyethyl piperazineethanesulfonic acid
Hex_n	Hexose oligosaccharides
HG	Homogalacturonan
His	Histidine (H)
His6-tag	Six Histidines tag
HTP	High-throughput
Ile	Isoleucine (I)
IMAC	Immobilized metal ion-affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IUBMB	Union of Biochemistry and Molecular Biology
K2HPO4	Potassium hydrogen phosphate
Kan	Kanamycin
kDa	Kilodalton
Kdo	2-keto-3-deoxy-D-manno octulosonic acid
Kg	Kilogram
KOH	Potassium hydroxide
L	Litre
LA	Enzymes with low activity
Lac	Laccases
LAHV	Low endogenous endo-1,4-β-xylanases activity and high viscosity
L-Araf	L-arabinofuranose
LB	Luria Bertani
LCCs	Lignin-carbohydrate complexes
Leu	Leucine (L)
LiP	lignin peroxidases
LPMO	Lytic polysaccharide mono-oxygenases
Lys	Lysine (K)
M	Molar
MALDI	Matrix-assisted laser desorption/ionization
MBP	Maltose-binding polypeptide
MCA	Methyl caffeate
MeGlcA	4-O-methyl-D-glucuronic acid
MES	2-(N-morpholino)ethanesulfonic acid
Met	Methionine (M)
MFA	Methyl ferulate
mg	Milligram
min	Minutes
mL	Milliliter
mM	Millimolar

Mn^{2+}	Manganese cation (2+)
Mn^{3+}	Manganese cation (3+)
MnP	Manganese peroxidases
MOS	Mannan-oligosaccharides
MpCA	Methyl p-coumarate
MSA	Methyl sinapate
MT	Enzymes active on methyl ester substrates
MW	Molecular Weight
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
nd	Not detectable
NDF	Neutral Detergent Fiber
Ni^{2+}	Nickel ion
nm	Nanometer
NRC	National Research Council
NSP	Non-starch polysaccharide
NZY5α	*E. coli* competent cells with similar properties to DH5α
ºC	Celcius degree
OD	Optical Density
OSE	diet supplemented with xylose (Experiment 1)
OTUs	Operational Taxonomic Units
p	p-value
p(F)	p-value from F-Ratio
PC	Phosphate citrate
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PCW	Plant cell wall
PD10	GE Healthcare dessalting columns
PDB	Protein data bank
PEG 3350	Polyethylene glycol 3350
$Pent_n$	Pentose oligosaccharides
pH	Negative decimal logarithm of the hydrogen ion activity in a solution
Phe	Phenylalanine (F)
pHTP1	*E.coli* expression vector containing a N-terminal Histidine tag
pHTP9	*E.coli* expression vector containing a N-terminal GFP fusion tag
PL	Polysaccharide Lyase
pNP	P-nitrophenol
pNP-Fe	P-nitrophenyl ferulate
pNPP	4-nitrophenol palmitate
PP	Enzymes that attack pNP-ferulate but not the methyl substrates
Pro	Proline (P)
R. albus	*Ruminococcus albus*
R. flavefaciens	*Ruminococcus flavefaciens*
r.m.s.d	Root mean square deviation
r2	Coefficient of determination

RGI	Rahmnogalacturonan I
RGII	Rahmnogalacturonan II
Rha*p*	Rhamnose
S	Syringyl
SAS	Statistical Analysis Software
SCFA	Short Chain Fatty Acids
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEM	Standard error of the mean
Ser	Serine (S)
SLHs	S-layer homology modules
SOS	Soybean meal oligosaccharides
T. turnerae	*Teredinibacter turnerae*
TCEP	Tris(2-carboxyethyl)phosphine
Thr	Threonine (T)
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
Trp	Tryptophan (W)
TrxA	Thioredoxin
Tyr	Tyrosine (Y)
U	Enzymatic unities
U/g	Enzymatic unities per gram
U/kg	Enzymatic unities per kilogram
UK	United Kingdom origin
Unk	Unknown
USA	American origin
v/v	Volume per volume
Val	Valine (V)
VFA	Volatile fatty acids
VP	Versatile peroxidases
w/v	Weight per volume
w/w	Weight per weight
×g	Ggravity or relative centrifugal force
XGA	Xylogalacturonan
XLN	1,4-β-endoxylanase
XOS	Xylo-oligosaccharides
XYL	1,4-β-xylanase
βG	β-Glucosidade
µg	Microgram
µL	Microliter
µM	Micromolar
µm	Micrometre

1. BIBLIOGRAPHIC REVIEW AND OBJECTIVES

1.1. Introduction

Plant cell walls (PCWs) and their carbohydrate components are the most abundant organic compounds found in nature and the most abundant source of organic carbon in the biosphere. These structures are mainly constituted by cellulose, hemicellulose (which include xyloglucans, xylans, mannans, and β-glucans), pectin and lignin. Recycling these highly relevant molecules is a process of growing interest, particularly in the biofuel and bioprocessing sectors. In nature, the recycling of organic carbon stored in the plant cell wall is performed by microbial enzymes that convert the cell wall polysaccharides to monosaccharides and oligosaccharides. These enzymes are generally referred as Carbohydrate-Active Enzymes (CAZymes) and include Glycoside Hydrolases (GHs), Polysaccharide Lyases (PLs), Carbohydrate Esterases (CEs), Glycosyl Transferases (GTs) and Auxiliary Activities (AAs). Additionally, CAZymes are frequently modular, containing one or more catalytic domains connected to non-catalytic carbohydrate-binding modules (CBMs). CAZymes and CBMs have been grouped into sequence-based families on the continuously updated Carbohydrate-Active enZymes database (www.cazy.org). The complete biodegradation of recalcitrant plant cell wall carbohydrates requires the cooperative attack between enzymes that break the polysaccharide backbone, such as cellulase, hemicellulases, pectinases, among others, along with several accessory enzymes, which remove the side chains and break crosslinks between hemicellulosic carbohydrates and other plant polymers, such as lignin. Among the accessory enzymes, feruloyl esterases and glucuronoyl esterases, which belong to CEs enzyme class, play a key role in enhancing the accessibility of backbone enzymes to their target bonds. CAZYmes play not only a relevant role in carbon turnover in nature but also display considerable biotechnological potential, in particular in animal nutrition. It is well established that soluble non-starch polysaccharides, such as arabinoxylans, β-glucans and pectins are considerably anti-nutritive for simple-stomach animals. Particularly, in poultry, elevated levels of PCW Non-Starch Polysaccharides (NSPs) lead to an increase in digesta viscosity with a consequent decreased in nutrient digestion and nutrient absorption. When birds are fed wheat or barley-based diets, the presence of soluble NSPs is considerable. Therefore, it is now well established that the dietary supplementation with exogenous enzymes (such as cellulases and hemicellulases) reduces the degree of polymerization of NSPs, improving the nutritive value of diets and, consequently, increases animal's performance. However, the biological role of CAZymes in animal nutrition remains poorly understood, in particular in what relates

to the modulation of the gastrointestinal microflora through the production of dietary prebiotics.

This thesis aims to clarify several unresolved hypotheses related with CAZymes and their biotechnological impact in animal nutrition and it is divided in 8 main chapters. The first chapter begins with a general review on plant cell wall structure and composition, followed by a description of the different mechanisms and enzymes that are required for the degradation of structural lignocellulosic material. Then, a brief description of its importance in poultry production is also presented. At the end of the bibliographic review, the main objectives of this work are clearly defined. The following chapters (2 to 7) are organized in papers based on scientific manuscripts, already published, in preparation to submission or submitted to international journals. The 8th chapter aims to provide an integrated discussion and derive conclusions of all the work presented in this thesis.

1.2. Plant cell wall

The plant cell wall (PCW) is the most abundant source of terrestrial biomass. It is a rigid and semi-permeable protective layer located outside the cell membrane that also provides support for the cell's structure. Its main function is to provide cells rigidity, strength and protection, as well as providing a barrier against the environment and the invasion of potentially pathogenic organisms. Thus, PCWs act as an exoskeleton and are crucial for the development and function of plants (Scheller and Ulvskov 2010; Ralet *et al*. 2016).

1.2.1. Plant cell wall structure

The PCW is multi-layered and it is divided into three sections. These layers are identified as the primary cell wall, the middle lamella and the secondary cell wall. While all plant cells have a middle lamella and a primary cell wall (Figure 1.1), secondary cell walls (Figure 1.2) are characteristic of differentiated cells (Caffall and Mohnen 2009).

The middle lamella is the first synthesized layer of the PCW. It is an outer cell wall layer and contains, predominantly, polysaccharides called pectins (Reiter 2002). These polysaccharides have the function of cementing the cell wall contacts of two adjoining plant cells (Lodish *et al*. 2000).

When going from the periphery to the center of the cell, the next layer is the primary cell wall. It is formed between the middle lamella and the plasma membrane (Figure 1.2) and it is primarily composed of cellulose microfibrils, hemicellulose and pectin polysaccharides (Figure 1.1). The primary cell wall is classified into Type I and Type II, based on the structure. Type I primary cell walls are present in dicots and to a certain degree in monocots as well (McCann and Carpita 2008). It consists of cellulose microfibrils, which are embedded in a network built of xyloglucans and, to a lesser extent, glucuronoarabinoxylans (GAXs), as

well as pectins such as homogalacturonan and rhamnogalacturonan I (Carpita and Gibeaut 1993; Hoffman *et al.* 2005). Type II cell walls are present in *Poaceae* and in related monocots. In these cell walls, cellulose is embedded in a network of highly abundant GAXs and, to a lower extent, pectins, glucomannans, and xyloglucans (Gordon *et al*. 1985; McCann and Carpita 2008). This layer surrounds growing cells or cells capable of growth and provides the strength and flexibility needed to allow for cell to expand (Lodish *et al*. 2000; Keegstra 2010).

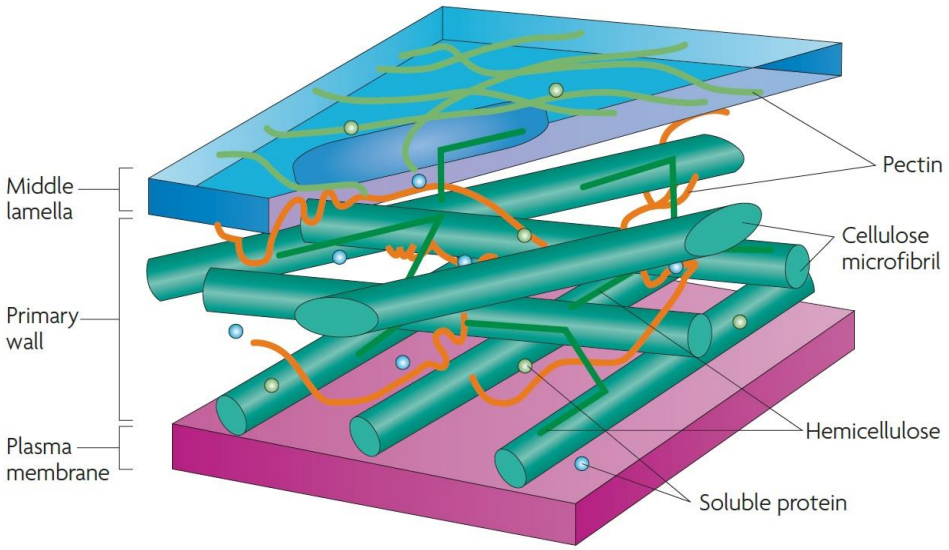


Figure 1.1. Schematic representation of the cell wall.

(Adapted from Sticklen 2008)

The third layer is the secondary cell wall which is formed between the primary cell wall and the plasma membrane, at later stages of plant differentiation (Figure 1.2). Secondary cell walls start developing when primary cell walls stop growing. They correspond to the most rigid and recalcitrant portion of PCWs and provide additional strength and support to the cell (Keegstra 2010). The secondary cell wall comprises cellulose and hemicellulose, such as (acetylated) GAXs in grasses and hard woods or mannans in soft woods (McCann and Carpita 2008). In addition to cellulose and hemicellulose, secondary walls are thickened structures containing lignin and surrounding specialized cells such as vessel elements or fiber cells (Keegstra 2010). Lignin strengthens the cell wall and aids in water conductivity in plant vascular tissue cells (Lodish *et al.* 2000). Through this mechanism, the cell wall loses its flexibility and becomes thicker to form a consolidated structure (Ralet *et al.* 2016).

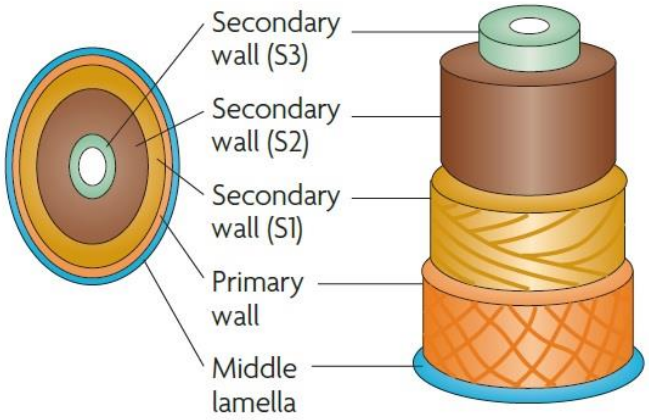


Figure 1.2. Cell wall structure.

(Adapted from Sticklen 2008)

1.2.2. Plant cell wall components

The primary cell wall polysaccharides can be divided into three groups: cellulose, hemicellulose and pectin. Cellulose represents the major component in cell wall, hemicellulose is the second most abundant structure and the last one is pectin polysaccharides (Harholt *et al*. 2010; Scheller and Ulvskov 2010). In the other hand, the secondary cell wall contains, in addition to cellulose and hemicellulose, lignin (Sticklen 2008). Quantitatively, cellulose is the largest fraction in secondary cell wall, followed by lignin and hemicellulose. Other components present in small amount are pectin, fats, oils, proteins and glycoproteins and extractives (Rose and Lee 2010; Shrotri *et al*. 2017).

1.2.2.1. Cellulose

Cellulose is a polysaccharide composed of 1,4-β-D-glucan chains that interact with each other via hydrogen bonds to form a crystalline microfibril. It is a linear polymer with a chain length of repeating units ranging from 140 or less to 10,000 or more depending on the source (Hallac and Ragauskas 2011). The cellulose chains are parallel and the successive glucose residues are rotated 180º, forming a flat ribbon in which cellobiose is the repeating unit (Figure 1.3) (Lodish *et al.* 2000; Somerville 2006). The chains pack together to form rodlike microfibrils, which are stabilized by hydrogen bonds between the chains (Lodish *et al.* 2000).

Cellulose is a polymer that can assume different crystalline structures and, until recently, it was assumed that there were six polymorphic forms of cellulose (O'Sullivan 1997). Cellulose I represents the native material present in plants, and it is this polymorph that is composed of two distinct crystal structures or types called cellulose Iα and Iβ. Cellulose Iα exists as a single-chain triclinic unit cell, whereas cellulose Iβ has a two-chain

monoclinic unit cell (Heiner *et al*. 1995; Brown Jr 1996; Somerville 2006). Cellulose I allomorphs can be converted to cellulose II as a result of acid regeneration or mercerization, realizing a form that can be more readily hydrolysed. Cellulose III_I and III_{II} can be produced by the ammonium treatment of cellulose I and II, respectively (Marrinan and Mann 1956; Hayashi *et al*. 1975), whilst these cellulose III allomorphs can be converted to their respective IV allomorphs (IV_I and IV_{II}) by heating in glycerol to 206 °C (Gardiner and Sarko 1985). In addition, cellulose chains may align in parallel (Type I) or antiparallel (Type II) orientation to each other (Sugiyama *et al*. 1991).

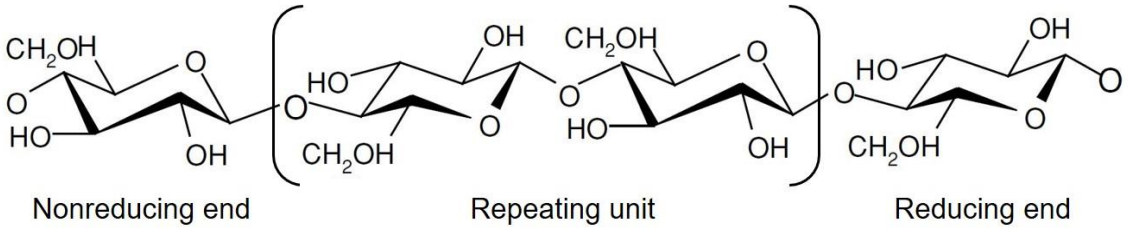


Figure 1.3. Cellulose representation.

(Adapted from Shrotri *et al.* 2017)

1.2.2.2. Hemicellulose

Hemicelluloses are heterogeneous polysaccharides that have different backbones and show structural side-chain variations as a result of different types and distributions of substituents along the backbone. In addition, hemicellulosic polysaccharides bind tightly to cellulose microfibrils via hydrogen bonds and most wall models have incorporated this interaction as one important feature of cell wall architecture (Keegstra 2010). These group of polysaccharides include, among other polysaccharides, xylans, xyloglucans, mannans, glucomannans and mixed-linkage glucans (Figure 1.4) (Ebringerová 2006; Scheller and Ulvskov 2010; Zhou *et al*. 2016).

Xylans represent the largest group of hemicellulose in the PCW of grass plants and can account for up to 50% (w/w) of the hetero-polysaccharides (Ebringerová 2006). Xylans are a diverse group of polysaccharides with the common feature of a backbone of 1,4-β-D-xylan residues (Figure 1.4a). A common modification of xylans is a substitution with 1,2-α-linked glucuronosyl which are often known as glucuronoxylans. Xylans also can contain arabinose attached at positions 2 and/or 3 and are known as arabinoxylans and glucuronoarabinoxylans (Scheller and Ulvskov 2010). Other residues, such as glucuronic acid and ferulic acid esters, are also attached in arabinoxylans, and these are particularly abundant in cereal grasses (Cosgrove 2005).

Xyloglucan is the most abundant hemicellulose in the primary walls of spermatophytes plants with exception of grasses (Scheller and Ulvskov 2010). Xyloglucan

has a backbone structure of 1,4-β-D-glucan residues, like cellulose (Figure 1.4b). However, xyloglucan is heavily decorated with side chains of α-D-xylose residues linked to the C-6 of backbone glucose residues (Caffall and Mohnen 2009). The xylose can also be serially appended with galactose (Gal) and fucose (Fuc) residues (Cosgrove, 2005).

Mannan is the major fraction of gymnosperms (Vries and Visser 2001). Mannan is divided into galactomannan (Figure 1.4c) and glucomannan (Figure 1.4d) based on structure properties (Moreira and Filho 2008) but both mainly consist of (1→4)-β-linked mannosyl units. Galactomannan has a mannose chain decorated with galactose residues connected by 1,6-α-glycosidic linkage (Chaubey and Kapoor 2001; Tamaki *et al*. 2010), whereas glucomannan possesses both glucose and mannose in the main chain and is decorated with galactose side chain in different degrees (Figure 1.4) (Maeda *et al*. 1980; Teleman *et al*. 2003).

Mixed-linkage β-glucans (1,3-1,4-β-glucans) (Figure 1.4e) are mainly present in the primary cell wall of cereal kernels, such as oat and barley (Scheller and Ulvskov 2010). In general, three to four (1→4)-β-linked glycosyl units are linked with each other via (1→3)-β-linkages, but longer (1→4)-β-linked segments have also been reported (Bulone *et al*. 1995; Fry *et al*. 2008). They are associated with cellulose microfibrils during cell growth (Ebringerová 2006).

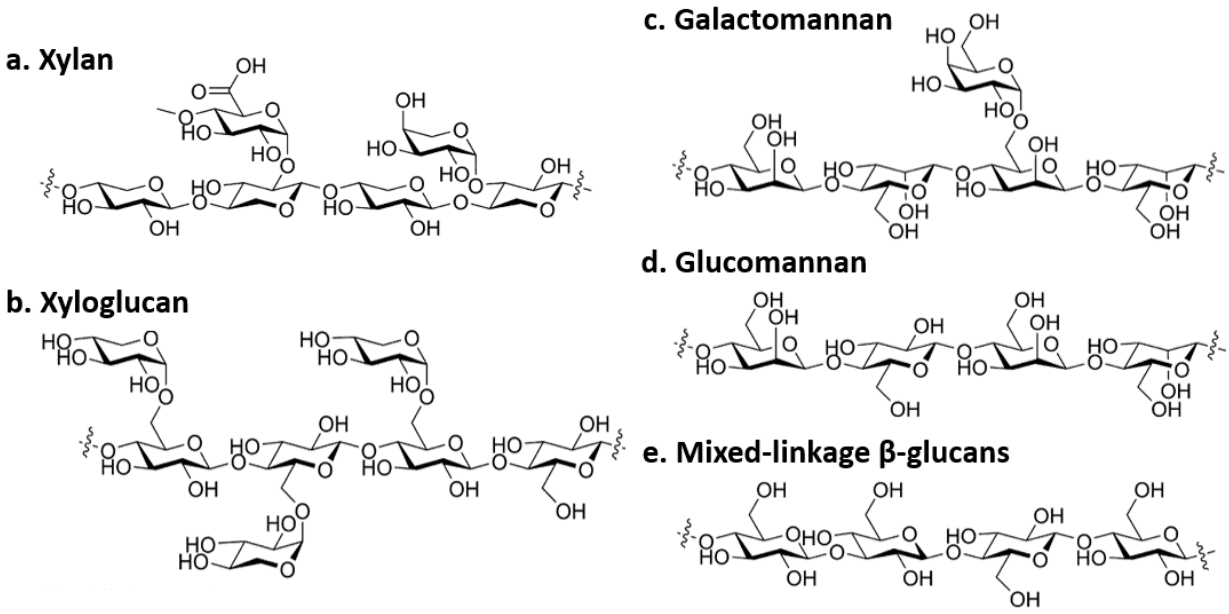


Figure 1.4. Structure of different types of hemicelluloses.

(Adapted from Shrotri *et al.* 2017)

1.2.2.3. Pectin

Pectins are also highly heterogeneous polysaccharides, traditionally characterized by being relatively easily to extract with hot acid or chelators and by containing a backbone 1,4-linked α-D-Gal*p*A residues (Vincken *et al*. 2003; Scheller and Ulvskov 2010). Various pectic polysaccharides can be detected in plant cell walls, including, homogalacturonans (HG), rhamnogalacturonans I and II (RGI and RGII), and substituted galacturonans (xylogalacturonan (XGA) and apiogalacturonan (AGA)) (Harholt *et al.* 2010). The ratio between HG, XGA, RGI and RGII is variable but typically HG is the most abundant polysaccharide, constituting about 65% of the pectin, while RGI constitutes 20% to 35% (Mohnen 2008). XGA and RGII are minor components, each constituting less than 10% (Zandleven *et al*. 2007; Mohnen 2008). The different pectic polysaccharides are not separated molecules but covalently linked domains (as shown in Figure 1.5) (Harholt *et al.* 2010). The complexity of the pectic polysaccharides and their conservation, to a greater or lesser degree, throughout the plant kingdom, infers specific and important biological functions in plant cell walls (Caffall and Mohnen 2009).

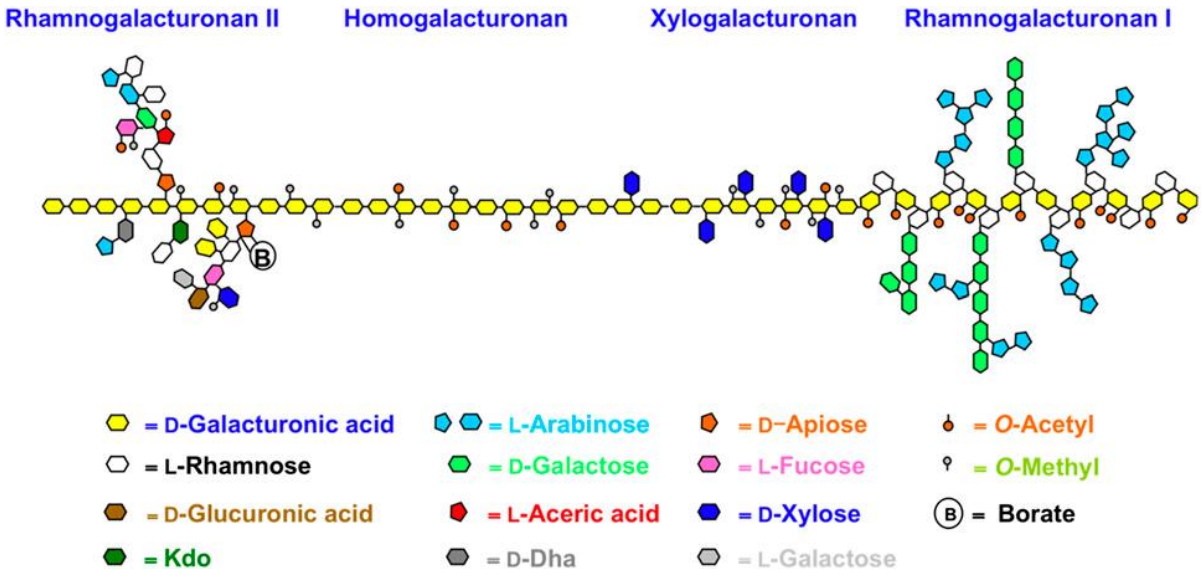


Figure 1.5. Schematic structure of pectin.

Pectin consists of four different types of polysaccharides (RGII, HG, XGA and RGI). (Kdo, 3-Deoxy-D-manno-2-octulosic acid; DHA, 3-deoxy-D-lyxo-2-heptulosaric acid) (Harholt *et al.* 2010)

Homogalacturonan (HG) is the simplest and the most abundant pectic polysaccharide and it consists of a linear backbone of 85–320 1,4-linked α-D-Gal*p*A residues (Figure 1.5) (Thibault *et al*. 1993; Hellín *et al*. 2005; Round *et al*. 2010). HG Gal*p*A residues may be methyl-esterified at the C-6 carboxyl or acetylated at the O-2 or O-3 (Figure 1.6) (Ridley *et al*.

2001). The pattern and degree of methylesterification and acetylation varies from source to source and determines the industrial applicability of pectin (Caffall and Mohnen 2009).

Rhamnogalacturonan I (RGI) has repeating units of [→2)-α-L-Rha*p*-(1→4)-α-D-Gal*p*A-(1→]$_n$, where *n* can be larger than 100 (McNeil *et al*. 1980; Visser and Voragen 1996; Vincken *et al*. 2003). Partial acetylation often occurs at the O-2 and/or O-3 positions of the Gal*p*A residues. The rhamnosyl residues can be substituted at O-4 with neutral sugars (McNeil *et al*. 1980; Lau *et al*. 1987). Depending on the plant species, the RGI backbone is decorated with galactans (1,4-β-D-galactose units) and/or arabinans (1,5-α-linked L-arabinofuranose units with additional L-arabinofuranose side-chains) (Figure 1.5) (Luis *et al.* 2018).

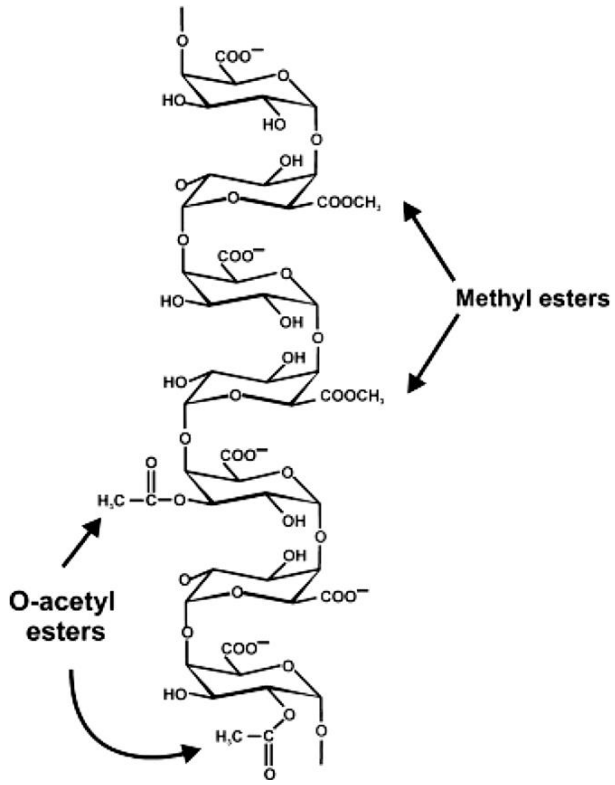


Figure 1.6. Homogalacturonan structure and representative sites of methylesterification at the C-6 and O-acetylation at the O-2 or O-3 of the carbohydrate ring.

(Ridley *et al.* 2001)

The structure of Rhamnogalacturonan II (RGII) is highly complex with 12 different types of glycoside residues, including the rare sugar species 2-O-methyl xylose, 2-O-methyl fucose, aceric acid, 2-keto-3-deoxy-D-lyxo heptulosaric acid (Dha), and 2-keto-3-deoxy-D-manno octulosonic acid (Kdo) (Figure 1.5) (Caffall and Mohnen 2009).

Finally, Xylogalacturonan (XGA) is HG modified by the addition of D-xylose residues at the O-3 of Gal*p*A backbone residues (Figure 1.5) (Cosgrove 2005; Caffall and Mohnen

2009). In addition, in apiogalacturonan (AGA) the galacturonan backbone is decorated with D-apiofuranose at O-2 and/or O-3 (Hart and Kindel 1970; Ovodov *et al*. 1971; Longland *et al*. 1989).

1.2.2.4. Lignin

Lignification is the last important process of cell wall development, accounting for increased mechanical strength and protection against pathogen attack (Wei *et al.* 2009). Lignin has also the function of protecting cellulose and hemicellulose from enzymatic hydrolysis (Schoenherr *et al*. 2018). Around 25% of all the lignin in wood is found in the middle lamella (Dinwoodie 2000), and the remaining 75% is part of secondary cell walls being deposited following the completion of the cellulosic frame work. Functionally, lignin reinforces plant cell walls through bonding with cellulose. In addition, lignin also enhances the waterproof nature of plant cell walls as a result of its hydrophobicity, while allowing the efficient transport of water in the vascular tissues (Zhou *et al.* 2016).

Lignin is a highly complex noncrystalline molecule comprised of a large number of phenylpropane monomers (Desch and Dinwoodie 1996). Primarily, three monolignols (monomers) p-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol are present in varying degrees in native lignin. These monolignols are often referred to as phenylpropanoids, which differ in the substitutions at the 3-C and 5-C positions in the aromatic ring (Faix 1991; Wong 2009). Lignin synthesis starts with the random self-replicating radical coupling of phenoxy radical to form an oligomeric product. After polymerization, these polymers are referred as p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) (from p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol, respectively) (Figure 1.7). Monolignols are linked either by C–C bond or C–O–C bond, and more than two third of monolignols are joined by ether linkages (Parthasarathi *et al*. 2011). The distribution of these monomers varies in different plant species and tissues. Generally, lignin from grasses is a roughly equimolar mixture of G, S and H units, whereas lignin from hardwood contains approximately equal quantities of G and S units but relatively small amounts of the H unit, and lignin from softwood is mainly composed of G units (up to 90%) (Faix 1991).

In the plant cell wall, lignin is built into a network with hemicellulose via ester and ether linkages. These linkages are formed between lignin and residues of hemicellulose, which comprise glucuronic acid or arabinosyl-ferulic acid substituents (Takahashi and Koshijima 1988; Jacquet *et al*. 1995; Lam *et al*. 2001).

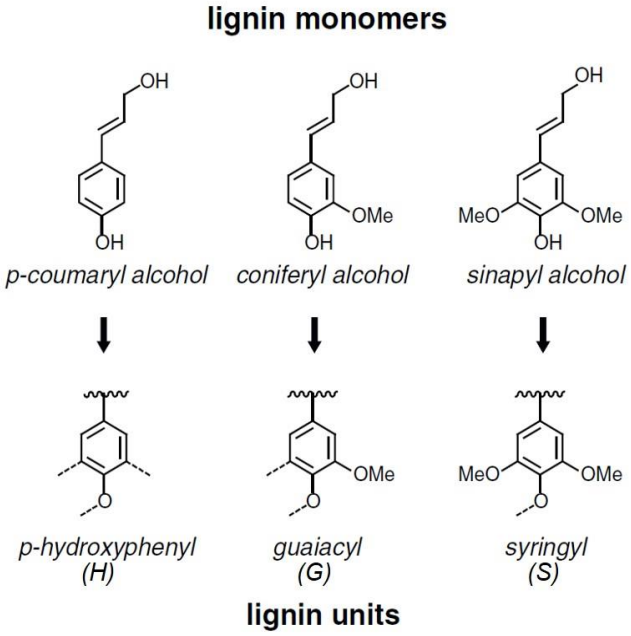


Figure 1.7. Chemical structure of monolignols and the corresponding units in lignin.

(Brown and Chang 2014)

1.2.3. Plant cell wall models

Presently, the organization and interactions established between the repertoire of plant cell wall components remains unclear. Several models have been proposed to represent plant cell wall organization while accounting a dynamic nature appropriate to allow cells to expand and grow. Keegstra and colleagues, in 1973, proposed that the cell wall matrix polymers (xyloglucan, pectin and glycoproteins) are covalently linked to form a giant macromolecular network. In this model, cellulose interacts with the matrix via H-bonding, predominantly to xyloglucan (Figure 1.8-A). More recently, Hayashi (1989) and Fry (1989) proposed that cellulose microfibrils may be tethered together directly via long xyloglucan chains. The cellulose-xyloglucan network is enmeshed in a non-covalently cross-linked pectic network (Figure 1.8-B). This model is currently the most popular one (Carpita and Gibeaut 1993), although, variations to this general organization have been proposed. Talbott and Ray (1992) proposed a model in which each cellulosic microfibril is coated by a series of progressively less-tightly bound polysaccharide layers (Figure 1.8-C) and the linkage between microfibrils is made indirectly via the lateral (non-covalent) associations between the distinctive polysaccharide layers. Ha *et al*. suggested that xyloglucan molecules are hydrogen bonded to and cross-link cellulose microfibrils and these cellulose-xyloglucan lamellae are separated by strata of pectic polysaccharides (Figure 1.8-D) (Cosgrove 2001).

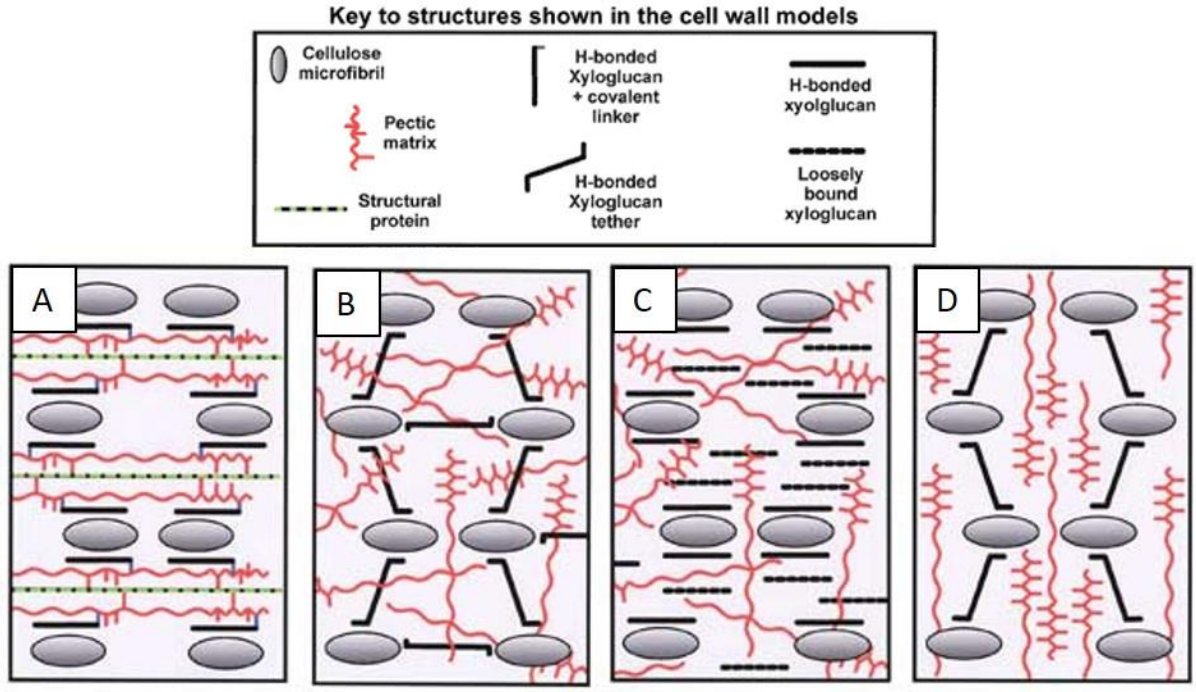


Figure 1.8. Models of cell wall structure.

A) The covalently cross-linked model; B) The tether model; C) The diffuse layer model, and D) The stratified layer model. (Adapted from Albersheim *et al.* 2007)

Clearly much research is still required to provide a complete description of the plant cell walls. It is, however, established that primary walls are dynamic structures whose composition and architecture changes during plant growth and development.

1.2.4. Plant cell wall degradation

As referred above, ligno-cellulosic biomass is the most abundant source of organic carbon in the biosphere and the recycling of plant cell wall structures is a general process of considerable biological importance (Hervé *et al.* 2010). The fixed carbon is recycled by microbial enzymes, generally termed as CAZymes, that convert cell wall polysaccharides to monosaccharides and oligosaccharides. Understanding how microbes deconstruct cell walls is also of growing industrial significance for the biofuel and bioprocessing sectors (Figure 1.9) (Sticklen 2008; Himmel and Bayer 2009; Hervé *et al*. 2010).

Microbial communities secrete a large repertoire of enzymes that act sequentially and in a synergistic manner to degrade plant cell walls (Wilson 2008; Wei *et al*. 2009). It is believed that synergy between different types of enzymes produced by bacteria and/or fungi is crucial to effectively degrade plant cell wall components (Wei *et al.* 2009). The plant cell wall degrading microorganisms have been shown to use two different approaches for cellulose degradation. Aerobic microorganisms secrete large quantities of modular enzymes, usually containing a catalytic module linked to one or more non-catalytic carbohydrate-

binding module (CBM) (Lynd *et al*. 2002; Hashimoto 2006). In aerobes, enzymes are either secreted into the extracellular milieu or are located on the outer membrane. Although these enzymes do not physically associate, they do display extensive biochemical synergy. In contrast, anaerobic bacteria and fungi organize enzymes in high molecular mass multienzyme complexes, called cellulosomes, which are usually attached to the outer surface of the microorganism (Wilson 2008). It is believed that the anaerobic environment impose a greater selective pressure for the evolution of these highly efficient nanomachines (Bayer *et al*. 2004). In previous studies, it has been proposed that cellulosomal systems may have greater activity against recalcitrant plant cell walls than non-aggregated cellulase enzymes (Wei *et al.* 2009).

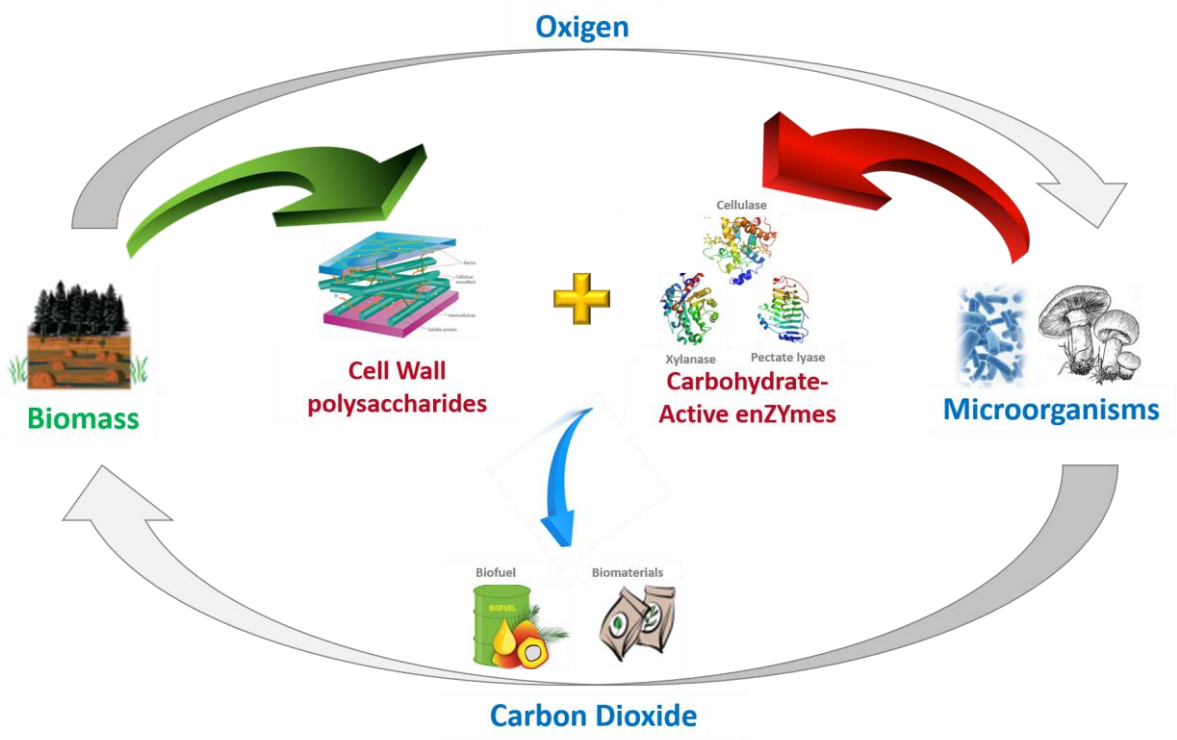


Figure 1.9. Representation of cell wall deconstruction.

Ligno-cellulosic biomass is the most abundant source of organic carbon in the biosphere. Its degradation is a process of growing interest, particularly in the biofuels and bioprocessing sectors, and can be performed by microbial enzymes, which are generally referred as Carbohydrate-Active Enzymes.

1.2.4.1. Carbohydrate-Active Enzymes (CAZymes)

Carbohydrate-Active Enzymes (CAZymes) comprise a diversity of enzymes classes including glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and glycosyl transferases (GTs). While the PLs, CEs, and GHs carry out the breakdown of polysaccharides, the GTs are mainly involved in the formation of the glycosidic

bond and thus in the biosynthesis of carbohydrates (Chakraborty *et al*. 2017). Recently, the CAZy database incorporated a new category, named Auxiliary Activities (AAs), which covers redox enzymes that act in conjunction with CAZymes. This category groups families of lytic polysaccharide monooxygenases (LPMOs) as well as ligninolytic enzymes (Levasseur *et al*. 2013).

The enzymes are classified in families according to their amino acid sequence similarities which reflects common structural folds. This classification usually reflects a common evolutionary origin and conservation on the catalytic mechanisms, protein fold and structural features, much better than the Enzyme Commission Number (EC) that is a numerical classification scheme for enzymes, based only on the chemical reactions they catalyse. While the structures of CEs, PLs, and CBMs are dominated by the α/β-hydrolase (Correia *et al.* 2008), parallel β-helix (Pickersgill *et al*. 1994), and jelly roll (or β-sandwich) (Czjzek *et al.* 2001) folds, respectively, there are a large number of different folds within the GHs (Gilbert 2010).

The classification of CAZymes is continuously updated in Carbohydrate-Active EnZymes database (CAZy; www.cazy.org) (Cantarel *et al.* 2009; Lombard *et al.* 2014). Currently, the database contains 165 sequence-based families of GHs, 37 families of PLs, 15 families of CEs, 107 families of GTs, 16 families of AAs and finally, 85 families of CBMs (data collected on July 2019).

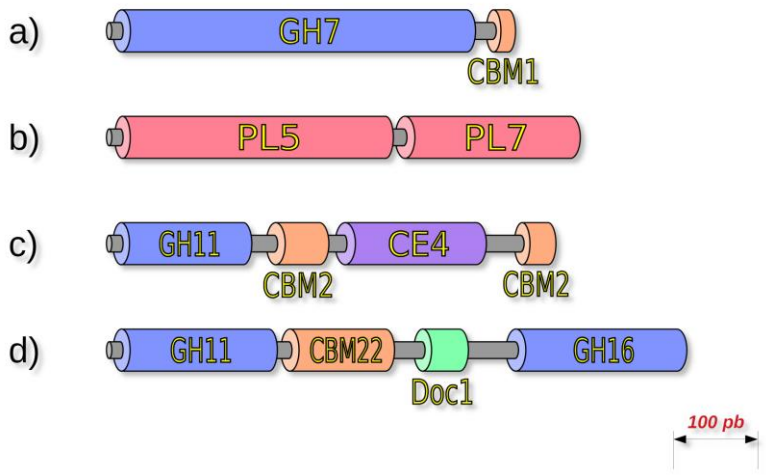


Figure 1.10. Modular architecture of Carbohydrate-Active enZymes (CAZymes).

Examples of modular CAZymes, composed by GHs, PLs, CEs, dockerin (Doc) and CBMs from different CAZy families. a) cellobiohydrolase I from *Hypocrea jecorina* (SP P00725); b) alginate lyase from *Sphingomonas* sp. A1 (GB BAB03312.1); c) xylanase from *Cellulomonas fimi* (GB CAA54145.1); d) xylanase D/licheninase from *Ruminococcus flavefaciens* (GB CAB51934.1).

(Adapted from Cantarel *et al*. 2009)

CAZymes are frequently modular, containing one or more catalytic domains connected by highly flexible linker sequences to the accessory non-catalytic carbohydrate-binding modules (CBMs) (Boraston *et al*. 2004; Hashimoto 2006; Guillén *et al*. 2010) (Figure 1.10). These CBMs, which do not exhibit catalytic activity, were also classified according to primary sequence similarity in families (Lombard *et al*. 2014).

1.2.4.1.1. Glycoside hydrolases (GHs)

Glycoside hydrolases or glycosidases (EC 3.2.1.-) are enzymes that catalyse the hydrolysis of the glycosidic linkage of glycosides between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. These enzymes are highly versatile tools in chemical glycobiology and important biocatalysts in industrial and biotechnological processes, catalysing the hydrolytic degradation of carbohydrates (Kötzler *et al*. 2014).

GHs can be classified in many different ways. The simplest classification method is based upon the substrate specificity of enzyme, as recommended by the International Union of Biochemistry and Molecular Biology (IUBMB) (Zhang *et al*. 2010). EC numbers do not specify enzymes, but enzyme-catalysed reactions. If different enzymes catalyse the same reaction, then they receive the same EC number. A necessary consequence of the EC classification scheme is that codes can be applied only to enzymes for which a function has been biochemically identified. Additionally, certain enzymes can catalyse reactions that fall in more than one class. These enzymes must bear more than one EC number (Withers and Williams 2017).

A new concept based on a sequence-based classification was introduced years ago by Henrissat in the form of a ‘subfamily’ categorization (Stam *et al*. 2006), which provides a useful tool for further correlating primary sequences of glycosidases with their substrate specificities. This is especially important for large glycosidase families where different enzymatic activities and substrate preferences are found, within the same family (Zhang *et al.* 2010). Classification of families into larger groups, termed 'clans' has also been proposed (Henrissat and Bairoch 1996). A 'clan' is a group of families that possess significant similarity in their tertiary structure, catalytic residues and mechanism. Families within clans are thought to have a common evolutionary ancestry (Withers and Williams 2017).

Finally, GHs can be grouped based on their active site topology and it has been found that active site architectures fall, generally, into three categories (which strongly correlate with the action pattern of the corresponding enzyme) (Davies and Henrissat 1995). Enzymes having pocket-shaped active sites usually recognize and cleave the nonreducing ends of saccharide structures in an ‘exo’ mode of action. The second type contains a cleft-shaped active site. These ‘open’ structures allow access to the polymeric substrate in a relatively random manner, thus their action mode is ‘endo’ and the internal linkages of the

polysaccharide chain are cleaved. The last type of active site topology is tunnel-like. This topology probably evolved from the ‘clefts’ by ‘closing’ them with long loops. The advantage of a ‘tunnel-like’ active site probably lies in the increase of processivity since polymeric substrates thread through the tunnel and the reaction product remains bound close to the enzyme active site, ready for the next cleavage reaction (Rouvinen *et al*. 1990).

Catalytic Mechanism

Glycoside hydrolases can be divided in two mechanistic classes. The first class of GHs are known as inverting enzymes and the second class are known as retaining enzymes (Figure 1.11) (Withers 1995). In the case of the inverting enzymes, the hydrolysis of a glycoside with net inversion of anomeric configuration is generally achieved via a one step, single-displacement mechanism involving oxocarbenium ion-like transition states (Figure 1.11a). The reaction typically occurs with general acid and general base assistance from two amino acid side chains, normally glutamic or aspartic acid (McCarter and Withers 1994; Withers 1995). In contrast, in retaining enzymes, the hydrolysis with net retention of configuration is most commonly achieved via a two-step, double-displacement mechanism involving a covalent glycosyl-enzyme intermediate (Figure 1.11b). In the first step, one of the carboxyl groups functions as a general acid catalyst, protonating the glycosidic oxygen concomitantly with bond cleavage. The second catalytic residue acts as a nucleophile, forming a covalent glycosyl-enzyme intermediate. In the second step, the side-chain carboxylate deprotonates the incoming water molecule, which attacks at the anomeric center and displaces the sugar. Both steps occur via transition states with substantial oxocarbenium ion character (McCarter and Withers 1994; Wang *et al*. 1994; Davies and Henrissat 1995; Withers 1995).

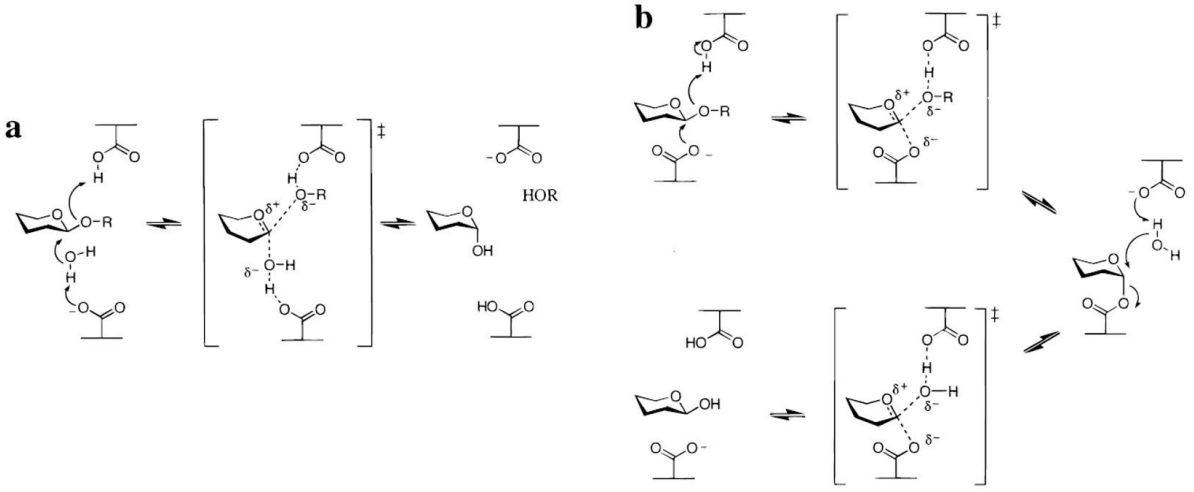


Figure 1.11. General mechanism for (a) inverting and (b) retaining glycosidases.

(Zechel and Withers 2000)

1.2.4.1.2. Polysaccharide lyases (PLs)

Polysaccharide lyases (EC 4.2.2.-) are a group of enzymes that cleave uronic acid-containing polysaccharides via a β-elimination mechanism to generate an unsaturated hexenuronic acid residue with a new reducing end at the point of cleavage (Yip and Withers 2006; Garron and Cygler 2010; Lombard *et al*. 2010). These enzymes act on various substrates, all polyanionic in nature, yielding products in which the non-reducing terminus is modified to form an unsaturated uronic acid. Most of the enzymes in this class present a randomly endolytic catalytic mechanism and, with the exception of xanthan lyases, cleave predominantly the main chain of polysaccharide structures (Sutherland 1995).

PLs classification in families and subfamilies are based on amino acid sequence similarities, intended to reflect their structural features (Lombard *et al.* 2010). Just as for GHs, the sequence-based families of PLs are frequently polyspecific (i.e. contain enzymes acting on different substrates or that generate different products) (Lombard *et al.* 2010). For the purpose of this family classification, the scope of the term PL is restricted to those enzymes which operate according to the general syn- and anti-elimination mechanisms described in Lombard *et al.* (2010), to produce a terminal hexenuronic acid moiety by β-elimination. This constitutes a clear distinction from the broader IUBMB classification of carbon-oxygen lyases acting on polysaccharides under EC 4.2.2.-, where other enzyme mechanisms have been described (Henrissat *et al*. 2019). These enzymes show a large variety of fold types (or classes), suggesting that PLs have been invented more than once during evolution from totally different scaffolds (Garron and Cygler 2010; Lombard *et al.* 2010).

Catalytic Mechanism

The catalytic mechanism involved in cleaving the polygalacturonic chain by PLs is broadly divided into three events. The first is the abstraction of the C-5 proton of the uronic acid by a basic amino acid (arginine, lysine, and histidine) side chain, the second is stabilization of the resultant anion into the C-6 carbonyl group by charge delocalization and the last step is the cleavage of the 1,4-α-glycosidic bond, facilitated by proton donation from an acidic amino acid (aspartate and glutamate) at the catalytic centre of the enzyme, producing an unsaturated hexenuronan moiety and a reducing end (Yip *et al*. 2004; Yip and Withers 2006). Depending on the monosaccharide composition of the substrate and its conformation at the active site of PL, the proton removed from C-5 and the departing oxygen on C-4 may lay either syn or anti to each other. Substrate recognition by PLs is often mediated by the interaction of bivalent cations (often Ca^{2+}), or positively charged amino acid side chains, with uronic acid groups in the substrate. The bivalent cation usually stabilizes the transient anion in the reaction pathway (Yip and Withers 2006; Lombard *et al*. 2010).

1.2.4.1.3. Carbohydrate esterases (CEs)

Carbohydrate esterases (CEs) are enzymes catalysing the O-de- or N-diacylation of substituted saccharides, i.e. esters or amides in which sugars play the role of alcohol and amine. For CEs also considered enzymes that hydrolyse esters in which sugars play the role of acid. The form of sugar acids involved in such esters is uronic acids (Biely 2012). A number of possible reaction mechanisms may be involved, although CEs generally hydrolyse ester linkages through a double displacement mechanism in which Ser (Schubot *et al.* 2001) or Asp (Fries *et al*. 2007) functions as the catalytic nucleophile. Exceptions to this mode of action are apparent in family CE4, where catalysis is metal dependent (Taylor *et al.* 2006).

CEs are currently classified in 15 families in the Carbohydrate-Active Enzyme (CAZy) database (CE1 to CE16 but CE family 10 has been nullified). Thus, a large variety of enzymes are grouped within the CAZy database and these assemble, modify and breakdown carbohydrates and glycoconjugates (Nakamura *et al*. 2017). Several esterases, mainly feruloyl esterases (FAEs), do not fit into the established CE families and have been separately classified based on biochemical or sequence similarities (Udatha *et al*. 2011). The range of biological and biotechnological applications of CEs is diverse. For instance, the majority of families include members that catalyse the removal of ester-based modifications from mono-, oligo- and polysaccharides. Therefore, by removing the acylated moieties of polysaccharides, these CEs could accelerate the degradation of these polymers facilitating the access of GHs, thus assisting in biomass saccharification (Christov and Prior 1993).

1.2.4.1.4. Auxiliary Activities (AAs)

Auxiliary Activities (AAs) cover redox enzymes that act in conjunction with CAZymes. The discovery that members of families CBM33 and family GH61 are lytic polysaccharide monooxygenases (LPMO), gave impetus to a reclassification of these families into a new category, the Auxiliary Activities. In addition, since lignin is invariably found together with polysaccharides in the plant cell wall and because lignin fragments are likely to act in concert with LPMOs, the authors decided to join the families of lignin degradation enzymes and LPMO families in this new CAZy class Auxiliary Activities, in order to accommodate a range of enzyme mechanisms and substrates related to ligno-cellulose conversion (Levasseur *et al.* 2013). Currently, AAs contain 9 families of ligninolytic enzymes and 6 families of LPMO. Like the traditional CAZy families, the new families are based on sequence similarity with one or several biochemically-characterized founding member(s), ensuring that members of a given family share the same three-dimensional structure (Levasseur *et al.* 2013). Today, it is widely recognized that LPMOs are powerful enzymes that oxidatively cleave glycosidic bonds in polysaccharides (Frommhagen *et al*. 2018), such as cellulose (AA9, AA10), hemicellulose (AA9, AA14) (Couturier *et al.* 2018), chitin (AA10, AA11) and starch (AA13) (Breslmayr *et al.*

2018). LPMOs have been shown to improve the activity of hydrolytic enzyme cocktails and are now seen as essential components for ligno-cellulosic degradation (Harris *et al*. 2010; Hu *et al*. 2015).

1.2.4.1.5. Glycosyltransferases (GTs)

The biosynthesis of disaccharides, oligosaccharides and polysaccharides involves the action of hundreds of different glycosyltransferases (GTs) (EC 2.4.x.y). Glycosyltransferases can be broadly defined as enzymes that catalyse the transfer of glycosyl residues from their specific donor to an acceptor molecule (Sinnott 1990). Glycosyltransferases can be classified as either retaining or inverting enzymes according to the stereochemistry of the substrates and reaction products (Sinnott 1990). The recommendations of IUBMB do not indicate the intrinsic structural features of the enzymes, nor do they adequately accommodate enzymes which act on several distinct substrates (Williams 2017). Thus, as it was described above for other classes of CAZymes, GTs were also classified into families based on amino acid sequence similarities (J.A. Campbell *et al*. 1997; Coutinho *et al*. 2003).

In contrast to the widely diversified structures of glycosidases, only two general folds have been found for the nucleoside phosphor sugar-dependent GTs characterized to date, namely, GT-A and GT-B (Hu and Walker 2002). Both contain α/β/α Rossmann folds that are typical of nucleotide-binding proteins (Zhang *et al.* 2010). The GT-A fold is a glycosyltransferase protein topology consisting of two closely abutting β/α/β Rossmann domains. In contrast, the GT-B fold is a glycosyltransferase protein topology consisting of two β/α/β Rossmann domains that face each other and are linked flexibly (Lairson *et al*. 2008). A new glycosyltransferase fold, GT-C, was recently proposed from detailed sequence and structure comparison of the glycosyltransferases listed in the CAZy database (Liu and Arcady 2003). The GT-C fold is a predicted protein topology for transmembrane glycosyltransferases that is not experimentally verified (Lairson *et al.* 2008). Nevertheless, GTs families can be classified into clans depending on their fold and the stereochemical outcome of the reactions that they catalyse. Among GT-A and GT-B superfamilies, the overall fold of the enzyme does not dictate the stereochemical outcome of the reaction that it catalyses, as examples of both inverting and retaining glycosyltransferases have been identified within both the GT-A and GT-B fold classes (Coutinho *et al.* 2003).

Catalytic Mechanism

Two stereochemical outcomes are possible for reactions that result in the formation of a new glycosidic bond: the anomeric configuration of the product can either be retained or inverted with respect to the donor substrate (Figure 1.12). In this way GTs are classified as inverting or retaining enzymes, depending on the outcome of the reaction (Lairson *et al.*

2008). Inverting GTs catalyses group transfer reactions with net inversion of stereochemistry at the anomeric reaction centre of the donor substrate, i.e., inverting glycosyltransferases most likely follow a single displacement mechanism where the acceptor performs a nucleophilic attack at carbon C-1 of the sugar donor somewhat analogous to the mechanism of inverting glycosidases. In contrast, retaining GTs catalyses group transfer reactions with net retention of stereochemistry at the anomeric reaction center of the donor substrate, i.e., retaining glycosyltransferases do not appear to operate via a two-step mechanism involving the formation of a glycosyl-enzyme intermediate analogous to glycosidases (Lairson *et al*. 2008; Williams 2017).

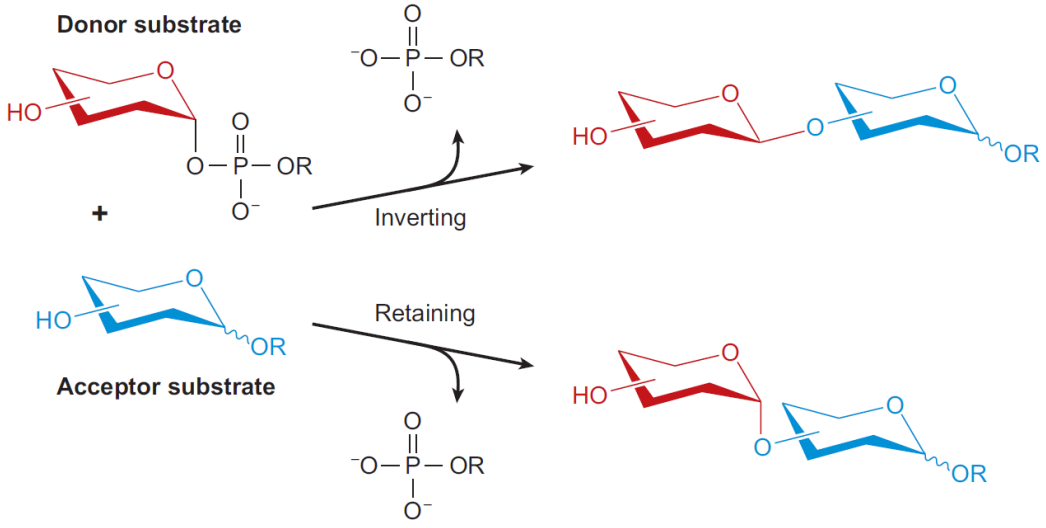


Figure 1.12. Stereochemical outcomes of glycosyltransferases.

(Lairson *et al.* 2008)

1.2.4.1.6. Carbohydrate-Binding Modules (CBMs)

Carbohydrate-binding modules (CBMs) play an important role in the ability of enzymes to degrade insoluble substrates but they have little or no effect on the activity. The main function of CBMs is to bind polysaccharides, bringing the catalytic domain into close and prolonged access with its substrate, leading to a more efficient carbohydrate hydrolysis, thereby increasing the catalytic efficiency (Gilbert 2010; Guillén *et al.* 2010).

Similar to CAZymes, CBMs have been classified in families based on amino acid sequence similarity. However, although ligand specificity varies widely in CBMs, common folds are observed in proteins with different specificities and belonging to different taxonomic groups. In addition, CBMs have also been classified into three types (A, B, and C) based on the topology of their ligand-binding sites and their mode of ligand recognition (Boraston *et al.* 2004). Type A CBMs bind strongly to insoluble polysaccharide surfaces, type B CBMs bind

to soluble glycan chains and type C CBMs bind to small saccharides (Boraston *et al.* 2004) (Figure 1.13).

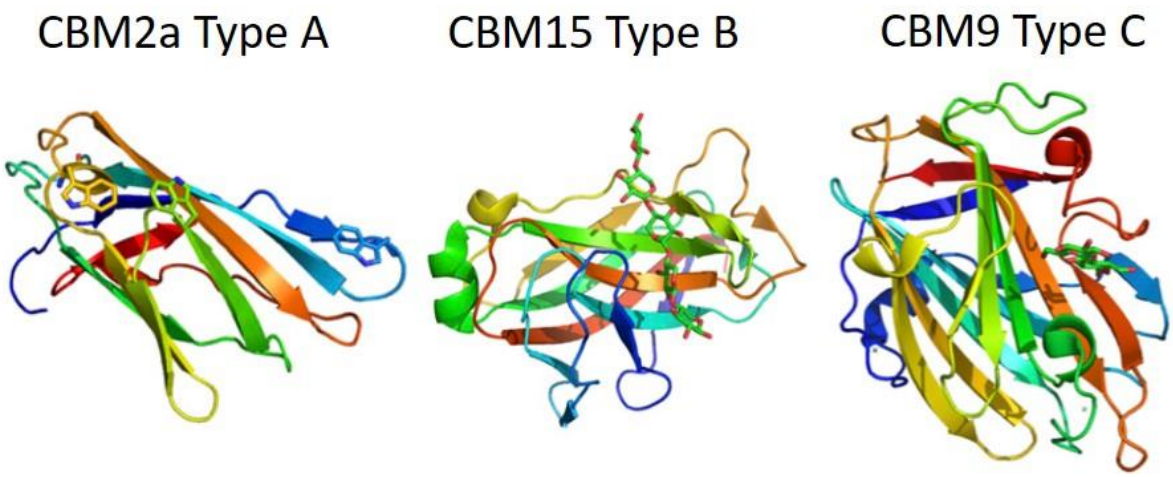


Figure 1.13. Examples of type A, type B and type C CBMs.

CBM2a is derived from the *Cellulomonas fimi* xylanase Xyn10A (Protein Data Bank [PDB] 1XG), CBM15 is a component of the *C. japonicus* xylanase Xyn10C (PDB 1GNY), and CBM9 is from a *Thermotoga maritima* GH10 xylanase (PDB 1I82). The folds are ramped from blue (N terminus) to red (C terminus) (Adapted from Gilbert 2010)

The CAZy database provides an updated list of the CBM-containing proteins, arranged by CBM family. Today, CBMs are categorized in various families and although the majority bind to cellulose, several CBM members present chitin binding activity or interact with hemicellulose, in particular with xylan, xyloglucan or mannan. In addition, some CBMs present a promiscuous specificity and can bind to several polymers, while others are specific for one or a restricted number of topological similar carbohydrates (Wilson 2008). In general, the ligand specificity of CBMs reflects the substrate cleaved by the cognate enzyme. Similar to the other CAZys, some of the CBM families can be grouped into super-families or clans.

1.2.4.2. Plant cell wall degrading enzymes

Each type of enzyme that can degrade structural polysaccharides is represented in multiple families that share sequence and structural similarities. Enzymes that are capable of hydrolytically cleaving glycosidic bonds in oligo- or polysaccharides (including cellulose and hemicellulose) are generally summarized GHs families (Henrissat and Bairoch 1993). Alternatively, enzymes that cleave pectin and lignin are dispersed in GHs, CEs, PLs and AAs families.

1.2.4.2.1. Cellulases

Cellulose deconstruction is mediated by three types of cellulases: endo-1,4-β-glucanases (E.C. 3.2.1.4), cellobiohydrolases or exo-1,4-β-glucanases (E.C. 3.2.1.91), and β-glucosidases (E.C. 3.2.1.21) (Ziegler *et al*. 2000; Ziegelhoffer *et al*. 2001). Endo-1,4-β-glucanases can hydrolyse the glycosidic bond at any site along a cellulose chain; cellobiohydrolases hydrolyse the cellulose chain only at one of its free termini (reducing or non-reducing) and then degrades the substrate in a processive manner and β-glucosidases hydrolyse only terminal, non-reducing glucose units from short cellodextrins (Figure 1.14) (Artzi *et al*. 2016).

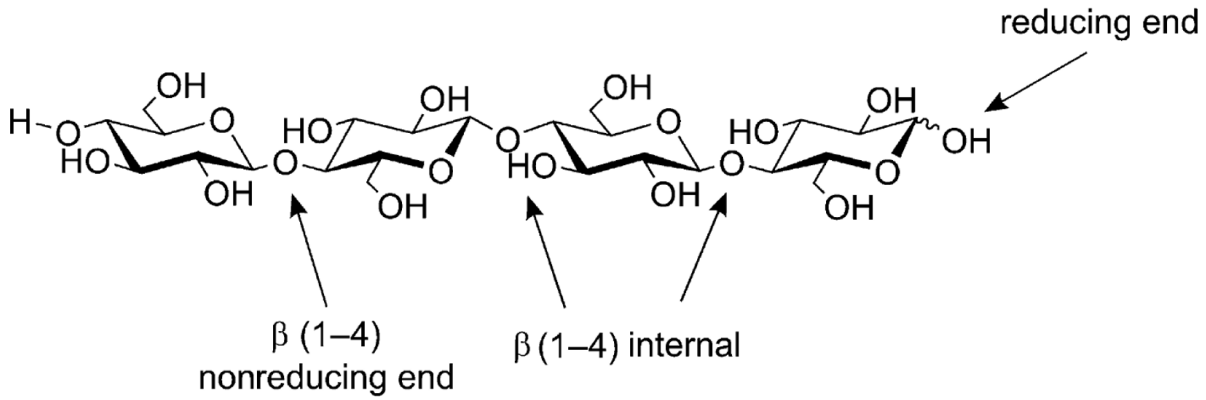


Figure 1.14. Structure of cellulose

(Adapted from Shrotri *et al.* 2017)

Classically, cellulose hydrolysis is viewed as a synergistic process. In this process, endo-1,4-β-glucanase initiate the attack by the random cleavage of different regions of crystalline cellulose, producing chain ends that act as substrates for cellobiohydrolase. This last enzyme firmly attaches to the cellulosic chain end and threads it through its active site, cleaving off cellobiose units. The cellobiohydrolase also acts on regions of amorphous cellulose with exposed chain ends without the need for prior endoglucanase activity. Finally, β-glucosidase breaks the bonds between the two glucose sugars of cellobiose to produce monomers of glucose (Figure 1.15) (Sticklen 2008). In other words, endo-1,4-β-glucanases are distributed throughout GH families 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 74, 124 and 148; cellobiohydrolases are found in GH families 5, 6 and 9; and β-Glucosidases are predominantly found in GH families 1, 2, 3, 5, 9, 16, 30, 39 and 116 (www.cazy.org).

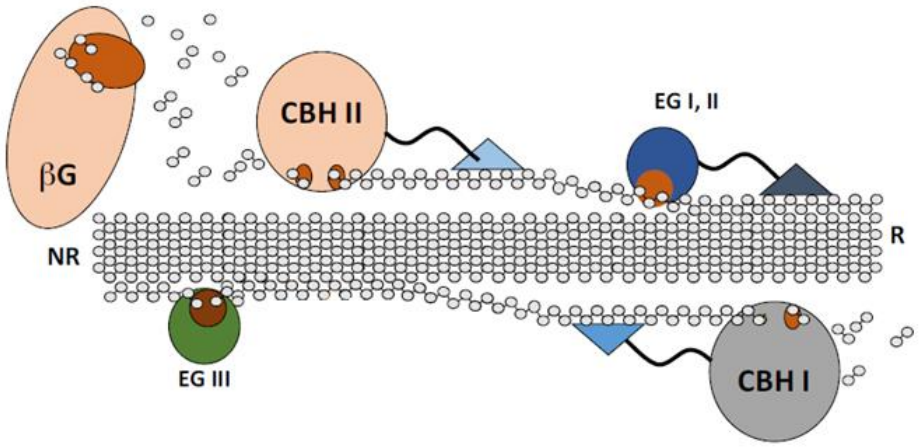


Figure 1.15. Illustration of cellulases action on cellulose.

(βG, β-Glucosidade; CBH, cellobiohydrolases; EG, endoglucanases; NR, non-reducing end of cellulose chain; R, reducing end of cellulose chain) (Adapted from Johansen 2016)

1.2.4.2.2. Hemicellulases

In general, it is assumed that the hemicellulose surrounding cellulose must be removed to allow the access of cellulases to their substrate chains. While cellulose consists of a single monosaccharide and type of bond, hemicelluloses are diverse and amorphous (Sticklen 2008). This group of polysaccharides includes a wide range of different sugars or non-sugar constituents connected via a diversity of covalent bonds (Artzi *et al.* 2016). The complex nature and structure of hemicellulose requires multiple enzymes to break down these polymers into their sugar monomers constituents.

The biodegradation of xylan depends of a repertoire of hydrolytic enzymes, including xylanases (endo-1,4-β-xylanases; E.C.3.2.1.8), β-xylosidases (xylan-1,4-β-xylosidases, E.C.3.2.1.37), α-glucuronidases (α-glucosiduronases; E.C.3.2.1.139), α-arabinofuranosidases (α-L-arabinofuranosidases; E.C.3.2.1.55) and acetyl xylan esterases (E.C.3.1.1.72) (Juturu and Wu 2012). All of these enzymes act cooperatively to convert xylan into its constituent sugars (Figure 1.16) (Belancic *et al.* 1995). The xylan is hydrolysed by xylanases from GH families 3, 5, 8, 9, 10, 11, 12, 16, 26, 30, 43, 44, 51, 62, 98 and 141; β-xylosidases from GH1, 2, 3, 5, 10, 30, 39, 43, 51, 52, 54, 116 and 120; α-glucuronidases from GH4 and GH67, α-arabinofuranosidases from GH families 2, 3, 43, 51, 54, 62 and 155; and acetyl xylan esterases from CE families 2, 3, 5, 10, 43, 51, 54, 62 and 155 (www.cazy.org).

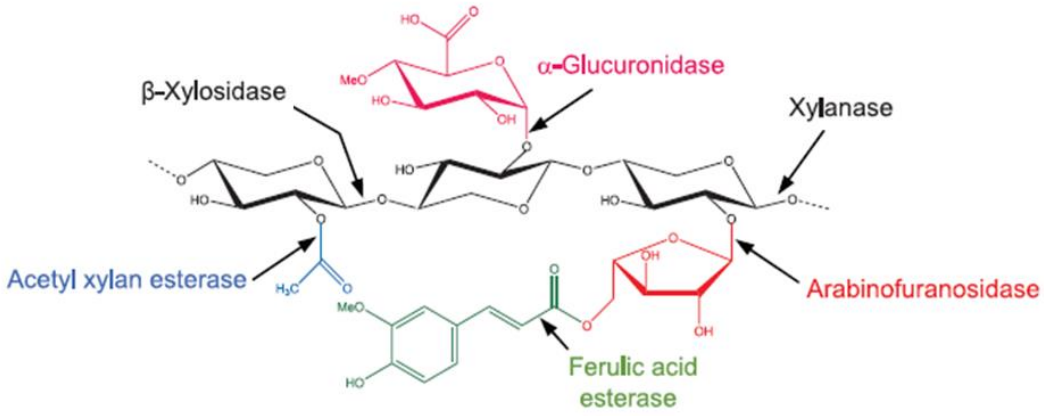


Figure 1.16. Structure of xylan and the xylanolytic enzymes involved in its degradation.

(Adapted from DeBoy *et al.* 2008)

The xyloglucan backbone is hydrolysed by specific endo-xyloglucanases (xyloglucan endo-hydrolases; E.C. 3.2.1.151) from GH families 5, 9, 12, 16, 26, 44, and 74. GH12 enzymes can tolerate the side chains in xyloglucan (Gilbert, 2010; www.cazy.org).

Mannans (including galactomannan and glucomannan) are susceptible to hydrolysis by three types of enzymes (Figure 1.17): β-Mannanases (mannan endo-1,4-β-mannosidases; EC 3.2.1.78), which catalyse the random hydrolysis of 1,4-β-mannosidic linkages within the backbones of mannans, galactomannans, and glucomannans; β-Mannosidases (1,4-β-mannoside mannohydrolases; EC 3.2.1.25), which cleaves the terminal 1,4-β-linkages from the nonreducing ends of oligosaccharides; and finally, α-Galactosidases (α-D-galactoside galactohydrolases; EC 3.2.1.22), which remove 1,6-α-linked galactose residues from galactomannan polymers (Stoll *et al*. 1999; Vries and Visser 2001; Mahammad *et al*. 2006). In resume, β-mannanases are GH5, 9, 26, 44, 113 and 134, while β-mannosidases are GH1, 2 and 5 enzymes and α-Galactosidases are GH4, 27, 31, 36, 57, 97 and 110 (Gilbert, 2010; Stoll *et al.* 1999; www.cazy.org).

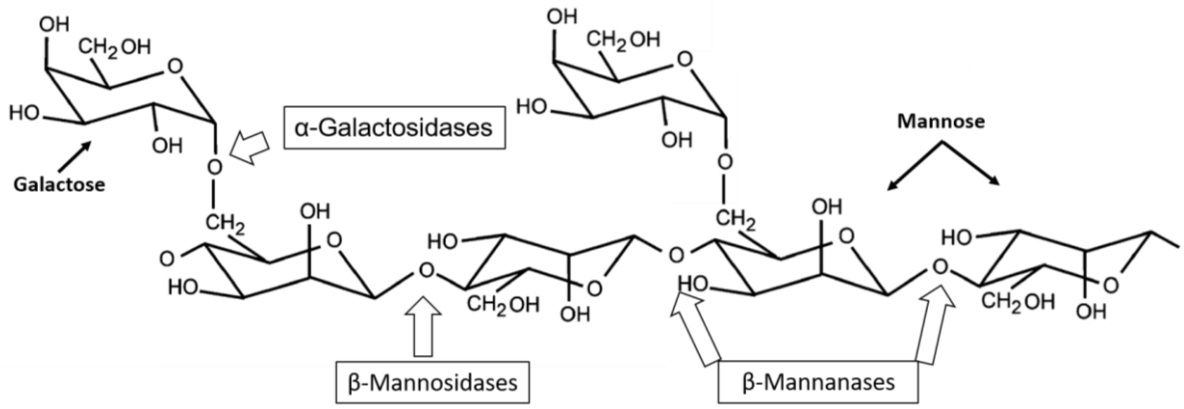


Figure 1.17. Enzymatic attack on galactomannan structure.

(Adapted from Mahammad *et al.* 2006; Moreira and Filho, 2008)

Finally, mixed-linked β-glucans are hydrolysed by endo-1,3-β-glucanases (laminarinases; EC. 3.2.1.39) from GH families 5, 16, 17, 55, 64, 81, 128, 152, 157 and 158; endo-1,4-β-glucanases (cellulases; EC 3.2.1.4) found in several GH families such as 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 74, 124 and 148; and lastly, endo-1,3-1,4-β-glucanases (licheninases; EC. 3.2.1.73) from GH families 5, 6, 7, 8, 9, 11, 12, 16, 17, 26 and 51 (www.cazy.org/).

1.2.4.2.2.1. Acessory enzymes

In addition to the enzymes described above, additional enzymes are required to cleave the linkage to side chains, to remove modifications (such as methyl esters and acetylation), or to split linkages to lignin. Among the accessory enzymes, feruloyl esterases and glucuronoyl esterases play a key role in enhancing the accessibility of enzymes and subsequent hydrolysis of hemicellulose fibers by removing the ester linkages between hemicellulosic carbohydrates and lignin(Figure 1.18) (Ishii 1997; Saulnier and Thibault 1999; Wong 2006).

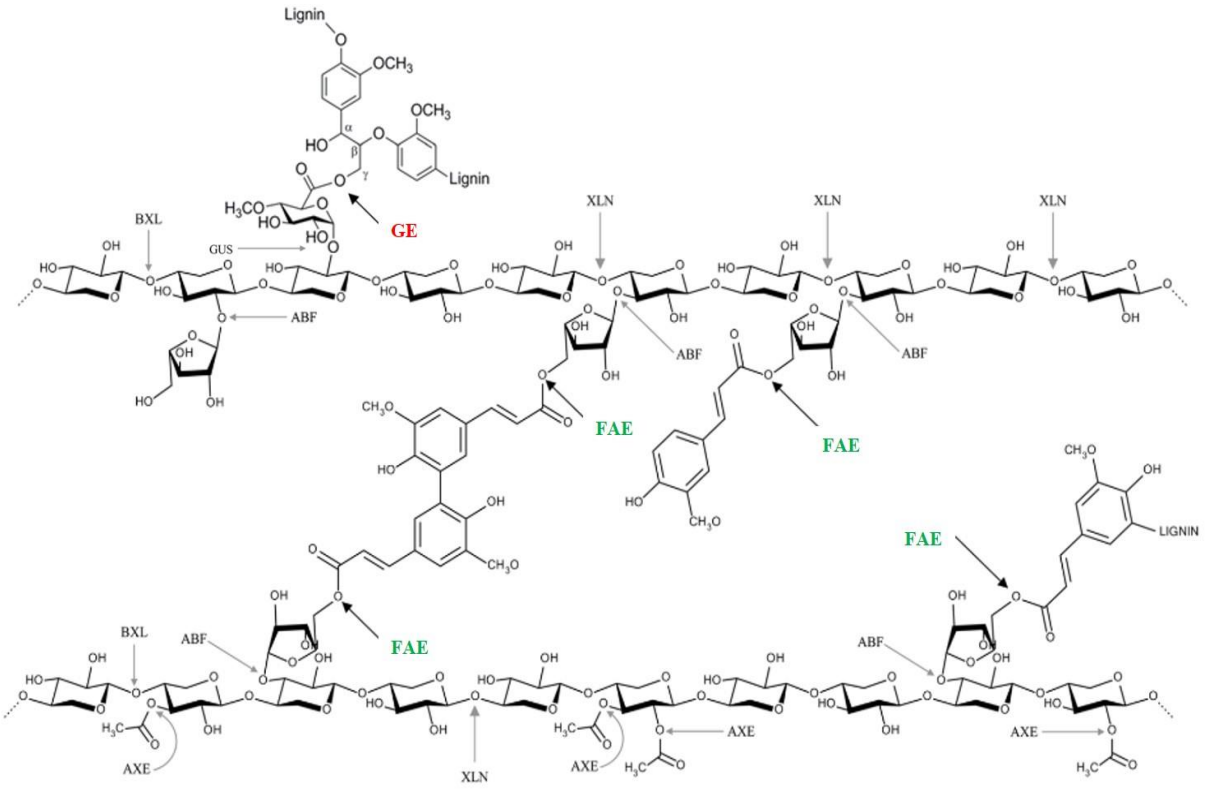


Figure 1.18. Representative ester linkages between xylans and lignin.

(ABF, α-arabinofuranosidase; AXE, acetyl xylan esterase; BXL, 1,4-β-xylosidase; FAE, feruloyl esterase; GE, glucuronoyl esterases; GUS, α-glucuronidase; XLN, 1,4-β-endoxylanase)

(Adapted from d'Errico *et al.* 2016; Dilokpimol *et al.* 2016, 2018)

a) Feruloyl esterases (FAEs)

Feruloyl esterases (FAEs; EC 3.1.1.73), also known as ferulic acid esterases, cinnamoyl esterases and cinnamic acid esterases, are able to catalyse the hydrolysis of the ester bond between hydroxycinnamic acids, such as ferulic acid (FA), and sugars present in plant cell wall (Williamson *et al*. 1998; Topakas *et al*. 2007; Faulds 2010). In other words, in xylan, FA is linked to O5 of terminal arabinose residues attached to the xylan backbone (Figure 1.18) (Wende and Fry 1997), and, in pectic hairy regions, it is attached to O2 of terminal arabinose residues or O6 of terminal galactose residues (Colquhoun *et al*. 1994). Thus, FAEs act synergistically with other carbohydrate-degrading enzymes, such as xylanases, in gaining access to their site of action during biomass conversion (Kroon *et al*. 1999; Wong 2006; Faulds 2010). These enzymes belong to the family 1 Carbohydrate Esterase (CE1) in the CAZy database (www.cazy.org).

b) Glucuronoyl esterases (GEs)

Glucuronoyl esterases (GEs; EC 3.1.1.-) are accessory enzymes that have the ability to catalyse the hydrolysis of ester bonds formed between hydroxyl groups of lignin alcohols and 4-O-methyl-D-glucuronic acid (MeGlcA) residues of xylans (Figure 1.18) (Watanabe and Koshijima 1988; Balakshin *et al*. 2011; Bååth *et al*. 2016). These enzymes are designated as family 15 Carbohydrate Esterase (CE15) in the CAZy database (www.cazy.org).

1.2.4.2.3. Pectinases

Pectic enzymes or pectinases are a heterogeneous group of related enzymes that hydrolyse pectic polysaccharides present mostly in plants. Pectin is an extremely complex polysaccharide, composed of as many as 17 different monosaccharides and more than 20 different linkages (Xia and Li 2018). Recently, Bonnin, Garnier and Ralet (2014) suggested a new classification of pectin degrading enzymes depending on their action sites in the pectic polymer. Therefore, the pectinases include homogalacturonan (HG) and rhamnogalacturonan (RG)-degrading enzymes (Figure 1.19).

HG-degrading enzymes consist of esterases, polygalacturonases, and lyases. Esterases include pectin methylesterases (EC 3.1.1.11; CE8), that catalyse the release of methanol from methyl-esterified Gal*p*A, and pectin acetylesterases (EC 3.1.1.1-; CE12), which remove acetylesters from pectin. Polygalacturonases include endo-polygalacturonases (EC 3.2.1.15; GH28) and exo-polygalacturonases (EC 3.2.1.67; GH4 and 28), both catalyse the hydrolysis of the α-(1–4) glycosidic bond between two adjacent D-Gal*p*A units. Xylogalacturonan hydrolase (EC 3.2.1.-; GH28) acts on xylogalacturonan by cleaving 1,4-α-D-galacturonan linkages between two xylosilated Gal*p*A units only to predominantly release the disaccharide Xyl-Gal*p*A. Pectin and pectate lyases split the glycosidic linkage between

two Gal*p*A units by catalysing a β-elimination reaction, thus introducing a double bond on the newly formed nonreducing Gal*p*A end. Pectin lyases (EC 4.2.2.10; PL1) and pectate lyases (EC 4.2.2.2; PL1,2,3,9 and 10) act on methylated and non-methylated substrates, respectively (Xia and Li 2018).

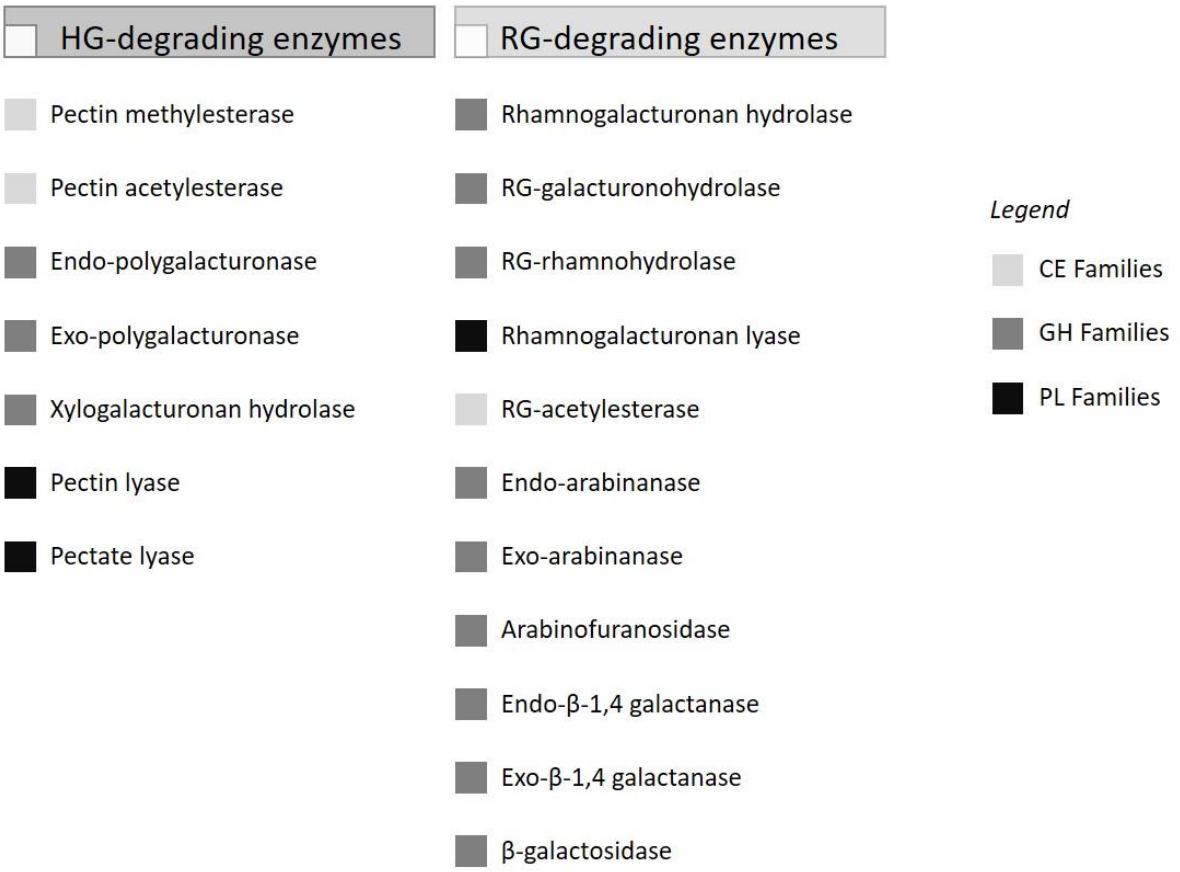


Figure 1.19. Summary of HG and RG-degrading enzymes.

RG-degrading enzymes include endo- and exo-enzymes. Exo-acting enzymes include RG hydrolase (EC 3.2.1.171; GH28), RG-galacturonohydrolase (EC 3.2.1.173; GH28), and RG-rhamnohydrolase (EC 3.2.1.174; GH78 and 106). Very few RG-lyases (EC 4.2.2.23; PL4,9 and 11) are reported, which ensure the endo-type eliminative cleavage of L-α-Rha*p*-(1–4)-α-D-Gal*p*A bonds of RGI domains. RG acetylesterases (EC 3.1.1.86; CE12) required for the deacetylation of the RG-I backbone were firstly described in *Aspergillus* species (Xia and Li 2018).

In plant species that the RGI backbone is decorated with galactans and/or arabinans (Luis *et al.* 2018) more enzymes are required for degradation of galactan and arabinan. The complete degradation of galactan requires a comprehensive set of different enzymes. Endo-1,4-β galactanases (EC 3.2.1.89; GH53 and 147) act randomly on the galactan core of AGI.

Exo-1,4-β-galactanases (EC 3.2.1.-; GH35) release galactose or galactobiose from the nonreducing end of 1,4-β-galactan. β-Galactosidases (EC 3.2.1.23; GH1, 23, 35, 39, 42, 50, 59 and 147) release the terminal nonreducing galactose from various substrates of low molar mass (Xia and Li 2018). In conclusion, several enzymes are required to degrade pectin and these enzymes belong to endo/exo-hydrolases type, lyases, and esterases. Alternatively, degradation of arabinan involves different enzymes that differ by their recognition sites in the polymer: endo-arabinanases, exo-arabinanases, and arabinofuranosidases. Endo-arabinanases (EC 3.2.1.99; GH43) randomly cleave the 1,5-α-linkages in the internal region of the arabinan backbone. Exo-arabinanases (EC 3.2.1.-; GH93) hydrolyse arabinan from the nonreducing end to release arabinobiose. Arabinofuranosidases (EC 3.2.1.55) act on α-L-arabinofuranosides, α-L-arabinans containing (1,3)- and/or (1,5)-linkages, arabinoxylans, and arabinogalactans. These enzymes are found in several GH families, including: 2, 3, 43, 51, 54, 62 and 155 (Xia and Li 2018).

1.2.4.2.4. Lignin degradation

Lignin is a complex substrate that is known to require a suite of oxidative enzymes and assorted small molecule co-factors for its complete degradation (Brown and Chang 2014). In contrast to hemicellulose and cellulose, lignin degradation results from an enzyme-mediated oxidative process, involving the initial transfer of single electrons to the intact lignin. Electrons are transferred to other parts of the molecule in uncontrolled chain reactions, leading to breakdown of the polymer. It is different from the carbohydrate hydrolysis because it requires oxidative reactions, based on an oxidizing power (e.g., hydrogen peroxide, H_2O_2), to completely break down the macromolecule. In general, it is a significantly slower reaction than the hydrolysis of structural carbohydrate (Clifford 2018).

Lignin degradation is driven by several fungi as well as several bacterial species (Breen and Singleton 1999; Singh Arora and Kumar Sharma 2010; Bugg *et al*. 2011; Huang *et al*. 2013). However, the most effective lignin-degrading microbes in nature are thought to be white rot fungi (D'Souza *et al*. 1999). In general, enzymes responsible for lignin degradation are known as ligninases (Butler and Day 1998). The most common fungal ligninases are copper-containing laccases (Lac; EC 1.10.3.2; AA1) and heme peroxidases, that include lignin peroxidases (LiP; EC 1.11.1.14; AA2), manganese peroxidases (MnP; EC 1.11.1.13; AA2), versatile peroxidases (VP; EC 1.11.1.16; AA2) and dye-decolorizing peroxidases (DyPs; EC 1.11.1.19) (Sugano 2009; Abdel-Hamid *et al*. 2013; Vrsanska *et al*. 2015).

Lignin peroxidases catalyse the oxidative depolymerization of lignin with H_2O_2 acting as the oxidizing agent. Manganese peroxidases use H_2O_2 to oxidize Mn^{2+}, which is present in wood and soils, thus generating reactive Mn^{3+} ions. Versatile peroxidases combine the

properties of lignin peroxidases and manganese peroxidases, conferring the catalytic versatility inferred by their name. They can oxidize Mn^{2+} to Mn^{3+} like manganese peroxidases, but can also oxidize non-phenolic compounds in the same manner as lignin peroxidases (Ruiz-Dueiias *et al*. 2001). In contrast to the other three classes, DyP-type peroxidases are not members of the classical plant/microbial peroxidase superfamily, due to differences in sequence, structure and function (Liers *et al*. 2010; Colpa *et al*. 2014). DyP-type peroxidases show considerable potential for the degradation of lignin because they oxidize dyes, non-phenolic lignin model compounds (including veratryl alcohol and adlerol) and non-phenolic lignin model compounds containing β-O-4 linkages (Kim and Shoda 1999; Liers *et al*. 2010).

Finally, laccases catalyse the oxidation of polyphenols and methoxy-substituted phenols by generating free radicals (Claus 2004). The oxidation is typically coupled with the reduction of molecular oxygen to water (Thurston 1994). In the presence of redox mediators, laccases can even catalyse the breakdown of non-phenolic lignin structures, including the cleavage of β-O-4 linkages (Bourbonnais and Paice 1990; Kawai *et al*. 2002). In addition to the peroxidases and laccases, fungi produce other accessory oxidases such as aryl-alcohol oxidase (AAO; EC 1.1.3.7; AA3), also called veratryl alcohol oxidase, glyoxal oxidase (GLOX; EC 1.2.3.15; AA5) and various carbohydrate oxidases (Gonzalo *et al*. 2016) that generate the hydrogen peroxide required by the peroxidases (Sigoillot *et al*. 2012; Abdel-Hamid *et al*. 2013).

1.2.4.3. Cellulosomes

The cellulosome is a highly efficient, highly organized, cell surface enzymatic system (Bayer *et al*. 2004) produced by anaerobic microorganisms to efficiently deconstruct complex plant cell wall polysaccharides, such as cellulose and hemicellulose (Bule *et al.* 2018). The cellulosome was first identified in the anaerobic thermophilic cellulolytic bacterium *Clostridium thermocellum* (Bayer *et al*. 1983; Lamed *et al*. 1983). It was described as an highly elaborate multi-enzyme complex which was later found in other cellulolytic bacteria (Lamed *et al*. 1987; Bayer *et al*. 2008), including *Clostridium cellulovorans* (Doi and Tamaru 2001; Doi *et al*. 2003), *Clostridium cellulolyticus* (Bélaich *et al*. 1997), *Clostridium josui* (Kakiuchi *et al.* 1998), *Clostridium acetobutylicum* (Sabathé *et al*. 2002), *Clostridium papyrosolvens* (Pohlschroder *et al*. 1995), *Clostridium clariflavum* (Artzi *et al*. 2014; Artzi *et al*. 2015), *Acetivibrio cellulolyticus* (Ding *et al*. 1999), *Bacteroides cellulosolvens* (Ding *et al*. 2000), *Ruminococcus albus* (Ohara *et al*. 2000), and *Ruminococcus flavefaciens* (Kirby *et al*. 1997; Ding *et al*. 2001).

1.2.4.3.1. Cellulosome composition and structure

Cellulosomes include a large structural subunit, that contains tandemly repeated cohesin modules, usually in multiple copies, that bind the dockerin modules located in cellulosomal enzymes, generally termed scaffoldin (Figure 1.20) (Bayer *et al*. 2004). Thus, cellulosome assembly depends on the specific protein-protein interactions established between cohesin and dockerin modules, which bind tightly to each other (Artzi *et al.* 2016). The specificity displayed by the cohesin–dockerin interaction defines the intimacy of the cellulosome structure and the mechanism of cell surface attachment (Bayer *et al*. 1994; Doi and Kosugi 2004). The scaffoldin also frequently includes different types of domains, such as CBMs and S-layer homology modules (SLHs). The CBM is the domain through which the complex usually recognizes and binds to the cellulosic substrate (Bayer *et al*. 2004). SLH motifs are found at the N-terminus of many S-layer proteins (the external constituent of many bacteria cell walls) and at the C-terminus of some exo-proteins of Gram-positive bacteria (Sára and Sleytr 2000). Three types of SLH domains have been identified according to their origin in S-layer proteins, exo-proteins, and porins (Engelhardt and Peters 1998). Cellulosomes can be attached to the bacterial cell surface, through the SLH domain, or can be released as cell-free cellulosomes (Hamberg *et al.* 2014; Xu *et al.* 2016).

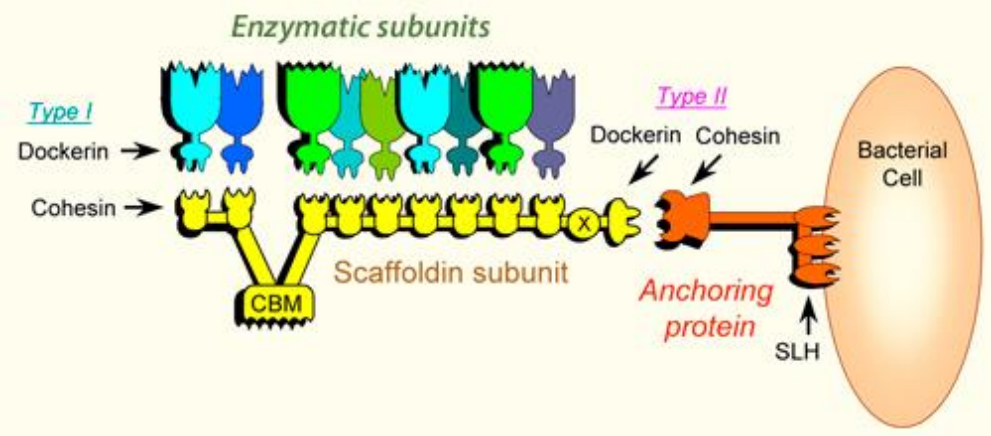


Figure 1.20. Schematic representation of cellulosome.

a) Scaffoldins

The cellulosome system is characterized by three major types of scaffoldins, referred as primary, anchoring and adaptor scaffoldins (Figure 1.21). Cellulosome assembly primarily relies on the function of primary scaffoldins (Zverlov *et al*. 2008), which incorporates the enzymatic subunits and usually bears a single cellulose-binding CBM (Bayer *et al*. 2004). Anchoring scaffoldins interact with the cell surface through specialized anchoring modules, either non-covalently through SLH domains or covalently through sortase motifs (Lemaire *et al*. 1995; Rincon *et al*. 2005), and bind to primary scaffoldins through specialized cohesin-

dockerin interactions. Finally, more complex and intricate cellulosomes contain adaptor scaffoldins that increase the repertoire or number of components into the complex by either connecting two scaffoldins or a scaffoldin and an enzyme (Figure 1.22) (Artzi *et al.* 2016).

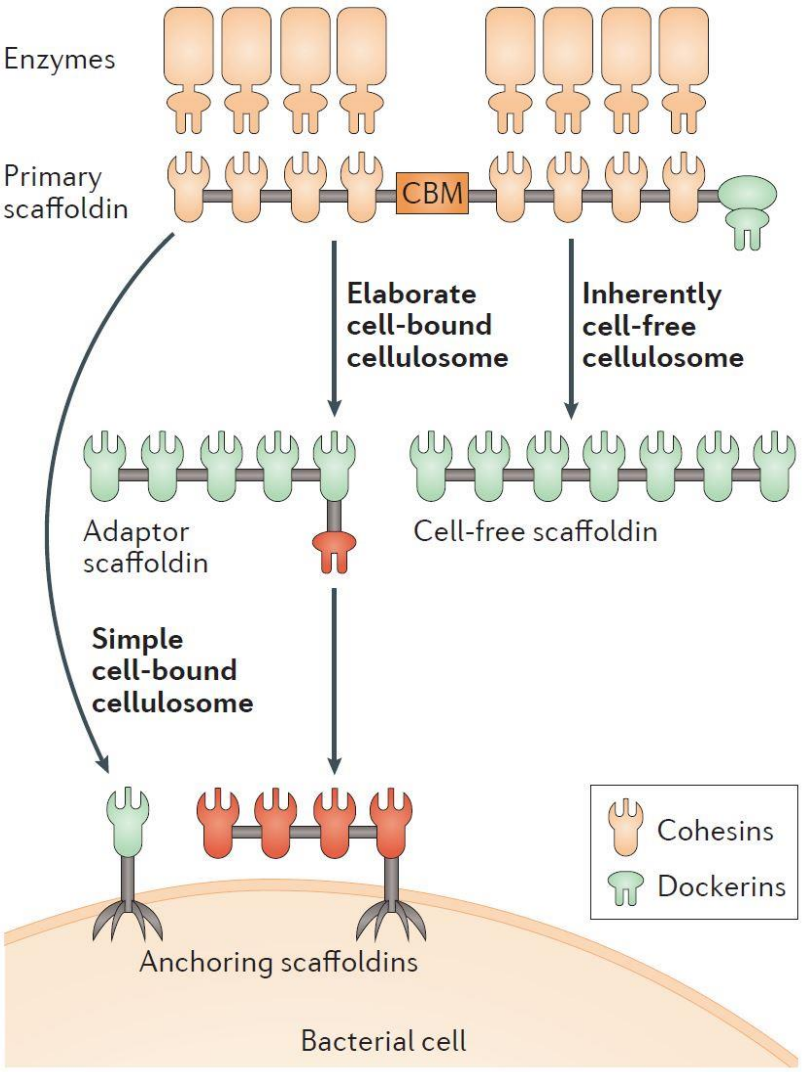


Figure 1.21. Types of cellulosome systems.

A primary scaffoldin can be bound directly to the cell surface through an anchoring scaffoldin or, in a more elaborate system, through an intermediary adaptor scaffoldin. Cellulosomes can also exist in an inherently cell-free state through attachment to an appropriate free scaffoldin (Artzi *et al.* 2016).

b) Cohesin-dockerin interaction

Cellulosome assembly is mediated by the highest affinity protein:protein interaction (>10^9 M^{-1}) between dockerin and cohesin modules (Francis Schaeffer *et al*. 2002; Miras *et al*. 2002). This non-covalent interaction is one of the strongest known in nature (Valbuena *et al*. 2009; Stahl *et al*. 2012; Gunnoo *et al*. 2016) and is very difficult to dissociate (Bhat and Wood 1992). Cohesin modules are the major building blocks of scaffoldins which are

responsible for organizing the cellulolytic subunits into the multi-enzyme complex. Dockerin modules located in the enzymes anchor the catalytic subunits into the scaffoldin by establishing specific interactions with cohesins. It is now well recognized that the intrinsic duplication in dockerin primary sequences confers an internal two-fold symmetry to this protein domains consisting of a duplicated F-hand motif (a calcium-binding loop preceding an helix) (Bayer 2019). Both cohesins and dockerins are highly homologous within the same species and the residues directly involved in protein:protein recognition are conserved within the same species (Fontes and Gilbert 2010) while varying among different species (Bayer 2019). The composition of different individual cellulosome molecules remains heterogeneous and usually reflect differences in the primary carbon sources the organism is using (Ciruela *et al*. 1998).

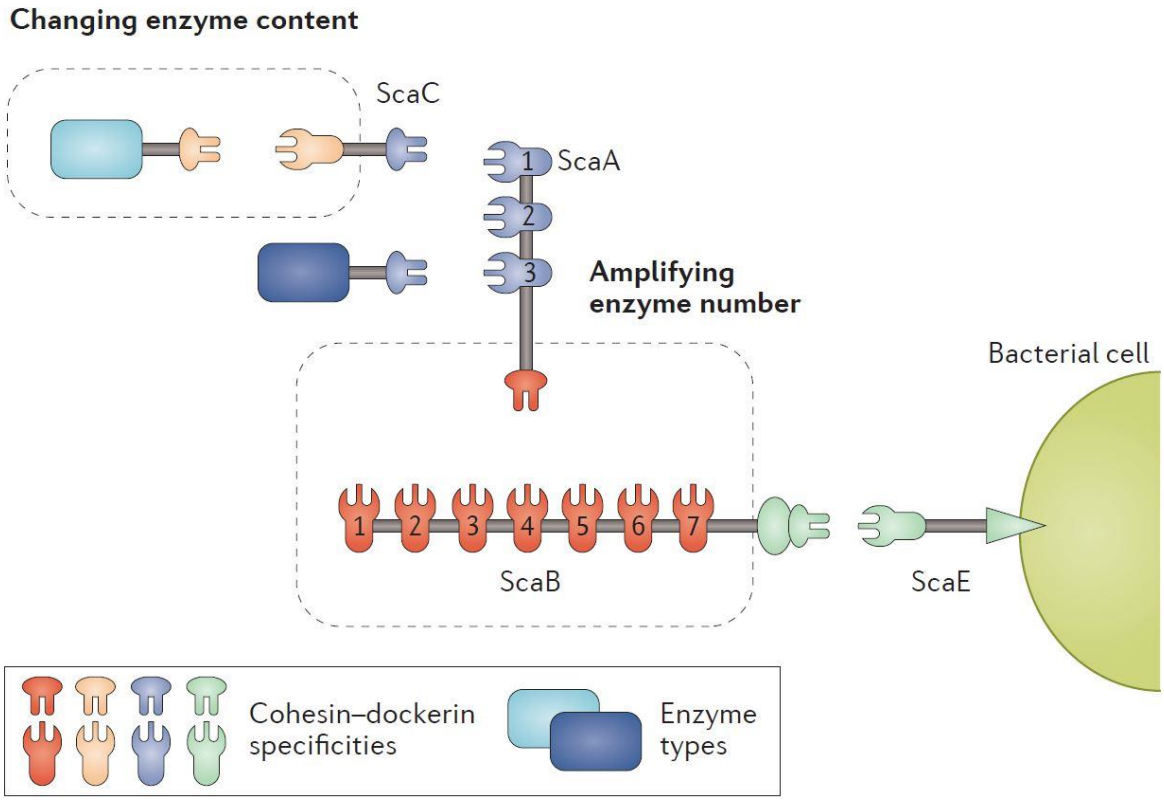


Figure 1.22. The cellulosome of *Ruminococcus flavefaciens* with alternative roles of adaptor scaffoldins.

(Artzi *et al.* 2016)

Three types of cohesin–dockerin interactions have been described and are based on the primary sequences of each cohesin–dockerin pair. In general, Type I interactions occur between dockerin-containing catalytic subunits and cohesins of the primary scaffoldin. Type II interactions occur between a dockerin located in the primary scaffoldin and cohesins of

anchoring scaffoldins (Figure 1.20), although there are exceptions (Artzi *et al.* 2015). Later, Type III interactions, which are completely unrelated to the type I and type II interactions initially described, were discovered and are characteristic of ruminal microbes in particular *R. flavefaciens* (Fontes and Gilbert 2010; Artzi *et al*. 2016). While the type-I dockerin modules exhibit a symmetrical dual binding mode to cohesins, it was recently discovered that type-III dockerins bind to cohesins in a different mode, possibly in a non-symmetrical manner and with both segments contributing to complex formation on an equal basis (Bayer 2019).

Type-I and II cohesins exhibits a jellyroll topology that folds into a nine-stranded β-sandwich (Noach *et al.* 2005). Type-III cohesin has the core nine-stranded jellyroll cohesin topology with two type-II like β-flaps but also displays a unique N-terminal loop and dominant α-helix region (Alber *et al.* 2009; Karpol *et al.* 2013; Salama-Alber *et al.* 2013; Smith and Bayer 2013). In contrast, the two 22-residues duplicated sequences characteristic of type-I dockerins, which are separated by a linker of 9–16 residues, fold into two well-conserved 12-residue loops which bind to calcium ions and two α-helices. The structure is a variation and subtype of the so-called "EF-hand" helix-loop helix motif. Type-II dockerin is made from two loop-helix motifs, termed F-hand motifs, separated by a 14-residue linker region. Like in type-I dockerin, Ca^{2+} ions are bound to a well-conserved 12-residue loop (Jarrett J. Adams *et al*. 2006). The structure of Type-III dockerins diverges from the other two. It exhibits two F-hand motifs but the second motif misses the 12-residue sequence found in the Ca^{2+}-binding loop (Karpol *et al.* 2013). It exhibits a total of five α-helices, unlike the other types which contain only three helices. Furthermore, the linker between the repeats is much longer as are the helices (Salama-Alber *et al.* 2013).

c) Enzymes

Cellulosomes combine an extensive and diverse repertoire of enzymes, including glycoside hydrolases, pectate lyases and carbohydrate esterases (for more details see chapter 1.2.4.1.CAZymes). Both, free and cellulosomal enzymes, contain very similar types of catalytic domains.

d) Carbohydrate-Binding Modules

The scaffoldin CBM plays three main roles in adhesion of cellulosomes to cellulose substrates: a proximity effects, substrate targeting and, eventually, microcrystalline cellulose disruption (Reinikainen *et al*. 1992; Din *et al*. 1994; Henshaw *et al*. 2004) (for more details see chapter 1.2.4.1.6.CBMs). Thus, CBMs keep the enzymes embedded within the scaffoldin in close proximity to the cellulose surface, promoting enzyme-cellulose associations and an increased enzyme local concentration. The known scaffoldins contain CBMs that belong only

to two families (family 2 and 3) (Poole *et al*. 1992; Gerngross *et al*. 1993; Morag *et al*. 1995; Dassa *et al*. 2012)

1.2.4.3.2. Cellulosome systems

Bacterial cellulosomal systems can be categorized into two major types: simple cellulosomes (Figure 1.23) that contain a single scaffoldin (exemplified by that of *C. cellulovorans*) and complex cellulosomes (Figure 1.24) that exhibit multiple types of interacting scaffoldins (such as in *C. thermocellum*) and, consequently, contain more enzymes in a single assembly than simple cellulosomes (Bayer *et al*. 2004; Artzi *et al*. 2016).

Simple cellulosome systems use a single primary scaffoldin which contain a single CBM, one or more X2 modules and numerous cohesins (5 to 9) (Figure 1.23). These scaffoldins incorporate the dockerin-bearing enzymes into the complex and, in several cases, have been shown to be associated with the cell surface, although the mechanism underlying this interaction remains uncovered. The simple cellulosomal system has been reported in *C. cellulolyticum, C. cellulovorans, and C. papyrosolvens* (Figure 1.23) (Bayer *et al*. 2004; Doi and Kosugi 2004; Desvaux 2005)

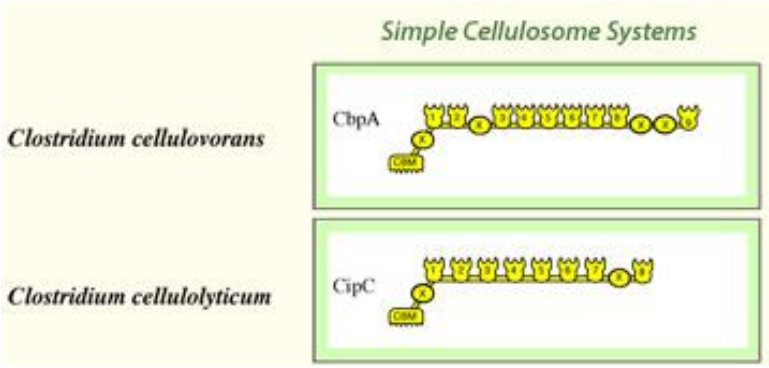


Figure 1.23. Schematic representation of the scaffoldins of simple cellulosome systems.
(Adapted from Bayer 2019)

In complex cellulosomal systems (Figure 1.24), more than one scaffoldin interlocks with other scaffoldins through different mechanisms to produce highly dynamic and diverse cellulosome architectures. At least one type of scaffoldin serves as primary scaffoldin that incorporates the enzymes directly into the cellulosome complex. In each species, another type of scaffoldin attaches the cellulosome complex to the cell surface via a specialized module or sequence, designed for this purpose. The complex cellulosome systems has been reported in *C. thermocellum* (Figure 1.24)*, A. cellulolyticus, C. clariflavum*, and other species (Bayer *et al*. 1998; Xu *et al*. 2003; Bayer *et al*. 2004; Xu *et al*. 2004; Artzi *et al*. 2014).

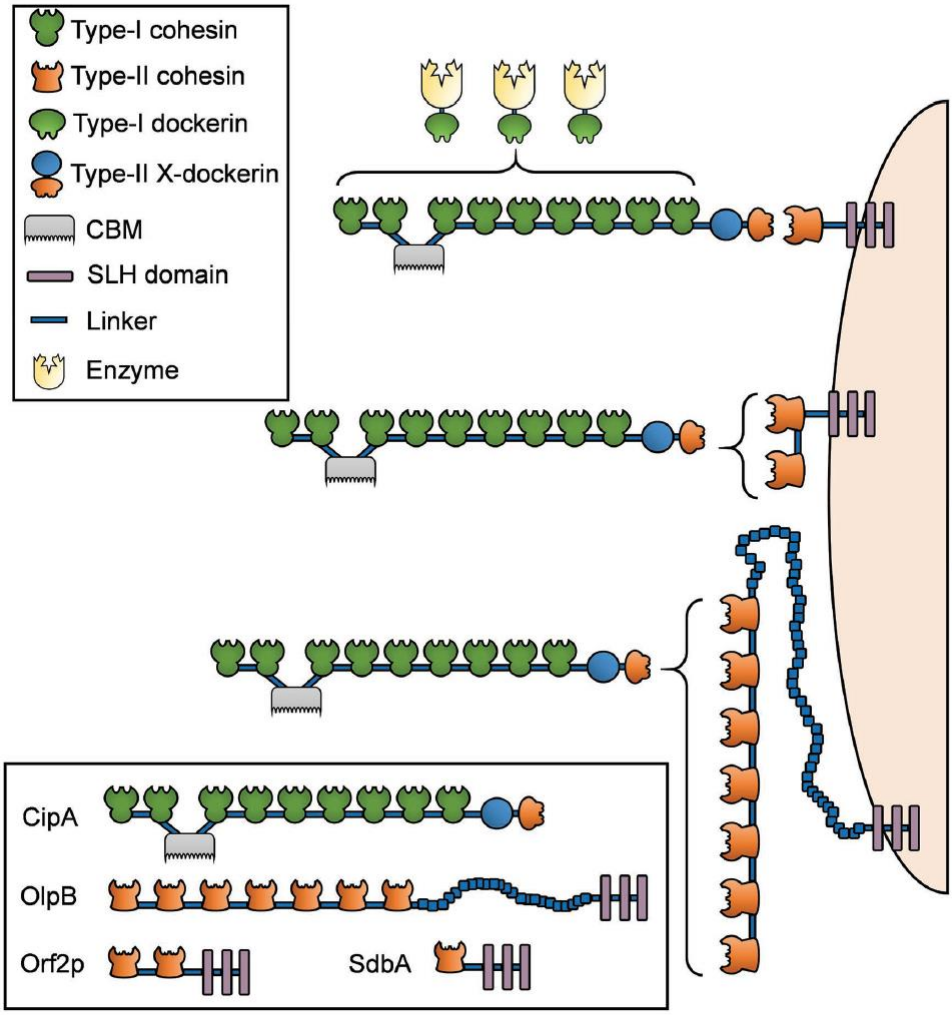


Figure 1.24. The cellulosome architecture of *C. thermocellum*.

(Gunnoo *et al.* 2016)

1.3. Supplementation of diets for poultry

1.3.1. Non-starch polysaccharides in poultry nutrition

Non-starch polysaccharides (NSPs) comprehend a large variety of structural carbohydrates, comprising some of the most representative compounds of the cell wall (Williams *et al*. 1997). NSPs can be divided into insoluble, such as cellulose, and the partly soluble ones, such as arabinoxylans, β-glucans and pectins (Figure 1.25) (Hetland *et al*. 2004). Monogastric animals, such pigs and poultry, do not express the enzymes to hydrolyse these carbohydrates in the upper portions of the gastrointestinal tract, and thus their digestion occurs by means of bacterial fermentation in the lower digestive compartments (Choct *et al*. 2010). Besides that, the concentrations of these components vary between different plants, plant parts and stages of maturity (Hetland *et al.* 2004).

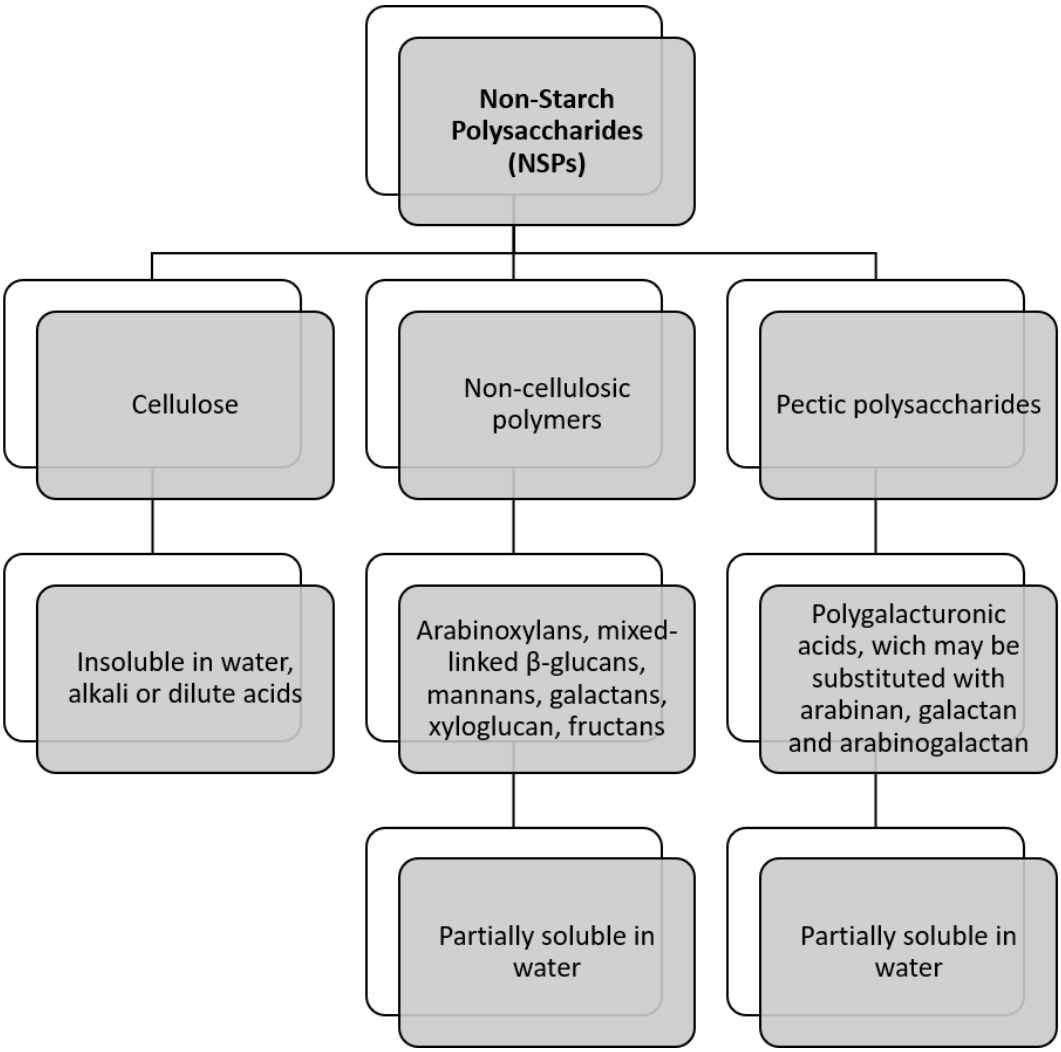


Figure 1.25. Classification of Non-Starch Polysaccharides (NSPs).

(Adapted from Choct *et al.* 2010)

Insoluble NSP affects gut functions and modulates nutrient digestion (Hetland *et al.* 2004). Thus, digestibility of starch is higher and digesta passage rate faster when a moderate level of insoluble NSP is present in the diet. The effect of insoluble fiber on gut functions stems from its ability to accumulate in the gizzard, which seems to regulate digesta passage rate and nutrient digestion in the intestine (Hetland *et al.* 2004). Thus, several studies (Hetland and Svihus 2001; Hetland *et al*. 2002) suggest that animal performance does not decrease when insoluble fiber is included in moderate levels to broilers or layers despite the clear reduction in nutrient concentration in the diet.

In contrast, soluble NSP lead to a higher digesta viscosity in the small intestine, thereby inhibit digestion and absorption (Smits and Annison 1996). Increased viscosity leads to changes in the physiology and the ecosystem of the gut (Angkanaporn *et al*. 1994). This is probably related to a slower digesta passage rate. A slow moving digesta with low oxygen tension in the small intestine could provide a relatively stable environment where

fermentative microflora can establish (Wagner and Thomas 1978). Choct *et al.* (1996) demonstrated a large increase in fermentation in the small intestine of broilers by adding soluble NSP to the diet. At first, it could be thought that increased production of volatile fatty acids (VFA) would increase the energy content of the feed, but due to the drastic change in the gut ecosystem, the net effect was a decreased nutrient digestion accompanied by poor bird performance (Choct *et al.* 2010). The water-holding capacity of soluble NSP also causes problems leading to sticky droppings (Hetland *et al.* 2004).

The digestibility of NSPs is affected by a multitude of factors which include animal species, age of animals, solubility, chemical structure, and their quantity in the diet (Choct *et al.* 2010). In addition, the anti-nutritive effect of NSPs on nutrient digestion is related with their inclusion levels, sources and composition (Murray *et al*. 1977; Freire *et al*. 2000) and with the capacity of the gut microflora of the chicken to digest large amounts of NSPs.

1.3.2. Enzymatic supplementation of cereal-based diets for poultry

In poultry, elevated levels of NSPs, in particular the soluble fraction, lead to decreased nutrient digestion and lower nutrient absorption (Antoniou *et al*. 1981; Choct and Annison 1990). As described above, soluble NSPs increase digesta viscosity, reducing digesta passage rate and affecting the interaction of the endogenous digestive enzymes with their target substrates, consequently affecting feed intake and digestibility, respectively (Pettersson and Åman 1989; Smits and Annison 1996; Józefiak *et al*. 2006). The prolonged digesta passage rate also promotes a modification in gut physiology such as an enlargement of the gastrointestinal organs (Smits and Annison 1996). Finally, high digesta viscosity favors the proliferation of anaerobic microbes in the upper parts of the GI tract thus affecting animal's health (Józefiak *et al*. 2007). This microflora can bind some proteins and form complexes that limit protein hydrolysis (Vahouny *et al*. 1980). The normal and healthy microflora is composed of facultative anaerobic microorganisms in jejunum and strict anaerobic microorganisms in caecum.

When birds are fed wheat or barley-based diets, the presence of soluble NSPs (arabinoxylans and β-glucans) is considerable. Therefore, it is now well established that the dietary supplementation with exogenous enzymes (such cellulases and hemicellulases) reduces the degree of polymerization of NSPs, improving the nutritive value of diets and, consequently, increasing animal's performance. These enzymes can partially hydrolyse NSPs, reducing the viscosity of the gut contents and resulting in significant improvements in nutrient absorption. Several studies have also demonstrated that exogenous enzyme supplementation can affect the intestinal morphology in birds fed barley-based diets or decrease the small intestinal fermentation attributed to high NSP diets (Brenes *et al*. 1993;

Choct *et al*. 1996; Williams *et al*. 1997). In addition, polysaccharide disruption releases endogenous trapped nutrients increasing feed metabolizable energy (Williams *et al*. 1997).

On the other way, is also well established that temporal response to enzymatic supplementation may vary among the sources of raw-materials used to prepare broilers diets (Cardoso *et al.* 2014). For example, broilers fed barley-based diets display an improved performance in response to β-glucanase supplementation in particularly at the early stages of their life (Newman and Newman 1988; Rotter *et al*. 1989; Nahas and Lefrancois 2001), when the young chicks have a poorly developed digestive system. Data presented by Cardoso *et al.* (2014) indicate that enzyme supplementation can be restricted to the first 11 days of the production cycle. Therefore, a later exposure to the exogenous microbial enzymes does not seem to be important to improve the nutritive value of barley-based diets for broilers. In contrast, in wheat-based diets, exogenous enzymes seem to be more important at later phases of the productive cycle (Fontes *et al*. 2004). In general, viscosity of wheat-based diets is usually not as pronounced as in barley or oat, and thus exogenous enzymes are not critical to reduce digesta viscosity, but rather to generate novel substrates, in particular xylo-oligosaccharides, which are used by the beneficial bacteria colonizing the final portions of the GI tract (Jamroz *et al*. 2002; Fontes *et al*. 2004).

1.3.3. Prebiotic supplementation of diets for poultry

Prebiotics are defined as nondigestible food ingredients that can be utilized by intestinal microflora, which beneficially affect the host (Gibson and Roberfroid 1995). However, more recent definitions state that a prebiotic is a selectively fermented ingredient that allows specific changes, in both the composition and activity of the gastrointestinal microbiota, which confers benefits to the host (FAO/WHO 2002). In this way, prebiotics alter the intestinal microbes and immune system to reduce colonization by the pathogens in certain conditions (Hajati and Rezaei 2010).

For a dietary substrate to be classed as a prebiotic, at least three criteria are required: (1) the substrate must not be hydrolysed or absorbed in the stomach or small intestine, (2) it must be selective for beneficial commensal bacteria in the large intestine such as the bifidobacteria, (3) fermentation of the substrate should induce beneficial luminal/systemic effects within the host (Scantlebury-Manning and Gibson 2004). Generally, prebiotics can be fermented by health-promoting bacteria in the intestine, producing lactic acid, short-chain fatty acid (SCFA), or some antibacterial substances, such as bacteriocin against pathogenic species (Bogusławska-Tryk *et al*. 2012). These products may not only benefit the intestinal microbial structure but also improve the integrity of intestinal epithelial cells, which further increase the absorption of nutrients and enhance the growth performance of animals (Lan *et al*. 2005).

Some oligosaccharides, carbohydrates composed of short chains of monosaccharides, enhance the growth of beneficial organisms in the gut and others are thought to function by competing with the pathogenic bacteria for attachment sites in the lumen. Thus, prebiotic oligosaccharides may improve animal health (Hajati *et al*. 2014). The prebiotic approach has not a long history of use in broilers. However, application studies have been increasing in the last years to assess their effect on gut health, performance, and reduction of pathogen shedding (Yang *et al*. 2009). The most commonly oligosaccharides with prebiotic characteristics in poultry production are fructo-oligosaccharides (FOS) (Kim *et al*. 2011; Światkiewicz *et al*. 2011), mannan-oligosaccharides (MOS) (Baurhoo *et al*. 2009; Xiao *et al*. 2012), xylo-oligosaccharides (XOS) (Courtin *et al.* 2008; Zhenping *et al.* 2013), galacto-oligosaccharides (GOS) (Faber *et al.* 2012) and soybean meal oligosaccharides (SOS) (Lan *et al*. 2007).

Application of prebiotics in diets could establish a healthy microbial community in the intestine of young broilers by enhancing the abundance of *Lactobacilli* and *Bifidobacteria* and reducing the titers of *Coliform* (Yang *et al*. 2008; Chee *et al*. 2010). Furthermore, the modulation of intestinal microbiota is associated with immune responses. First, inhibiting pathogen colonization by prebiotics can decrease detrimental molecules produced by pathogenic bacteria and, second, prebiotics can act as non-pathogenic antigens themselves. It is believed that prebiotics may be recognized by receptors of immune cells, which consequently modulate host immunity beneficially (Teng and Kim 2018).

1.3.3.1. Xylo-oligosaccharides (XOS)

Xylo-oligosaccharides (XOS) and arabinose-substituted xylo-oligosaccharides (arabinoxylan-oligosaccharides) (AXOS) are produced by the hydrolytic degradation of arabinoxylans, by endo-1,4-β-xylanases that cleave the internal β-xylosidic glycosidic linkage bonds of the linear polyxylose chains (Jommuengbout *et al*. 2009; Broekaert *et al*. 2011; De Maesschalck *et al*. 2015). Therefore, XOS are putative prebiotics and consist of a mixture of oligosaccharides formed by xylose residues linked through (1–4)-β-linkages. The number of xylose residues involved in their formation can vary between 2 and 10 (Aachary and Prapulla 2011; Morgan *et al*. 2017). The structure of XOS vary by the degree of polymerization (numbers of xylose residues in their backbone), degree of substitutions (ratio of arabinose to xylose), monomeric units and types of linkages. XOS generally consist of mixtures of oligosaccharides formed by xylose residues linked through 1,4-β-linkages. Some studies have suggested the beneficial effects of XOS on broiler performance when added to the feed (De Maesschalck *et al.* 2015). It was also proposed that XOS would improve humoral immunity in poultry (Zhenping *et al.* 2013).

1.4. Objectives

The work presented in this thesis aims to elucidate several unresolved questions concerning the use of high-throughput techniques for CAZymes characterization and the applicability of CAZymes to improve the nutritive value of cereal-based diets for animals. The specific aims of this study are as follows:

- To evaluate the efficacy of 1,3-1,4-β-glucanases and 1,4-β-glucanases to improve the nutritive value of β-glucan-containing diets for monogastric animals (Chapter 2);
- To investigate endogenous factors, specifically the levels of non-starch polysaccharides and xylanase activity, that might affect the nutritive value of wheat for poultry (Chapter 3);
- To study the effect of xylo-oligosaccharides generated by the *in viv*o activity of microbial xylanases to improve the performance of broilers fed wheat or corn-based diets (Chapter 4);
- To develop a large library of Carbohydrate-Active Enzymes for plant cell wall deconstruction (Chapter 5);
- To determine the structural determinants of substrate specificity in family 15 Carbohydrate Esterases (Chapter 6);
- To characterize the subtle changes at the catalytic site of feruloyl esterases that modulate the specificity to hydroxycinnamic acids (Chapter 7).

2. NOVEL EVIDENCE SUPPORTING THE ROLE OF 1,3-1,4-B-GLUCANASES AND NOT 1,4-B-GLUCANASES TO IMPROVE THE NUTRITIVE VALUE OF B-GLUCAN-CONTAINING DIETS FOR MONOGASTRIC ANIMALS

V. Cardoso[a,b], T. Ribeiro[a], J.L.A. Brás[a,b], V.O. Fernandes[a,b], H. Santos[a], M.M. Lordelo[c], L.M.A. Ferreira[a,b] and C.M.G.A. Fontes[†a,b]

[a]CIISA - Faculdade de Medicina Veterinária, Universidade de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; [b]NZYTech genes & enzymes, Campus do Lumiar, Estrada do Paço do Lumiar, 1649-038 Lisboa, Portugal; [c]Instituto Superior de Agronomia, Tapada da Ajuda, 1349-017 Lisboa, Portugal.

Adapted from: Cardoso *et al*. *Submitted for publication*

Abstract

Dissecting the role of exogenous feed enzymes in animal nutrition is urgently needed to optimize feed science technology. We hypothesized that 1,3-1,4-β-glucanases (EC 3.2.1.73) but not cellulases/1,4-β-glucanases (EC 3.2.1.4) are the required enzymes to improve the nutritive value of barley-based diets for simple stomach animals. Previously we observed that a family 8 cellulase, which highly efficiently cleaves anti-nutritive β-glucans *in vitro,* was unable to improve the nutritive value of a barley-based diet. Here we have extended these studies by analysing the capacity of the previously proven 1,3-1,4-β-glucanase 16A (*Ct*Glc16A) versus a highly general cellulase/1,4-β-glucanase 5E (*Ct*Cel5E), to improve the nutritive value of a barley-based diet for poultry. While *Ct*Glc16A effectively improved animal performance, cellulase *Ct*Cel5E was unable to affect the nutritive value of the cereal-based diet. Further analysis indicated that *Ct*Cel5E remained active during passage through the GI tract, suggesting that inefficiency of cellulase *Ct*Cel5E was predominantly due to its “trapping” by the abundant cellulosic substrates rather than resulting from enzyme inactivation. This novel evidence supports the role of 1,3-1,4-β-glucanases, but not cellulases/1,4-β-glucanases, to improve the nutritive value of β-glucan-containing diets. Although conceptually cellulases/1,4-β-glucanases could depolymerize β-glucans *in vivo,* the formation of unproductive interactions with cellulosic carbohydrates limits the capacity of the enzyme to attack the anti-nutritive β-glucans. The fact that distantly related cellulases from families GH8 and GH5 remain inefficient suggests that, in general, cellulases/1,4-β-glucanases (EC 3.2.1.4) will be ineffective to improve the nutritive value of β-glucan-containing diets for monogastric animals.

2.1. Introduction

Barley β-glucans are well-known soluble non-starch polysaccharides (NSPs), which upon solubilization, form a highly viscous solution in the animals´ gastrointestinal (GI) tract presenting significant anti-nutritive properties (Khoury *et al*. 2011). β-glucans consist of a backbone of glucose residues connected by intercalated 1,4-β- or 1,3-β-glycosidic bonds, generating a linear polysaccharide more soluble than cellulose (Xue *et al.* 2003). In barley β-glucans, 1,4-β linkages predominate with a ratio of 1,3-β to 1,4-β linkages of approximately of 1 to 2.5 (Xue *et al.* 2003; Jamar *et al.* 2011). In Nature, Glycoside Hydrolases (GH; EC. 3.2.1.-) that participate in the hydrolysis of barley 1,3-1,4-β-glucans may cleave 1,4-β-glycosidic bonds in the context of neighbouring 1,3-linkages or 1,4-β-glycosidic bonds, located in 1,4-β-invariable regions of the polysaccharide. The first enzymes, which target mixed linkages of the polysaccharide, are generally termed as 1,3-1,4-β-glucanases (EC 3.2.1.73), while the later are referred as cellulases or 1,4-β-glucanases (EC 3.2.1.4) and have evolved to degrade primarily cellulose. In addition, enzymes that participate in the endo-hydrolysis of 1,3 or 1,4 β-glucan linkages, when the glucose residue whose reducing group is involved in the linkage to be hydrolysed, is itself substituted at C-3, have been identified and are termed 1,3(1,4)-β-glucanases (EC 3.2.1.6). Based on primary sequence homology, GHs are organized in families (Henrissat 1991; Henrissat and Daviest 1997). Presently there are 162 families of GHs in the constantly updated Carbohydrate Active enZyme (CAZy) database (www.cazy.org, data collected on April 2019). Thus, 1,3-1,4-β-glucanases (EC 3.2.1.73), which display a strict specificity for 1,3-1,4-β-glucans, are currently grouped in GHs families 5, 6, 7, 8, 9, 11, 12, 16, 17, 26 and 51. The less diverse but 1,3-1,4-β-glucan specific 1,3(1,4)-β-glucanases (EC 3.2.1.6) belong to families 9 or 16. In contrast, the more generic cellulases/1,4-β-glucanases (EC 3.2.1.4), which also cleave 1,3-1,4-β-glucans in regions lacking 1,3-β-linkages, belong to GHs families 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 74, 124 and 148 (www.cazy.org).

It is well established that 1,3-1,4-β-glucanases (EC 3.2.1.73) or 1,3(1,4)-β-glucanases (EC 3.2.1.6) do not hydrolyse cellulose, but are specific enzymes attacking precisely β-glucans. In contrast, cellulases/1,4-β-glucanases (EC 3.2.1.4) are able to degrade both cellulosic substrates and also the more soluble 1,3-1,4-β-glucans, by acting on the abundant 1,4-β-linkages of the mixed linked carbohydrate (Xue *et al.* 2003). Exogenous enzymes used to supplement barley-based diets are tested by their 1,3-1,4-β-glucanase activity, although it is generally unknown which enzymes contribute most for this activity, if the less specific cellulases/1,4-β-glucanases (EC 3.2.1.4) or the strictly specific 1,3-1,4-β-glucanases (EC 3.2.1.6 or EC 3.2.1.73), that do not cleave other polysaccharides apart from β-glucans. To attempt clarifying which activity is more relevant to depolymerize β-glucans in barley-based

diets, we recently compared the capacity of two highly pure *Clostridium thermocellum* GHs , recombinantly produced in *Escherichia coli*, the GH16 1,3-1,4-β-glucanase A (Ribeiro *et al.* 2012), *Ct*Glc16A, and the GH8 1,4-β-glucanase A, *CtCel8A*, also known as cellulase 8A (Cornet *et al.* 1983), to improve nutritive value of barley-based diets for poultry. The data revealed that the GH8 cellulase/1,4-β-glucanase A is, *in vitro*, very effective in the hydrolysis of barley β-glucans, but, *in vivo,* it was unable to depolymerize the anti-nutritive polysaccharide. Therefore, it was suggested that cellulases/1,4-β-glucanases (EC 3.2.1.4) are not adequate feed enzymes because they may be "trapped" by the highly abundant cellulose, thus becoming less available to perform the required β-glucan depolymerization (Fernandes *et al.* 2016). However, this conclusion was restricted to the context of GH8 cellulases/1,4-β-glucanases. It is indeed possible that cellulases from other families, such as GHs families 5, 6, 7, 9, 10, 12, 26, 44, 45, 48, 51, 74, 124 and 148, may access the polysaccharides through different mechanisms.

In the ecological niches where cellulose is degraded, GH5 cellulases are one of the most prevalent biocatalysts attacking cellulose. Here, we have compared the capacity of the specific GH16 1,3-1,4-β-glucanase from *C. thermocellum*, *Ct*Glc16A (Ribeiro *et al.* 2012), in parallel with the GH5 1,4-β-glucanase from *C. thermocellum* , *Ct*Cel5E, also known as cellulase 5E (Taylor *et al.* 2005), to improve the nutritive value of a barley-based diet for poultry. The two enzymes were recombinantly expressed in *E. coli*, purified to become free of other side activities, and used to supplement a barley-based diet that was fed to broiler chicks from day 1 to 28.

2.2. Materials and Methods

2.2.1. Gene isolation and cloning into a prokaryotic expression vector

The thermostable 1,3-1,4-β-glucanase of *C. thermocellum*, termed *Ct*Glc16A (Ribeiro *et al.* 2012), is a modular enzyme containing an N-terminal glycoside hydrolase family 16 catalytic domain and a C-terminal dockerin. The gene was isolated and cloned into pET21a, as described previously (Ribeiro *et al.* 2012), generating pGH16. The thermostable 1,4-β-glucanase of *C. thermocellum,* termed *Ct*Cel5E, is the catalytic domain of the bi-functional cellulase *Ct*Lic26A-Cel5E, a modular enzyme containing an N-terminal glycoside hydrolase family 26 catalytic domain, an internal 1,4-β-glucanase A (*Ct*Cel5E catalytic domain), an internal family 11 carbohydrate-Binding module (CBM11) and a C-terminal dockerin (Taylor *et al.* 2005). The gene encoding *Ct*Cel5E (residues 325-630) was isolated from *C. thermocellum* genomic DNA through PCR using the primers 5´-CTCGCTAGCAAGGCTGTCGACCCCTT-3´ and 5´-CACCTCGAGATTAAATAGTGCATTGAGG-3´. The DNA amplifications were performed using Supreme NZYProof DNA Polymerase (NZYTech genes & enzymes, Portugal) and the

primers incorporated NheI/XhoI restriction sites such that the recombinant glycoside hydrolases contained a C-terminal His6 tag. The obtained PCR products were cloned into pGem T-easy and sequenced to ensure that no mutations had occurred during the amplifications. The obtained pGem T-easy recombinant plasmids were digested with *Nhe*I and *Xho*I and the excised Clostridia DNA fragment was cloned into the similarly restricted expression vector pET21a to generate pGH5. Recombinant *Ct*Glc16A and *Ct*Cel5E were produced from *E. coli* BL21 (DE3) after transformation with plasmids pGH16 and pGH5, respectively. Both recombinant proteins contained an engineered C-terminal His6 tag allowing the direct purification of the two polypeptides by immobilized metal affinity chromatography (IMAC).

2.2.2. Expression and purification of *Ct*Glc16A and *Ct*Cel5E

To express *Ct*Glc16A and *Ct*Cel5E, *E. coli* BL21 (DE3) harbouring, respectively, the pGH16 and pGH5 recombinant plasmids, were cultured in Luria-Bertani broth containing 100 μg mL^{-1} of ampicillin at 37 ºC to mid-exponential phase (A_{550} 0.5). At this point, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM and the cultures were incubated for further 5 h. Cells were collected by centrifugation and the resultant cell pellet was resuspended in 10 mM Imidazole buffer (50 mM NaHepes, pH 7.5, 10 mM Imidazole, 1 M NaCl, 5 mM $CaCl_2$,) followed by disruption through ultrasonication. Cell debris were removed by centrifugation at 4 ºC, 13000 ×g during 30 min. The obtained His6-tagged recombinant protein extracts were purified by IMAC using 5 mL HiTrap chelating columns (GE Healthcare, USA) as described by Fontes *et al*. (2004). The purity of the resulting proteins was analysed by SDS-PAGE (Laemmli 1970).

2.2.3. Biochemical properties of *Ct*Glc16A and *Ct*Cel5E

Unless otherwise stated, enzyme assays were determined following the method described by Fontes *et al*. (2000) by measuring the release of reducing sugars resulting from carbohydrate hydrolysis in Phosphate Citrate (PC) buffer (64 mM K_2HPO_4 and 12 mM citric acid; pH=6.5) at 40 ºC. The substrate used in this study was barley β-glucan at 0.25% (w/v) final concentration and substrate saturation conditions were confirmed by measuring enzyme activities at different time points. Reactions were stopped by adding a DNSA-based solution (1% DNSA, 1% NaOH and 0.2% phenol) following the method described by Miller (1959). The pH profiles and thermostability of *Ct*Glc16A and *Ct*Cel5E enzymes were explored as described by Fernandes *et al*. (2016). Resistance to proteolysis was tested essentially as described previously (Fontes *et al.* 1995); basically, both proteins were incubated with porcine pancreatin (Sigma P-1500) at 37 ºC and the residual enzyme activity was measured . Qualitative analysis of 1,3-1,4-β-glucanase activity of digesta samples was assessed in agar

plates, using barley β-glucan at 0.1% (w/v) mixed with agar at 2% (w/v) final concentration, in 10 mM Tris-HCl (pH 7.5).

2.2.4. Animals and diets

One hundred and sixty 1-d-old Ross 308 male broiler birds were assigned to 40 pens of 4 birds each. Chicks were wing-banded for individual identification. The 40 pens were randomly assigned to 4 treatments consisting of a barley-based diet not-supplemented with exogenous enzymes (treatment NC) or supplemented with the commercial enzyme Rovabio™ Excel AP (Adisseo, France; treatment PC), the 1,3-1,4-β-specific glucanase *Ct*Glc16A (treatment *Ct*Glc16A) or the broadly specific 1,4-β-glucanase *Ct*Cel5E (treatment *Ct*Cel5E). All the three enzymes were incorporated at a calculated dose of 1400 U/kg of feed. The calculated 1400 U/kg of feed of the commercial enzyme corresponded to the manufacturers recommended dose of 50 g of enzyme per ton of feed. The commercial enzyme mixture was a blend of *Talaromyces versatilis* 1,4-β-xylanase (EC 3.2.1.8) and 1,3-1,4-β-glucanase (EC 3.2.1.73). The basal diet, which composition is displayed in Table 2.1, contained 618 g/kg of barley and was formulated to ensure a nutrient availability as defined by the NRC (1994). Experimental diets were pelleted. Pelleting temperature was 60 ºC and pellet diameter was 3 mm.

The duration of the trial was 28 days. Throughout the experiment, chicks were given free access to water and feed, which were provided with drinking nipples and hanging feeders, respectively. Broilers were raised in wired floor pens that were located in an environmentally controlled room adjusted daily to the recommended temperatures, according to standard brooding practices. Feed consumption and body weight (BW) was determined weekly and mortality was recorded daily. At the end of the trials, one bird per pen was slaughtered by cervical dislocation. The weight of the crop, gizzard and liver, and the length of the duodenum, jejunum, ileum and caecum were determined. Digesta samples from duodenum and jejunum, and ileum compartments were collected to determine contents enzyme activity as described above. The bird experiment was conducted in accordance with the Ethics Committee of Interdisciplinary Centre of Research in Animal Health (CIISA, Faculty of Veterinary Medicine, University of Lisbon, Portugal) and approved by the Animal Care Committee of the National Veterinary Authority (Direcção Geral de Veterinária, Lisboa, Portugal), following the appropriate European Union guidelines (Council Directive 2010/63/UE).

Table 2.1. Ingredient composition and calculated analysis of the barley-based feed.

Ingredients	**g/ kg**
Barley	615.0
Soybean meal 47%	294.0
Soybean oil	59.0
Salt	2.5
Calcium carbonate	8.4
Dicalcium phosphate 18%	17.4
DL-Methionine	1.7
Mineral and vitamin premix[1]	2.0
Nutrient content	
Energy (MJ ME/kg DM)	12.1
Crude protein	208.0
Ether extract	75.0
Crude fiber	51.0
Methionine	4.7
Methionine + Cysteine	8.2
Lysine	10.9
Threonine	7.6
Calcium	9.2
Available phosphorus	4.1

[1]Mineral-vitamin premix provided the following per kilogram of feed: retinol, 2.7 mg; cholecalciferol, 0.05 mg; α-tocopherol, 20 mg; nicotinic acid, 30 mg; cyanocobalamin, 0.12 mg; calcium pantothenate, 10 mg; menadione, 2 mg; thiamin, 1 mg; riboflavin, 4.2 mg; pyridoxine hydrochloride, 1.7 mg; folic acid, 0.5 mg; biotin, 0.5 mg; Fe, 80 mg; Cu, 10 mg; Mn, 100 mg; Zn, 80 mg; Co, 0.2 mg; I, 1.0 mg; Se, 0.3 mg; monensin, 100 mg/kg.

2.2.5. Statistical Analyses

Data related to bird performance were subjected to ANOVA according to the general linear models procedure of SAS (2004) in order to detect significant differences between treatment groups. Chi-squared test has been performed to statistically assess if the presence of the enzymatic activity in gastrointestinal contents was different among groups. The experimental unit was the cage of 4 animals (n=10). Unless otherwise stated, differences were considered significant when $P<0.05$.

2.3. Results and Discussion

2.3.1. Biochemical properties of *Ct*Glc16A and *Ct*Cel5E

The capacity of *Ct*Glc16A and *Ct*Cel5E to function in the conditions prevalent in the GI tract of monogastric animals was assessed essentially as described by Fernandes *et al.* (2016). The data, presented in Figure 2.1, revealed that the thermostability of the two enzymes is remarkably similar (Figure 2.1B), reflecting their thermophilic origin. Thus, the activity of *Ct*Glc16A and *Ct*Cel5E are only severely compromised at temperatures above 80 ºC. In contrast, differences were found in the pH profiles, with *Ct*Glc16A being more active on the neutral to alkaline range, while *Ct*Cel5E activity prevails on the neutral to acidic range (Figure 2.1A). However, the two enzymes display a broad pH range of activity and remain significantly active at pHs between 4 and 8.5. In addition, *Ct*Glc16A and *Ct*Cel5E display complete resistance to proteolytic inactivation after a 30 min incubation with pancreatic proteases (data not shown), which suggests the retention of their molecular integrity during exposition to the proteolytic activity of the animal duodenum. Thus, the combination of these studies suggests that both *Ct*Glc16A and *Ct*Cel5E will be fully active in the conditions prevalent in the monogastric animal´s GI tract and thus express the required biochemical properties to actively contribute to the depolymerization of 1,3-1,4-β-glucans *in vivo*.

2.3.2. The efficacy of *Ct*Glc16A and *Ct*Cel5E to improve the nutritive value of a barley-based diet

The capacity of a family 16 1,3-1,4-β-glucanase, *Ct*Glc16A, and a typical family 5 cellulase/1,4-β-glucanase, *Ct*Cel5E, to improve the nutritive value of a barley-based diet for broilers was evaluated. The two enzymes were recombinantly produced in *E. coli* and purified. The 1,3-1,4-β-glucanase activity of both thermostable enzymes as well as, the positive control mixture were evaluated *in vitro* against purified barley β-glucan. Identical levels of the three enzyme preparations were used to supplement the barley-based diet displayed in Table 2.1. Data concerning the performance of birds fed on the four diets under study is presented in Table 2.2. At days 14, 21 and 28 birds fed on the CtGlc16A diet display a higher body weight than the negative control group consisting on the animals fed on the diet not supplemented with exogenous enzymes. In contrast, animals fed on diets receiving the recombinant *Ct*Cel5E display, throughout the entire period of the trial, a body weight identical to the animals of the negative control group. During the period of the experiment, body weight of animals supplemented with the commercial enzyme mixture (positive control) and the recombinant *Ct*Glc16A were similar, although at day 28 body weights of positive control animals was also similar to the negative control group. These data suggest that while *Ct*Glc16A enzyme was able to improve the nutritive value of the barley-based diet, cellulase/1,4-β-glucanase *Ct*Cel5E, which *in vitro* is very effective in degrading barley β-

glucans, was unable to affect the performance of the *Ct*Cel5E group. The reported differences in body weight result from similar differences measured in weight gain for the period of 7-28 days of the trial. Feed intake and feed conversion ratio (FCR) were generally not affected by the addition of exogenous enzymes, although during the 7-14 days period animals of the positive control and *Ct*Glc16A groups had smaller FCR's than the negative control animals. Together, the data confirm that *Ct*Glc16A is highly effective in improving the nutritive value of the barley-based diet for broilers while it is suggested that *Ct*Cel5E is unable to significantly affect animal performance. Thus, the data suggest that cellulase *Ct*Cel5E is unable to significantly depolymerize *in vivo* barley β-glucans.

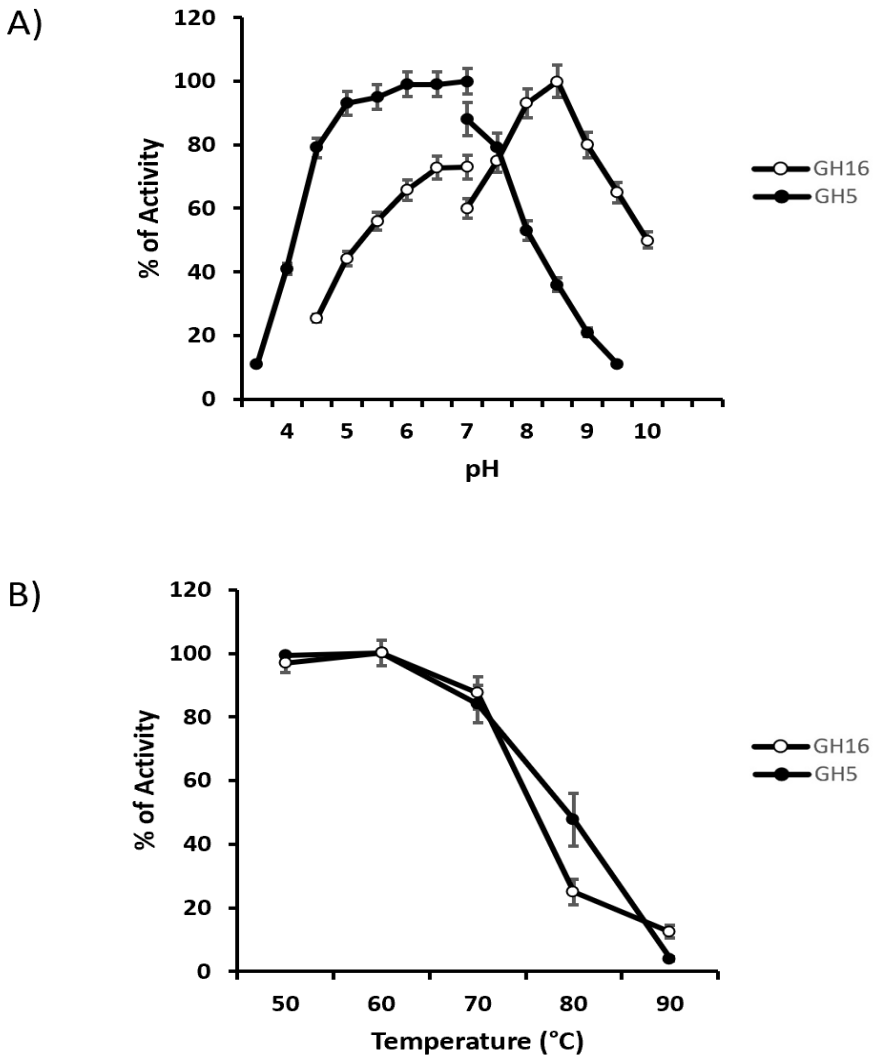


Figure 2.1. pH and temperature profiles of *Ct*Glc16A and *Ct*Cel5E.

In Panel A (pH), enzymes were incubated at standard conditions in MES, Tris or $NaHCO_3$ buffers presenting different pHs, followed by determination of β-glucanase activity. In Panel B (temperature), *Ct*Glc16A and *Ct*Cel5E were incubated for 30 min at different temperatures, and the residual activity was determined at a neutral temperature. CtGlc16A (○) and CtCel5E (●).

Table 2.2. Growth performance parameters of broilers fed on a barley-based diet

NC (non-supplemented with exogenous polypeptides) or PC (supplemented with a commercial enzyme mixture), *Ct*Glc16A (with the recombinant 1,3-1,4-β-glucanase) or *Ct*Cel5E (with the recombinant 1,4-β-glucanase).

	NC	PC	*Ct*Glc16A	*Ct*Cel5E	SEM	P-value
Body Weight (g)						
0d	46,7	46,7	46,8	46,5	0,44	0,943
7d	120	128	127	123	3,8	0,439
14d	256^{b}	321^{a}	329^{a}	267^{b}	9,2	0,001
21d	534^{b}	629^{a}	655^{a}	562^{b}	17,0	0,001
28d	927^{b}	1031^{ab}	1077^{a}	938^{b}	34,4	0,006
Weight Gain (g)						
0-7d	73,6	81,5	80,2	76,6	3,87	0,474
7-14d	135^{b}	192^{a}	194^{a}	144^{b}	8,30	0,001
14-21d	278^{b}	311^{ab}	335^{a}	295^{b}	9,35	0,001
21-28d	835	395	425	376	21,9	0,382
0-28d	881^{b}	981^{ab}	1035^{a}	892^{b}	34,9	0,006
Feed Intake (g)						
0-7d	107	114	108	108	5,21	0,780
7-14d	227^{b}	285^{a}	284^{a}	226^{b}	9,03	0,001
14-21d	471	500	508	473	16,6	0,310
21-28d	728	757	752	724	27,8	0,760
0-28d	1552	1656	1652	1532	45,8	0,104
Feed Conversion Ratio						
0-7d	1,45	1,40	1,35	1,47	0,065	0,547
7-14d	1,72^{a}	1,49^{b}	1,47^{b}	1,59^{ab}	0,050	0,004
14-21d	1,69	1,62	1,51	1,61	0,048	0,088
21-28d	1,89	2,14	1,78	2,00	0,233	0,689
0-28d	1,76	1,72	1,60	1,73	0,068	0,315

Means in the same line with different letter superscripts (a, b, c) are significantly different (P<0.05) or tend to be significantly different (P<0.1).

In the last years we have been using purified recombinant enzymes to dissect the biological role of different GHs to improve the nutritive value of cereal-based diets for monogastric animals. Concerning barley-based diets, it is well known that current enzyme mixtures provided commercially display a large repertoire of plant cell wall degrading enzymes, including cellulases/1,4-β-glucanases and 1,3-1,4-β-glucanases. In Nature cellulases/1,4-β-glucanases are highly prevalent enzymes as they attack the widely abundant cellulosic substrates that accumulate in different ecological niches, providing a major contribution to carbon recycling. It is well recognized that cellulases also attack mixed

linked β-glucans, in particular the homogenous 1,4-β regions of the polysaccharide. Until recently, the role of cellulases/1,4-β-glucanases to improve the nutritive value of barley-based diets for monogastric animals remained relatively obscure. It was anticipated that the major contribution of exogenous enzymes used to supplement barley-based diets for monogastric animals would consist in the reduction of the degree of polymerization of barley β-glucans, which may be achieved by a cellulase randomly cleaving a few 1,4-β-glycosidic linkages within the polysaccharide. The reduction of the degree of polymerization of the carbohydrate would lead to a reduction in the digesta viscosity that arises as a consequence of the solubilization of the large β-glucan polymers. However, a recent study revealed that indeed *in vivo* a GH8 cellulase was unable to affect the nutritive value of a barley-based diet for poultry (Fernandes *et al.* 2016). It was suggested that although the GH8 enzyme was active in the animals' digestive compartments it was, *in vivo*, trapped by the highly abundant cellulosic substrates present in the diets thus suggesting that the enzyme would become unavailable for β-glucan hydrolysis. Data presented here aims to clarify if this is a common phenomenon observed generally for cellulases and, as such, a similar experiment was undertaken using the most prevalent cellulases in Nature, the GH5 enzymes. The data suggest that similarly to GH8, the GH5 enzyme used in this study, *Ct*Cel5E from *C. thermocellum*, is unable to improve the nutritive value of barley-based diets for monogastric animals. In contrast, the 1,3-1,4-β-glucanase *Ct*Cel16A effectively improves the nutritive value of barley-based diets for broilers and most probably acts by reducing the degree of polymerization of the anti-nutritive soluble β-glucans present in the cereal (Marquardt *et al*. 1996; Silva and Smithard 2002; Lázaro *et al*. 2003).

2.3.3. Is *Ct*Cel5E stable during passage through the animal GI tract?

The inability of *Ct*Cel5E to improve the nutritive value of barley-based diets suggests that the enzyme may be inactivated during passage through the animal GI tract. Thus, samples from the various GI compartments were collected and the respective 1,3-1,4-β-glucanase activity measured. The data (not shown) revealed that, although 1,3-1,4-β-glucanase activity was detected in the crop and caecum of the majority of animals of all groups, only birds receiving diets supplemented with the exogenous enzyme mixtures, including *Ct*Cel5E, display significant levels of enzyme activity in the other regions of the GI tract, including the small intestine. This observation confirms the biochemical data reported above suggesting that *Ct*Cel5E is stable and remains active in the environmental conditions prevailing in the GI tract.

Data presented here suggest that the inefficacy of cellulase/1,4-β-glucanase *Ct*Cel5E to improve the nutritive value of barley-based diets for broilers is not related to the inactivation of the enzyme during passage through the GI tract. A similar observation was

made in a previous study conducted with the GH8 cellulase *Ct*Cel8A, that was found to be unable to improve the nutritive value of barley-based diets for poultry although remaining fully functional in the conditions prevailing in the GI tract (Fernandes *et al.* 2016). The observation that the capacity of the GH8 enzyme to reduce the viscosity of a β-glucan solution *in vitro* was severely reduced in the presence of cellulosic substrates, suggested that, *in vivo,* cellulases may be involved in non-productive interactions with the highly abundant cellulose, thus becoming unable to attack the anti-nutritive β-glucans. In contrast, since *Ct*Glc16A is unable to interact with cellulose as the catalytic site of 1,3-1,4-β-glucanases will not interact with linear 1,4-β-glucans, the enzyme will be exclusively involved in the development of productive interactions with mixed linked glucans that consist on their target substrates. Overall, data presented here suggest that in general cellulases/1,4-β-glucanases, i.e. enzymes that attack 1,4-β-glucans, in particular cellulose, will not attack 1,3-1,4-β-glucans in the context of complex carbohydrate mixtures rich in cellulose as they become “trapped” by their target cellulosic substrates. Thus, the inefficacy revealed by purified cellulases to improve the nutritive value of barley-based diets for poultry does not result from the inactivation of cellulases neither of their incapacity to act in the animal GI tract, but rather from the formation of unproductive interactions with cellulose instead of attacking the anti-nutritive 1,3-1,4-β-glucans.

2.4. Conclusions

The use of exogenous enzymes to improve the nutritive value of animal feed has attracted wide interest. However, a clarification of the different biological roles played by enzymes under the context of a variety of diets and animal species is urgently required. For example, it is now well established that microbial glycoside hydrolases improve the nutritive value of barley-based diets for monogastric animals. Nevertheless, the nature of the biocatalysts that effectively contribute to reduce the detrimental effects associated with the presence of β-glucans in barley (if the naturally widely abundant cellulases, or the highly specific β-glucanases, or both), remains obscure. Taken together, the results reported here and by Fernandes *et al*. (2016) put in evidence that cellulases/1,4-β-glucanases (EC 3.2.1.4) from CAZy families GH8 (Fernandes *et al.* 2016) and GH5 (this study) are not efficient enzymes to improve the nutritive value of barley-based diets for poultry. This observation suggests that eventually all cellulases will be unable to effectively hydrolyse *in vivo* the anti-nutritive β-glucans due to their preferential substrate specificity for cellulosic carbohydrates. It is suggested that in the context of animal diets that are generally richer in cellulose than β-glucans, cellulases/1,4-β-glucanases (EC 3.2.1.4) will made unproductive interactions with their targeted cellulosic substrates leaving β-glucans undegradable. In contrast, enzymes targeting specifically β-glucans, such as the GH16 1,3-1,4-β-glucanase (EC 3.2.1.73)

reported in this study, will not be “trapped” by other competing polysaccharides present in the cereal-based diet and thus will productively contribute to degrade the anti-nutritive carbohydrates present in barley-based diets. The significance of these observations suggests that effective microbial mixtures used to supplement barley-based diets for monogastric animals should contain primarily 1,3-1,4-β-glucanases (EC 3.2.1.73) but not cellulases/1,4-β-glucanases (EC 3.2.1.4). Although cellulases/1,4-β-glucanases effectively depolymerize β-glucans in conditions where they have no access to cellulosic substrates, presence of significant levels of cellulose in common animal diets limit the scope of cellulases/1,4-β-glucanases to hydrolyse the anti-nutritive β-glucans. In future, the potential “trapping” of exogenous feed enzymes by competing substrates present in animal diets should be considered when evaluating the efficacy of feed enzymes.

3. VARIATION IN LEVELS OF NON-STARCH POLYSACCHARIDES AND ENDOGENOUS ENDO-1,4-B-XYLANASES AFFECT THE NUTRITIVE VALUE OF WHEAT FOR POULTRY

V. Cardoso[*,1], E.A. Fernandes[†,1], H.M.M. Santos[*], B. Maçãs [‡], M.M. Lordelo[†], L.M.A. Ferreira[*], C.M.G.A. Fontes[*,2] and T. Ribeiro[*]

[†]Instituto Superior de Agronomia, Universidade de Lisboa, 1349-017 Lisboa, Portugal; [*]CIISA, Faculdade de Medicina Veterinária, Universidade de Lisboa, Pólo Universitário do Alto da Ajuda, Av. da Universidade Técnica, 1300-477 Lisboa, Portugal; [‡]INRB/INIA, Estrada Gil Vaz, 7350-228 Elvas, Portugal.

Adapted from: Cardoso, *et al.* (2018). *British Poultry Science*. 59(2):218-226

Abstract

1. Endo-1,4-β-xylanase is known to improve the nutritive value of wheat-based diets for poultry by degrading dietary arabinoxylans. However, broilers response to supplementation of wheat-based diets with exogenous endo-1,4-β-xylanase is not always observed.

2. In this study, 108 different wheat lots were analysed for levels of extract viscosity as well as for endogenous endo-1,4-β-xylanase activity and the impact of these two variables in animal performance was tested.

3. Results revealed that endogenous endo-1,4-β-xylanase activity and extract viscosity content varied widely among different wheat lots. Thus, a trial was conducted to evaluate the efficacy of exogenous enzyme supplementation in broiler diets using wheats with different levels of extract viscosity and endogenous endo-1,4-β-xylanase activity.

4. The data revealed that exogenous enzyme supplementation was only effective when the wheat present in the diet had high levels of extract viscosity (14.8cp) with low endogenous endo-1,4-β-xylanase activity (347.0 U/kg). Nevertheless, it is apparent that exogenous microbial xylanases reduce digesta extract viscosity and feed conversion ratio independently of the endogenous properties presented by different wheat lots.

5. The data suggest that extract viscosity and/or endogenous endo-1,4-β-xylanase activity affects the response to enzyme supplementation by poultry fed on wheat-based diets.

3.1. Introduction

Wheat contains a significant proportion of non-starch polysaccharides (NSPs), specifically arabinoxylans, which display considerable antinutritive properties for poultry (Choct and Annison 1992; Choct 2006). Soluble arabinoxylans can form highly viscous solutions as they can absorb up to ten times their weight in water (Choct 1997). High digesta viscosity decreases the rate of diffusion of substrates and digestive enzymes, decreasing the efficacy of the repertoire of biocatalysts secreted to the gastrointestinal (GI) tract of poultry and consequently reducing feed digestibility (Johnson and Gee 1986; Ikegami *et al*. 1990). In addition, it is apparent that high digesta viscosity decreases the digesta passage rate, which results in a decrease in feed intake. Besides the direct influence in digestion, high viscosity can also cause an enlargement of the digestive organs and influence the microbiome that colonizes the poultry GI tract with a direct influence in performance and health. It is well established that soluble NSPs may be used as substrates for fermentation in the small intestine favouring the development of abnormal anaerobic microflora in this GI portion (Choct 1997). The conjugation of all these adverse effects resulting from anti-nutritive NSPs is known to impair poultry performance.

Supplementation of wheat-based diets with exogenous endo-1,4-β-xylanases is today a common practice to improve the nutritive value of wheat-based diets for poultry (Silva and Smithard 2002). Activity of exogenous microbial endo-1,4-β-xylanases added to poultry diets reduces the degree of polymerization of the recalcitrant arabinoxylans and as such the deconstruction of the antinutritive carbohydrates results in lower digesta viscosities. Bedford (1996) reported that even small reductions in digesta viscosity can promote nutrient digestibility to a considerable extent due to an improvement in the efficacy of endogenous digestive enzymes. The decrease in viscosity also leads to a faster digesta passage rate, which increases the feed intake capacity. On the other hand, nutrient absorption is promoted as a result from a closer contact with the intestinal lumen. It is also apparent that the xylo-oligosaccharides resulting from the degradation of arabinoxylans by the direct effect of exogenous endo-1,4-β-xylanases have prebiotic properties that are highly beneficial to the animal (Apajalahti and Bedford 1999). Thus, as a result of the above mentioned effects, addition of endo-1,4-β-xylanases to poultry diets leads to an improved animal performance.

In particular in Europe, but also worldwide, wheat is commonly used for broiler feed production. Although supplementation of wheat-based diets for poultry with exogenous microbial endo-1,4-β-xylanases is a current practice, it is apparent that in a wide range of situations endo-1,4-β-xylanase supplementation fails to have an impact in animal performance (Murphy *et al*. 2009). This is not restricted to wheat-based diets as it is known that, in certain cases, the inclusion of exogenous carbohydrate-active enzymes in diets containing a high percentage of barley or rye also fails to have any effect in animal

performance (Bedford 2000). Previous results suggested that a variety of factors may influence the direct efficacy of exogenous enzymes used to supplement poultry diets, such as the cereal cultivar, growing conditions, energy and protein content and extract viscosity, among others (Pirgozliev *et al*. 2003). In addition, relationship between energy availability in cereal grains and content of soluble NSP is inversely related in poultry (Villamide *et al*. 1997). Since levels of NSP may vary between cereal lots, cereals may express different nutritive values. Several factors may affect NSP content in grains, which include cereal genotype and growing conditions, length of the cereal storage period, grain cultivar, growing season, or soil type, among others (Villamide *et al.* 1997).

In a previous study, Ribeiro and colleagues (2011) reported that response by broilers to supplementation of barley-based diets with exogenous endo-1,3-1,4-β-glucanases is dramatically affected by levels of endogenous endo-1,3-1,4-β-glucanases present in the barley grain. Thus, it was revealed that when the barley lot used to prepare broiler diets contains high levels of endogenous endo-1,3-1,4-β-glucanase activity, exogenous enzyme supplementation is redundant and does not improve animal performance (Ribeiro *et al.* 2011). Different reports confirm that wheat also contains a variety of endogenous endo-1,4-β-xylanases, which are required to modulate cell wall differentiation during plant grow (Dornez *et al*. 2006; Gebruers *et al*. 2010). In addition, data reported by Gebruers and colleagues (2010) suggested that there may be substantial differences in levels of endo-1,4-β-xylanase activity between different wheat types and varieties. However, little research attention has been focused on the characterization of the levels of endogenous endo-1,4-β-xylanases in wheat and their effects on the efficacy of exogenous microbial arabinoxylan degrading enzymes used to supplement wheat-based diets for poultry. The goal of the present study was to characterize wheat extract viscosity and endo-1,4-β-xylanase activity in a wide range of wheat lots. In addition, the direct effects of these two factors in the efficacy of exogenous enzymes used to supplement wheat-based diets for poultry was evaluated.

3.2. Materials and Methods

3.2.1. Quantification of endo-1,4-β-xylanase and viscosity in different wheat lots

One hundred and eight wheat lots from different origins were evaluated for viscosity and endo-1,4-β-xylanase activity. This range of wheat lots was from different origin and/or varieties harvested between 2008 and 2009 (E1-E17, SA1-SA17, UK1-UK19) or harvested between 2007 and 2012 (W1 – WW55). Wheat lots coded as “W” were lots commercially available and most of them without information regarding origin or variety. Wheat enzyme activity was measured with an endo-1,4-β-xylanase assay kit (Megazyme, Ireland). The method followed the manufacturers’ protocol with the exception of the incubation period, which was extended for 2 hours. To measure the viscosity of the different wheat lots,

samples were milled at 0.5 mm and 15 g was then mixed with 15 mL of phosphate citrate buffer (pH 6.5) by vigorously shaking for 5 min. Samples were then centrifuged for 10 min at 7500 ×g and the viscosity of the supernatant was measured in a Brookfield viscometer (model LVDVCP-II, Brookfield Engineering Laboratories, Middleboro, MA) with a cup maintained at 24 °C. Analysis of wheat dry matter (method 934.01), crude fat (method 920.39), crude protein (method 954.01), Neutral Detergent Fiber (method 2002.04) and Acid Detergent Fiber and Acid Detergent Lignin (method 973.18) were performed according to the methods of A.O.A.C. (1980). Measurements were performed in triplicate.

3.2.2. Animals, Diets and Management

Animal experiments were conducted in accordance with the principles and specific guidelines of the European Union (1986) and approved by the Animal Care Committee of the National Veterinary Authority (Direcção-Geral de Veterinária, Lisboa, Portugal). Three hundred and sixty 1-d-old chicks (Ross 308) were individually weighed and randomly divided into 24 floor pens with 15 animals each. Water and mashed feed were available ad libitum throughout the experiment (35 days) and were provided from automatic drinking nipples and hanging feeders, respectively. The floor pens were located in an environmentally controlled room, which was adjusted daily to the recommended temperatures according to standard brooding practices. The composition of the basal diet used in this study is presented in Table 3.1 and was formulated to contain adequate nutrient levels as defined by the NRC (1994). The study consisted of four treatments in a 2x2 factorial arrangement with two different wheat lots, termed HALV (for high endogenous endo-1,4-β-xylanases activity and low viscosity) and LAHV (for low endogenous endo-1,4-β-xylanases activity and high viscosity), which were used to prepare two different basal diets that were supplemented with (+) or without (-) a microbial endo-1,4-β-xylanase preparation (Avizyme 40000G, DuPont Industrial Biosciences – Danisco Animal Nutrition, Marlborough, United Kingdom). The exogenous enzymes were incorporated following the manufacturer's recommendation (62.5 g per ton of feed). Feed consumption and individual Body Weight (BW) were recorded weekly. Feed conversion ratios were calculated by dividing the weight gain per pen by the feed consumption in the respective period. Bird mortality was recorded daily.

Table 3.1. Ingredient composition (g/kg) and calculated analysis of the wheat-based diets prepared in this study.

A total of four diets were prepared using a wheat variety displaying high endo-1,4-β-xylanase activity and low viscosity (termed HALV variety) or a wheat variety expressing low endo-1,4-β-xylanase activity and high viscosity (termed LAHV variety). The HALV and LAHV basal diets were supplemented or not supplemented with an exogenous xylanase mixture.

Ingredient	**g/kg**
Wheat	600
Soybean meal (47%)	290
Corn	41.7
Soybean oil	27.5
Sodium chloride	3.0
Calcium carbonate	14.3
Dicalcium phosphate (18%)	18.0
L-lysine	1.5
DL-Methionine	2.0
Mineral and vitamin premix[a]	2.0
Calculated nutrient content	
Energy (MJ of ME/kg DM)	12.6
CP (%)	19.9
Ether extract (%)	4.70
Crude cellulose (%)	2.93
Ash (%)	6.55

[a] Mineral-vitamin premix provided the following per kilogram of diet: vitamin A, 9,000 IU; vitamin D3, 2,100 IU; vitamin E, 20 mg; nicotinic acid, 30 mg; vitamin B12, 0.12 mg; calcium pantothenate, 10 mg; vitamin K3, 2 mg; thiamin, 1 mg; riboflavin, 4.2 mg; vitamin B6, 1.7 mg; folic acid, 0.5 mg; biotin, 0.5 mg; Fe, 80 mg; Cu, 10 mg; Mn, 100mg; Zn, 80 mg; Co, 0.2 mg; I, 1.0 mg; Se, 0.3 mg; monensin, 100 ppm.

3.2.3. Analysis of digesta samples

At the end of the experiment, two animals per pen were electrically stunned and killed by cervical dislocation. The liver and the empty crop, gizzard, duodenum, jejunum, ileum and caecum were weighed and length of the latter four was also measured. Contents of the crop, duodenum, jejunum, ileum and caecum were collected and stored at -20 °C for later analysis of endo-1,4-β-xylanase activity. To measure the viscosity of the small intestine contents, samples were collected from the duodenum plus jejunum and ileum, centrifuged for 10 min at 12000 g and viscosity of the supernatant was subsequently measured at 24 °C as described above. Qualitative analysis for endo-1,4-β-xylanase activity detection was performed in agar

plates, using wheat arabinoxylan (Megazyme) at 0.1% final concentration (w/v) in 10 mM Tris pH 7.5 as described in Ponte *et al.* (2004) and Mourão *et al.* (2006). Catalytic activity was detected after 16 h of incubation at 37 °C using 20 µl of supernatant of the GI content in wells performed in agar plates.

3.2.4. Statistical analysis

Data were subjected to ANOVA assuming a 2x2 factorial arrangement of treatments, using the General Linear Models procedure of SAS software (SAS Institute, 2012). The linear model included the main effects of wheat viscosity and exogenous enzyme supplementation, plus their interaction. In case the interaction was significant ($P<0.05$), comparisons were made between the means for combinations of wheat viscosity-exogenous enzyme supplementation, while in the case of non-significant interactions the comparisons were established among means for each of the main effects. The experimental unit was a pen of 15 birds (n=6). Unless otherwise stated, differences were considered significant at $p < 0.05$. The relationship between endo-1,4-β-xylanase and viscosity in wheat lots was assessed, and regression analyses were conducted to test for linearity of association between level of endo-1,4-β-xylanase and viscosity in wheat lots.

3.3. Results and Discussion

3.3.1. Variation of endogenous endo-1,4-β-xylanase activity and extract viscosity in different wheat lots

A previous study by Gebruers and colleagues (2010) revealed that up to 70% of the endo-1,4-β-xylanase activity detected in wheat originated in endogenous endo-1,4-β-xylanases synthesised in the grain. Here, to evaluate how endo-1,4-β-xylanase activity varies in wheat, a range of different wheats were selected. In total, one hundred and eight wheat lots were analysed for endo-1,4-β-xylanase activity and wheat extract viscosity (Table 3.2). The results, presented in Table 3.2, shown that levels of endogenous endo-1,4-β-xylanase activity in wheat varied from 27.3 to 1981.3 U/kg, which represents a 70-fold variation. In contrast, viscosity of the different wheat batches ranged from 1.49 to 39.1 cP, which is a lower than 30-fold variation (Table 3.2). Taken together these results demonstrate that both levels of endogenous endo-1,4-β-xylanase activity as well as extract viscosity vary significantly in wheat.

Table 3.2. Endo-1,4-β-xylanase activity (U/kg of wheat) and viscosity (cP) of different wheat lots[a,b].

Sample	Variety	Endo-1,4-β-xylanase	Viscosity
E1	Almansor (CV)	36.9	13.7
E2	Roxo (CV)	nd	8.53
E3	TNMU/6/CEP80111/CEP81165/5/MRNG/4 (BL)	27.3	16.2
E4	PSN/BOW/ROEK/3/Milan (BL)	nd	8.72
E5	Trident (Australia) (CV)	30.8	10.9
E6	THB/Maya/NAC/3/RABE/4/Milan (BL)	nd	12.6
E7	Munia/Altar84/Amsel (BL)	nd	13.6
E8	Anapo (CV)	34.7	6.31
E9	BOW/URES/2*Weaver (BL)	nd	18.8
E10	PF74354//LD/ALD/4/2*BR12*2/3/JUP/PAR214-6/… - BL	nd	33.9
E11	N5732/HER//SUDAN□11 – BL	47.6	8.05
E12	Patanegra (CV)	nd	16.2
E13	Anforeta (Italy) (CV)	nd	5.16
E14	Andana (Italy) (CV)	nd	10.6
E15	Buon-pastor (Italy) (CV)	44.0	4.68
E16	Randa (Italy) (CV)	111.3	9.75
E17	Salvia (Italy) (CV)	136.1	6.95
SA1	Unk (BL)	nd	12.6
SA2	Unk (BL)	nd	8.20
SA3	Unk (BL)	nd	19.4
SA4	Unk (BL)	nd	6.72
SA5	Unk (BL)	nd	16.4
SA6	Unk (BL)	nd	20.0
SA7	Unk (BL)	nd	7.33
SA8	Unk (BL)	nd	24.2
SA9	PF74354//LD/ALD/4/2*BR12*2/3/JUP… - (BL)	nd	36.5
SA10	Unk (BL)	nd	34.1
SA11	Unk (BL)	56.8	7.77
SA12	Unk (BL)	nd	19.5
SA13	Unk (BL)	59.4	6.20

SA14	Unk (BL)	nd	16.1
SA15	Unk (BL)	77.4	5.11
SA16	Unk (BL)	77.0	8.46
SA17	Unk (BL)	56.4	8.42
UK1	English (BL)	289.8	4.60
UK2	English (BL)	237.9	4.42
UK3	English (BL)	268.5	3.17
UK4	English (BL)	208.6	9.72
UK5	English (BL)	193.9	9.48
UK6	English (BL)	104.1	8.94
UK7	English (BL)	171.6	3.48
UK8	English (BL)	139.0	5.96
UK9	English (BL)	166.0	8.76
UK10	English (BL)	146.5	6.13
UK11	English (BL)	97.0	5.31
UK12	English (BL)	141.1	7.72
UK13	English (BL)	190.8	3.75
UK14	English (BL)	154.9	5.90
UK15	English (BL)	237.4	6.56
UK16	English (BL)	99.4	9.50
UK17	English (BL)	113.3	9.74
UK18	English (BL)	216.3	6.90
UK19	English (BL)	111.7	6.34
W1	Unk – PT	335.9	8.85
W2	Unk – PT	258.2	18.5
W3	Marialva – PT	189.7	6.41
W4	Patanegra – PT	261.9	19.8
W5	Unk	485.2	12.4
W6	Unk – UK	1587.7	7.74
W7	Unk	171.9	1.49
W8	Unk	1981.3	8.83
W9	Unk	548.5	8.68
W10	Unk	407.8	4.39
W11	Unk	680.4	9.39
W12	Unk – FR	386.8	8.43
W13	Unk– PT	471.6	10.3

W14	Unk– UK	635.2	17.6
W15	Marialva– PT	298.7	3.94
W16	Unk– PT	428.9	8.92
W17	Unk– PT	157.7	39.4
W18	Unk– PT	195.0	24.9
W19	Unk– PT	362.2	27.3
W20	Unk– PT	232.8	22.6
W21	Unk– PT	339.6	18.6
W22	Unk– PT	448.2	31.5
W23	Unk– PT	510.8	2.15
W24	Soft Red Winter – USA	419.3	.
W25	Unk	444.4	8.61
W26	Unk	195.0	15.4
W27	Unk	405.9	5.19
W28	Unk	499.0	8.81
W29	Unk	522.6	7.61
W30	Unk	256.4	.
W31	Unk – ES	639.3	11.9
W32	Unk– PT	1468.7	6.29
W33	Unk– PT	668.0	11.1
W34	Unk– PT	699.0	10.7
W35	Unk – PT	499.0	15.4
W36	Unk– PT	584.5	5.65
W37	Unk– PT	556.5	9.43
W38	Unk– PT	598.7	6.36
W39	Unk– PT	542.5	11.3
W40	Unk	1210.4	7.92
W41	Unk– PT	678.3	26.5
W42	Unk– PT	791.4	8.48
W43	Unk– PT	500.9	24.3
W44	Unk– PT	396.3	28.0
W45	Unk– PT	454.1	22.4
W46	Unk– PT	392.5	13.0
W47	Unk– PT	686.6	4.14
W48	Unk– PT	572.5	5.80
W49	Unk– PT	962.7	8.23

W50	Unk	544.0	6.87
W51	Unk	446.3	39.1
W52	Unk	489.2	11.2
W53	Unk	347.0	14.8
W54	Unk	499.0	10.5
W55	Unk	499.0	9.21

[a] Wheat W50 and W53 were used in the current animal experiment and were termed HALV and LAHV, respectively. [b] The Unk lots consist of varieties of unknown origin; (BL) refers to breeding lines; (CV) refers to cultivar lines; "E" and "SA" samples are from Portugal or Mediterranean areas; "UK" samples are United Kingdom varieties; Varieties named with "PT" refer to Portuguese origin; Varieties named with "UK" refer to United Kingdom origin; Varieties named with "FR" refer to French origin; Varieties named with "ES" refer to Spanish origin; Varieties named with "USA" refer to American origin. Nd = not detectable.

In the current study, a dramatic variation in the levels of endo-1,4-β-xylanase activity of different wheat lots was observed. It is well established that endo-1,4-β-xylanases play important biological roles in plants, mainly by modulating plant cell wall development and enlargement. In addition, in the particular case of wheat, endo-1,4-β-xylanase is strictly controlled by the activity of several wheat endo-1,4-β-xylanase inhibitors that were previously shown to affect poultry nutrition both *in vitro* (Smeets *et al*. 2014) or *in vivo* (Ponte *et al.* 2004). The large variation in endo-1,4-β-xylanase activity observed in the different wheat lots analysed here may result either from variation in the concentration of the carbohydrate-active enzymes present in different varieties, or different degrees of endo-1,4-β-xylanase inhibition, or both. Since levels of the endo-1,4-β-xylanase inhibitors are different between cultivars (Dornez *et al.* 2006), it is possible that this factor is an important modulator of endo-1,4-β-xylanase activity, even though this hypothesis remains to be analysed.

An elevated endogenous endo-1,4-β-xylanase activity in wheat could lead to an endogenous degradation of the anti-nutritive NSPs, leading to a lower wheat extract viscosity. Thus, to test this hypothesis the correlation between the two variables studied, endogenous endo-1,4-β-xylanase activity and cereal extract viscosity, was tested. The linear regression analysis between the levels of wheat endogenous endo-1,4-β-xylanase activity and wheat extract viscosity yielded an r^2 of 0.9, which indicates that there is no relationship between these two parameters (Figure 3.1). Thus, the two variables are not associated, suggesting that although endo-1,4-β-xylanase activity in planta is higher in different wheat lots the enzyme is not endogenously available for arabinoxylan hydrolysis and, therefore, will not impact wheat extract viscosity. This observation does not preclude the capacity of endogenous endo-1,4-β-xylanases to affect wheat viscosity *in vivo*, as it will be analysed below.

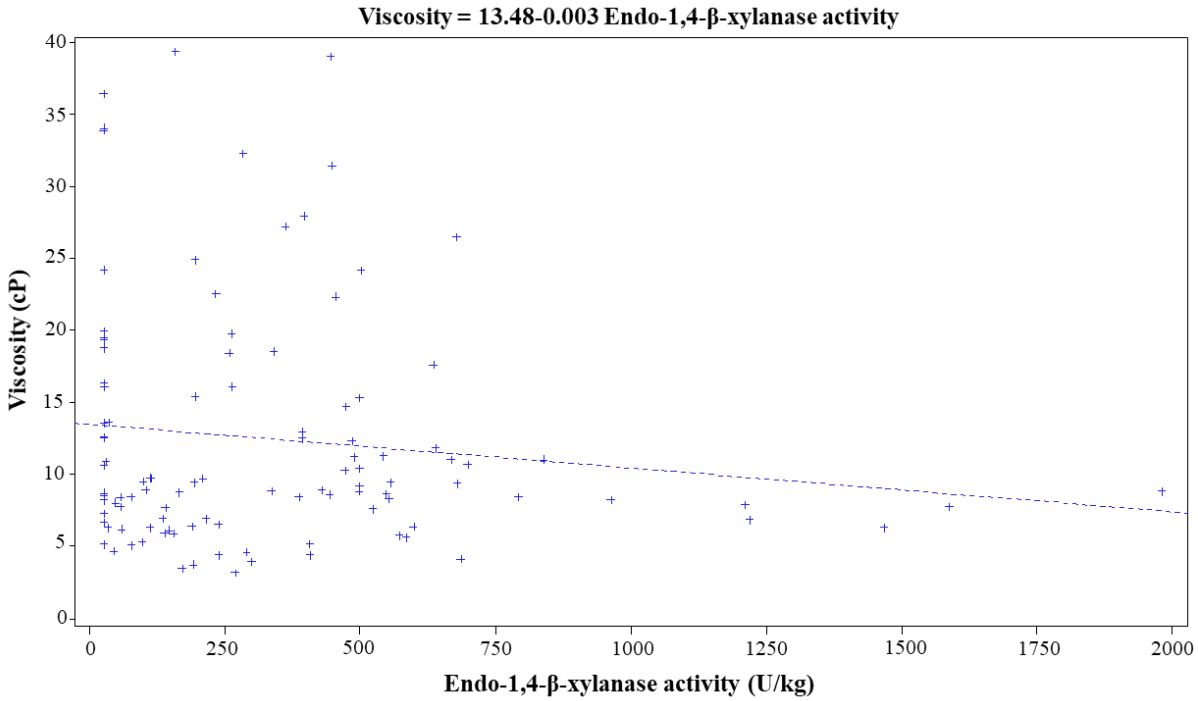


Figure 3.1. Relation between cereal extract viscosity and endogenous endo-1,4-β-xylanase activity in different wheat lots.

It is now well established that a range of factors affect the response of wheat-based diets to enzyme supplementation. However, there is still a paucity of information about the biological mechanisms that limit the effectiveness of the exogenous feed enzymes used in poultry production (Bedford 1997). In recent years, the effects of various factors in the modulation of animal response to enzyme supplementation, such as starch content, Hagberg falling number (Hetland *et al*. 2007), variety (McCracken *et al*. 2001; McCracken *et al*. 2008) and chemical composition (Pirgozliev *et al.* 2003), among others, have been studied. Nevertheless, it seems that the majority of these factors do not display a repeatable and significant effect on wheat quality and nutritive value (Bedford 1997). In addition, a considerable amount of research investigated the correlation between levels of NSPs in different wheat lots/varieties and poultry nutrition (Choct 1997; Williams *et al*. 1997; Bedford 2000). Thus, there is a considerable degree of evidence suggesting that levels of arabinoxylans, the most important NSP in wheat, is one of the most important factors negatively affecting animal performance. Similarly, wheat viscosity, which primarily results from soluble NSP content, was reported to be a major factor affecting wheat nutritive value (Choct and Annison 1990; Bedford and Morgan 1996; Bedford 2000) and has been inversely correlated with animal performance and feed efficiency (Bedford 1996). These observations suggest that variations in levels of NSPs may account for one of the most important factors that modulate response to exogenous enzyme supplementation, being a response to

enzyme supplementation only evident in wheat lots containing a significant level of arabinoxylans.

3.3.2. Levels of endogenous endo-1,4-β-xylanases and wheat extract viscosity affect the response to exogenous enzyme supplementation

Data presented above revealed that there is a considerable variation in levels of wheat endo-1,4-β-xylanase activity and extract viscosity in different cereal lots, suggesting that the two factors may be implicated in the variable poultry response to exogenous enzyme supplementation when fed wheat-based diets. Thus, to test how endogenous endo-1,4-β-xylanase activity and extract viscosity can influence the nutritive value of wheat-based diets for poultry, two wheat lots were selected for a broiler performance trial. These lots were selected from a group of few that were available in sufficient amounts to perform the animal experiment and consisted of a wheat lot presenting HALV and a wheat lot displaying LAHV. The two wheat lots selected correspond to wheats W50 and W53, respectively (Table 3.2). The chemical composition of these two different wheat lots is presented in Table 3.3. The data (Table 3.3) revealed that from a nutrient point of view the two wheats were similar. Thus, to investigate the influence of wheat lot in the response to exogenous enzyme supplementation, two basal diets were prepared using wheat batches HALV (W50) and LAHV (W53), following the formula presented in Table 3.1. The two basal diets were supplemented (+) or nor not supplemented (-) with an endo-1,4-β-xylanase commercial mixture. The four diets were randomly assigned to feed broiler chicks raised from day 1 to 35 under conventional brooding practices.

BW, weight gain, feed intake and feed conversion ratio of birds during the progression of the trial are displayed in Table 3.4. Mortality was low (1.7%) and was not related to treatments. Throughout the trial, there was a significant interaction ($P<0.05$) between wheat lot and exogenous endo-1,4-β-xylanase for BW, such that animals from treatments HALV-, HALV+ and LAHV+ had a higher BW ($P<0.05$) than animals from group LAHV- (Table 3.4). Overall, the data suggest that the commercial enzyme has a positive effect on BW only when birds were fed with the LAHV wheat diet. In contrast, BW of birds fed the HALV wheat diet was not improved by the addition of additional exogenous endo-1,4-β-xylanases (Table 3.4). Similarly, the interaction between wheat lot and exogenous supplementation was significant for broiler weight gain, where birds fed LAHV- displayed lower weight gains when compared with animals fed on the HALV-, HALV+ and LAHV+ diets (Table 3.4). However, the difference was only significant ($P<0.05$) when the entire experimental period is considered as a reflection of differences observed in the first 3 weeks of the trial. In the last 2 weeks of the experiment, there was no effect of wheat lot, exogenous endo-1,4-β-xylanase or the interaction of the two variables, suggesting that weight gain differences mainly occur in

younger birds. In addition, feed intake was consistently lower in the LAHV- treatment, but the interaction was only significant in the first 2 weeks of the experiment ($P<0.05$). Overall, broilers receiving the LAHV diet displayed a decreased feed intake when compared with animals of the HALV groups at the initial stages of the experiment ($P<0.05$), but differences were not significant after day 7 (Table 3.4). No interaction between wheat source and xylanase supplementation was observed for FCR ($P>0.05$). Over the entire experimental period (0-35 days) and during the fourth week of the experiment FCRs were improved ($P<0.05$) by the addition of exogenous endo-1,4-β-xylanase to the diet (Table 3.4). This suggests that addition of the exogenous enzyme had a positive impact in FCR irrespective of the wheat lot.

Table 3.3. Nutritional composition of the two different wheat lots used in the animal experiment.

Wheat HALV displays high endo-1,4-β-xylanase activity and low viscosity (Lot W50) while wheat LAHV displays low endo-1,4-β-xylanase activity and high viscosity (Lot W53).

Chemical Composition (%)	Wheat lot [a]	
	HALV (W50)	LAHV (W53)
Dry Matter	86.8	87.1
Crude protein	12.2	12.1
Ether extract	2.70	2.70
Ash	1.90	2.10
NDF	16.0	16.9
ADF	3.20	3.41
ADL	0.91	1.20

Taken together, the results suggest that exogenous microbial endo-1,4-β-xylanases were mainly effective in improving the nutritive value of the LAHV diet, suggesting that the higher levels of endogenous endo-1,4-β-xylanases combined with the lower viscosity of the HALV wheat lot resulted in a higher nutritive value of this diet per se. However, it remains to be established which factor was more responsible for the lower response to enzyme supplementation when enzymes were used to supplement wheat lot HALV, i.e., if it is the higher levels of endogenous enzymes or the lower viscosity presented by this cereal grain. Nevertheless, addition of exogenous enzymes to the two wheat-based diets leads to a significant improvement of FCR, suggesting that in general the exogenous endo-1,4-β-xylanases are effective but to a lower degree in HALV diets.

Table 3.4. Performance of broilers fed with diets prepared with wheat.

HALV (high endo-1,4-β-xylanase activity and low viscosity) or LAHV (low endo-1,4-β-xylanase activity and high viscosity) not supplemented (-) or supplemented (+) with a commercial endo-1,4-β-xylanase mixture.

	HALV		LAHV		SEM	Significance		
	-	+	-	+		W	E	W*E
Body Weight (g)								
0d	46.4	45.7	45.8	45.1	0.04	0.14	0.08	0.91
7d	161^{a}	156^{a}	145^{b}	156^{a}	1.9	<0.01	0.13	<0.01
14d	430^{a}	423^{a}	387^{b}	421^{a}	7.3	0.01	0.08	<0.01
21d	884^{a}	869^{ab}	812^{b}	876^{ab}	17.9	0.08	0.19	0.04
28d	1538^{a}	1536^{a}	1401^{b}	1537^{a}	26.6	0.02	0.02	0.02
35d	2226^{a}	2222^{a}	2073^{b}	2242^{a}	32.1	0.05	0.02	0.01
Weight Gain (g)								
0-7d	115^{a}	110^{a}	99^{b}	111^{a}	1.7	<0.01	0.04	<0.01
7-14d	269^{a}	268^{a}	242^{b}	265^{a}	4.2	<0.01	0.01	0.01
14-21d	454^{a}	449^{ab}	425^{b}	455^{a}	7.8	0.15	0.12	0.02
21-28d	614	615	596	623	9.7	0.61	0.18	0.17
28-35d	669	666	676	683	12.0	0.31	0.88	0.69
0-35d	2123^{ab}	2110^{ab}	2035^{b}	2139^{a}	26.8	0.27	0.09	0.03
Feed Intake (g)								
0-7d	161^{a}	150^{ab}	139^{b}	158^{a}	4.0	0.11	0.33	0.01
7-14d	407^{a}	398^{ab}	363^{b}	393^{ab}	8.9	0.01	0.24	0.04
14-21d	682	675	638	664	10.1	0.01	0.36	0.12
21-28d	980	950	943	960	19.3	0.48	0.74	0.24
28-35d	1206	1161	1188	1182	23.6	0.94	0.30	0.41
0-35d	3436	3334	3271	3357	49.3	0.17	0.88	0.07
Feed Conversion Ratio (g/g)								
0-7d	1.40	1.36	1.41	1.42	0.02	0.15	0.66	0.24
7-14d	1.52	1.49	1.50	1.49	0.04	0.81	0.65	0.78
14-21d	1.50	1.53	1.50	1.46	0.03	0.34	0.77	0.34
21-28d	1.61	1.55	1.58	1.54	0.02	0.43	0.01	0.61
28-35d	1.79	1.72	1.77	1.73	0.03	0.72	0.07	0.59
0-35d	1.62	1.58	1.61	1.57	0.01	0.38	0.01	1.00

Means in the same line with different letter superscripts (a, b, c) are significantly different (P<0.05) or tend to be significantly different (P<0.1).

Previous studies revealed that cereal variety has a direct effect in the effectiveness of an enzyme preparation used to supplement wheat-based diets for broilers (Dusel *et al*. 1998) Several factors may modulate the nutritive value of wheat for broilers. It is widely known that variability in wheat NSP levels influences wheat nutritive value, with high NSP levels yielding higher endo-1,4-β-xylanase responses in terms of animal performance parameters (Bedford 1997; Bedford and Schulze 1998; Bedford 2000; Bedford 2006). Levels of soluble NSPs are directly related with wheat extract viscosities, as digesta viscosity results from the solubilization of the antinutritive NSPs during digestion. Thus, results from the present experiment are not entirely surprisingly as it was expected that a high viscosity wheat would lead to an impaired animal performance (Dusel *et al.* 1998). However, the data suggest that levels of endogenous endo-1,4-β-xylanase may also affect the degree of performance response to exogenous enzyme supplementation as endogenous enzymes may be present in sufficient levels to degrade the antinutritive arabinoxylans. Nevertheless, diet supplementation with exogenous endo-1,4-β-xylanases may be predominantly effective only when wheats used to prepare the broiler feed present high viscosities, regardless of their endo-1,4-β-xylanase content. This hypothesis remains to be investigated and would involve the use of wheat lots displaying high endo-1,4-β-xylanase activity while expressing low or high extract viscosities.

3.3.3. Diet impact on organ size, digesta viscosity and endo-1,4-β-xylanase activity

Relative weight of gizzard and ileum decreased ($P<0.05$) with the addition of endo-1,4-β-xylanase to the two diets (Table 3.5), but no interaction with wheat source was observed ($P>0.05$). On the other hand, birds receiving LAHV diets displayed smaller livers. None of the main effects nor the interaction between them influenced the organ relative length (Table 3.5). Several authors have reported that the addition of plant cell wall degrading enzymes decreases digesta viscosity and therefore improves feed passage rate and nutrient absorption, which may lead to a decrease in the relative weight of the digestive tract and a possible improvement in carcass yield (Pettersson and Åman 1993; Viveros *et al*. 1994; Ribeiro *et al*. 2011). In the current study, a decrease in the digestive organ size was observed following endo-1,4-β-xylanase supplementation but only in the gizzard and ileum.

The main effect of wheat variety did not influence the viscosity of digestive contents. Nevertheless, exogenous endo-1,4-β-xylanase supplementation markedly decreased ($P<0.01$) hindgut content viscosity. Thus, the results suggest that exogenous endo-1,4-β-xylanases effectively contribute to reduce foregut and hindgut viscosities, most probably by contributing to a significant deconstruction of wheat arabinoxylans. However, endogenous endo-1,4-β-xylanases present in wheat can also contribute to depolymerize feed arabinoxylans but probably in a lower degree. These results suggest that small reductions in

the degree of polymerization of the NSPs can have a great impact on digesta viscosity and consequently in animals productive performance, as already suggested by other authors (Choct and Annison 1990; Bedford and Morgan 1996). When exogenous endo-1,4-β-xylanases were added to the LAHV diet, there was a marked decrease in ileum viscosity that was not seen when endo-1,4-β-xylanases were added to the HALV diet, which seems to reveal that endo-1,4-β-xylanases are active and remain stable under the different conditions of the GI tract.

Table 3.5. Relative weight and length of the gastrointestinal tract and viscosity of digesta. Samples of broilers fed with diets prepared with wheat HALV (high endo-1,4-β-xylanase activity and low viscosity) or wheat LAHV (low endo-1,4-β-xylanase activity and high viscosity) not supplemented (-) or supplemented (+) with a commercial endo-1,4-β-xylanase mixture.

	HALV		LAHV		SEM	Significance		
	-	+	-	+		W	E	W*E
Relative weight (g/kg)								
Crop	2.47	2.32	2.17	2.35	0.14	0.32	0.92	0.23
Gizzard	12.21	10.78	11.42	11.07	0.41	0.54	0.04	0.19
Liver	27.24	30.28	25.76	26.67	1.12	0.03	0.09	0.35
Duodenum	5.79	5.52	6.28	5.61	0.27	0.30	0.09	0.47
Jejunum	11.37	11.02	12.08	11.22	0.38	0.24	0.12	0.50
Ileum	8.58^{ab}	8.13^{b}	9.42^{a}	8.33^{ab}	0.29	0.08	0.01	0.28
Caecum	2.02	1.94	2.10	2.01	0.12	0.55	0.48	0.95
Relative length (cm/kg)								
Duodenum	12.22	12.10	12.82	12.87	0.38	0.08	0.92	0.82
Jejunum	30.65	30.91	32.35	31.44	0.81	0.18	0.69	0.48
Ileum	30.80	31.62	32.64	30.73	0.68	0.48	0.43	0.05
Caecum	8.37	7.77	8.31	8.42	0.35	0.40	0.49	0.31
Content Viscosity (cP)								
Duodenum+ jejunum	5.96	3.54	6.01	4.09	0.36	0.40	<0.01	0.48
Ileum	7.42^{ab}	6.61^{b}	9.74^{a}	5.48^{b}	0.78	0.42	<0.01	0.02

Means in the same row with different superscripts (a,b) are statistically different (P<0.05) or tend to be statistically different (P<0.1).

Endo-1,4-β-xylanase activity was detected in the crop and caecum of broilers representative of all treatments. In contrast, enzyme activity was detected in the gizzard, duodenum and jejunum-ileum of all birds that were fed the diets supplemented with

exogenous endo-1,4-β-xylanase (Table 3.6 and Figure 3.2), while endo-1,4-β-xylanase activity was only detected in the GI compartments of a limited number of animals of the non-supplemented groups. Thus, results of the present study revealed endo-1,4-β-xylanase activity exclusively in all the GI compartments when endo-1,4-β-xylanase was supplemented to the diets, suggesting that endogenous endo-1,4-β-xylanase activity of wheat is not detectable using the arabinoxylan-agar plate assay. This is in accordance with a previous study (Ribeiro *et al.* 2011), which indicated that the use of barley with high β-glucanase activity led to improvement of animal performance despite the absence or low levels of catalytic activity in the GIT compartments. The presence of endogenous plant cell wall hydrolases in cereals is important for the mobilization of reserves from storage organs and seeds during germination, and also for plant cell wall restructuration during growth and in response to microbial attack (McLauchlan *et al.* 1999). Although a great part of endo-1,4-β-xylanases may be inactivated by wheat endo-1,4-β-xylanase inhibitors proteins, the high levels at which these enzymes are present in some cereal lots are believed to positively influence animal performance (Courtin *et al*. 2005; Dornez *et al*. 2009).

Table 3.6. Number of birds with enzyme activity in GI tract.

Out of six animals analysed per treatment, fed with diets prepared with wheat HALV (high endo-1,4-β-xylanase activity and low viscosity) or wheat LAHV (low endo-1,4-β-xylanase activity and high viscosity) not supplemented (-) or supplemented (+) with a commercial endo-1,4-β-xylanase mixture and which present endo-1,4-β-xylanase activity in digesta samples.

	HALV		LAHV			
	-	+	-	+	Chi-square	p(F)
Crop	3	5	4	4	1.50	0.682
Gizzard	0	4	2	4	7.54	0.057
Duodenum	1^x	6^y	2^x	6^y	14.8	0.002
Jejunum	1^x	6^y	1^x	6^y	17.1	0.001
Ileum	1^x	6^y	0^x	6^y	20.6	0.001
Caecum	6	6	6	6	-	-

Values in the same line with different letter superscripts (x, y) are significantly different (P<0.05).

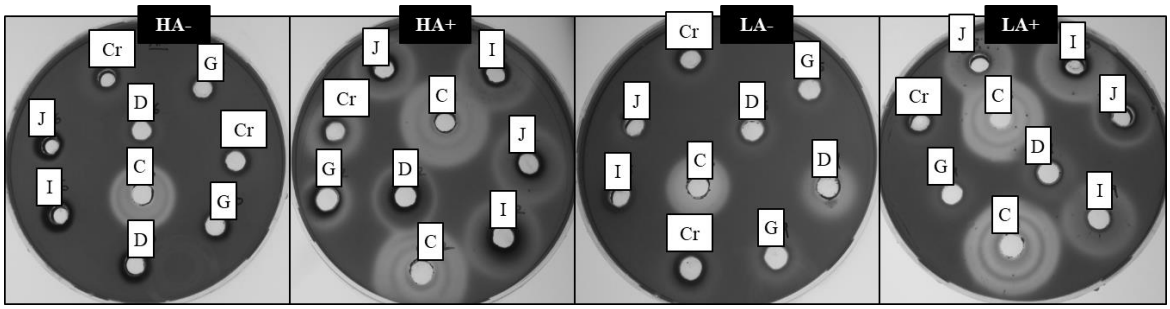


Figure 3.2. Detection of endo-xylanase activity in the gastrointestinal contents of broilers. Treatments consisted of a diet based in wheat with high endo-1,4-β-xylanase activity and low viscosity (HA), with (+) or without (-) xylanase supplementation and a diet based in wheat with low endo-1,4-β-xylanase activity and high viscosity (LA), with (+) or without (-) xylanase supplementation. Cr=crop; G=gizzard; D=duodenum; J=jejunum; I=ileum; Cc=caecum.

3.4. Conclusions

Overall, results presented here revealed that levels of wheat extract viscosity and endo-1,4-β-xylanase, although not correlated, vary widely in different wheat lots. Exogenous enzyme supplementation improved the nutritive value of the diet incorporating the wheat lot displaying higher extract viscosity (14.8 cp) and lower endogenous endo-1,4-β-xylanase activity (347.0 U/Kg). In contrast, when the wheat lot presented lower viscosity (6.9 cp) and higher levels of endogenous xylanase activity (544.0 U/Kg), broiler response to feed supplementation with exogenous microbial xylanases was clearly diminished. Notwithstanding the fact that variation of endogenous endo-1,4-β-xylanase activity in the two wheat lots selected for the animal trial described here was marginal, in particular considering the large variation in endogenous xylanase activity observed in the 108 wheat lots tested, it is clear that, as observed previously for endogenous β-glucanase activity in barley (Ribeiro *et al.* 2011), endogenous xylanase activity may influence response to exogenous enzyme supplementation. Thus, wheat lots presenting higher levels of endogenous xylanase activity may be less responsive to the supplementation with exogenous enzymes. However, it remains to be investigated which factor, endogenous endo-1,4-β-xylanase activity or intrinsic extract viscosity, influences mostly the effective response to exogenous enzyme supplementation of broilers fed on wheat-based diets.

4. XYLO-OLIGOSACCHARIDES DISPLAY A PREBIOTIC ACTIVITY WHEN USED TO SUPPLEMENT WHEAT OR CORN-BASED DIETS FOR BROILERS

T. Ribeiro[1,*], V. Cardoso[1,§,*], L.M.A. Ferreira [1], M.M. Lordelo[2], E. Coelho[3], A.S.P Moreira[3], M.R.M. Domingues[3], M.A. Coimbra[3], M.R. Bedford[4,†] and C.M.G.A. Fontes[1,†,§]

[1] CIISA – Faculdade de Medicina Veterinária, Universidade de Lisboa, Pólo Universitário do Alto da Ajuda, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; §NZYTech – genes & enzymes – Estrada do Paço do Lumiar, Campus do Lumiar, Edif. E, R/C, 1649-038 Lisboa, Portugal; [2] Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal; [3] QOPNA, Departamento de Química, Universidade de Aveiro, 3810-193 Aveiro, Portugal; [4] AB Vista, Woodstock Court, Blenheim Rd. Marlborough, UK.

Adapted from: Ribeiro *et al.* (2018) *Poultry Science.* 97(12):4330-4341.

Abstract

It is now well established that exogenous 1,4-β-xylanases improve the nutritive value of wheat-based diets for poultry. Among other factors, the mechanism of action of exogenous enzymes may involve a microbial route resulting from the generation of prebiotic xylo-oligosaccharides in the birds' gastrointestinal tract. In a series of three experiments, the effect of xylo-oligosaccharides (XOS) on the performance of broilers fed wheat or corn-based diets were investigated. In experiment 1, birds receiving diets supplemented with XOS displayed an increased weight gain (P=0.08). The capacity of XOS to improve the performance of animals during a longer trial (42 days) was investigated (Experiment 2). The data revealed that diet supplementation with XOS, tested at two incorporation rates (0.1 and 1 g/kg), or with an exogenous 1,4-β-xylanase resulted in an increased nutritive value of the wheat-based diet. An improvement in animal performance was accompanied by a shift in the microbial populations colonizing the upper portions of the gastrointestinal tract. XOS were also able to improve the performance of broilers fed a corn-based diet, although the effects were not apparent at incorporation rates of 10 g/kg. Together these studies suggest that in some cases the capacity of 1,4-β-xylanases to improve the nutritive value of wheat-based diets is more related to their ability to produce prebiotic XOS than to their ability to degrade arabinoxylans. The extremely low quantities of XOS used in this study also challenge the depiction of a prebiotic being a quantitatively fermented substrate. These data also bring into question the validity of the "cell wall" mechanism, as XOS elicited an effect with clearly no

action on endosperm cell wall integrity and yet the performance effects noted were equivalent or superior to the added enzymes.

4.1. Introduction

Cereal-based diets contain soluble Non-Starch Polysaccharides (NSPs) that, due to its intrinsic viscosity, display significant anti-nutritive properties for poultry (Bedford *et al*. 1991; Bedford and Classen 1992; Chesson 1993; Choct 2006). It is now well established that addition of 1,4-β-xylanases or 1,3-1,4-β-glucanases to wheat- or barley-based diets, respectively, leads to a significant depolymerization of the indigestible NSPs resulting in improved bird performance. There has been considerable debate about the mechanisms by which exogenous enzymes improve the nutritive value of cereal-based diets for poultry. In diets containing a high proportion of soluble NSPs, exogenous enzymes reduce the resulting high digesta viscosity, thus promoting feed intake and the efficacy of endogenous digestive enzymes and leading to an improvement in nutrient digestibility (Bedford and Morgan 1996; Bedford 2000). In low viscosity diets the action of exogenous enzymes has been attributed to their ability to degrade cereal cell walls, thus enabling enhanced access to their contents by digestive enzymes (Bedford 2000). Another, more subtle, action mechanism could involve the gut microbiota route (Apajalahti and Bedford 1999; Fontes *et al*. 2004; Figueiredo *et al*. 2012). In the case of wheat-based diets, which are rich in arabinoxylans (Fengler and Marquardt 1988), the resulting effect of exogenous 1,4-β-xylanases would be the generation of a range of xylo-oligosaccharides (XOS) (Biely *et al*. 1997). Such oligosaccharides would have a beneficial effect on the microflora colonizing the distal portion of the gastrointestinal (GI) tract (Apajalahti and Bedford 1999). Thus, the beneficial effects resulting from inclusion of 1,4-β-xylanases in wheat-based diets could result from the production of XOS in addition to the direct activity on the soluble, viscous arabinoxylans.

XOS are xylose oligomers with a number of residues that can vary from 2 to 15, or even 20. It has been suggested that XOS may improve gut health and stimulate the animals´ immune response, thus functioning as typical prebiotics (De Maesschalck *et al.* 2015). In general, prebiotics stimulate selectively the growth and/or activity of those gut bacteria that contribute to health and well-being (Gibson and Roberfroid 1995; Grizard and Barthomeuf 1999; Gibson *et al*. 2004). Prebiotics are not degraded or absorbed in the small intestine and are thus available for fermentation in the lower gut by bacterial species adept at using such carbohydrate sources. This typically shifts the composition of the intestinal microbiome towards a relative increase in Bifidobacterium and/or Lactobacillus species and a decrease in pathogenic bacteria. In addition, prebiotics may stimulate production of anti-bacterial metabolites such as lactic acid, volatile fatty acids, and bacteriocins (Ššković *et al*. 2001). The presence of SCFA in the intestines contributes to a lower pH, a better bio-availability of

calcium and magnesium, and the inhibition of potentially harmful bacteria (Teitelbaum and Walker 2002; Wong *et al*. 2006). SCFA are further used by the host organism as an energy source. A variety of compounds have been proposed to act as prebiotic compounds. These include inulin, fructo-oligosaccharides, galacto-oligosaccharides, isomalto-oligosaccharides, amylase-resistant gluco-oligosaccharides, lactosucrose, mannan-oligosaccharides (MOS), and soybean oligosaccharides(Monsan and Paul 1995; Orban *et al*. 1997; Patterson *et al*. 1997; Piva 1998; Collins and Gibson 1999; Macfarlane *et al*. 2006; Baurhoo, Letellier, *et al*. 2007; Baurhoo, Phillip, *et al*. 2007).

Several studies have demonstrated that different *Bifidobacterium* species, when grown in pure cultures, can efficiently utilize XOS (Okazaki *et al*. 1990; Yamada *et al*. 1993; Jaskari *et al*. 1998; van Laere *et al*. 2000; Crittenden *et al*. 2002; Palframan *et al*. 2003; Moura *et al*. 2007). Some strains of Bifidobacterium species grow more efficiently on XOS than on xylose (Okazaki *et al*. 1990; Crittenden *et al*. 2002; Palframan *et al*. 2003). This suggests that Bifidobacterium species possess a specific oligosaccharide uptake mechanism for XOS (Palframan *et al.* 2003), which may provide a competitive advantage over strains that can only take up monosaccharides. Less efficiently, XOS can be utilized by some *Bacteroides* species and Lactobacillus species, but not by a range of other intestinal bacteria including *Escherichia coli* and *Clostridia* (Yamada *et al*. 1993; van Laere *et al*. 2000; Moura *et al*. 2007). The increase in the levels of colonic or caecal *Bifidobacterium* spp has been reported in several studies using XOS on rats or on mice (J.M. Campbell *et al*. 1997; Hsu *et al*. 2004; Santos *et al*. 2006). In these studies, increases in *Bifidobacterium* spp levels were also accompanied by significant decreases in levels of Enterobacteriaceae and/or sulfite-reducing bacteria. Here, in a set of three experiments, we have investigated the prebiotic activity of XOS. The data suggest that, per se, XOS can contribute to improve the nutritive value of both wheat- and corn-based diets for poultry. The implication of the data in view of the putative role of 1,4-β-xylanases used to supplement cereal-based diets for poultry is discussed.

4.2. Materials and Methods

4.2.1. Enzymes and oligosaccharides

A series of three experiments were performed to probe the capacity of 1,4-β-xylanases and xylo-oligosaccharides (XOS) to improve the nutritive value of wheat and corn-based diets for broilers. The enzyme selected for these studies was DuPont/Danisco Xylanase 40000G® (Marlborough, United Kingdom), a dried *Trichoderma reesei* fermentation product displaying a minimum xylanase activity of 40000 U/g. The XOS preparation (XOS95P®) and purified D-xylose used to supplement the poultry diets were acquired from Shandong Longlive Bio-technology Corporation, China.

4.2.2. Oligosaccharide and monosaccharide composition analysis

Monosaccharides were released from XOS by hydrolysis with 1 M H_2SO_4 for 2.5 h at 100 °C. Neutral sugars were analyzed as their alditol acetates by gas-chromatography-flame ionization detection (Coelho *et al*. 2016). The hydrolysis was performed in duplicate. For the determination of free sugars composition, the samples were not submitted to the acid hydrolysis step prior to the derivatization to alditol acetates. The XOS profile was determined by reduction and acetylation of the sample as performed for monosaccharides. The reduced and acetylated XOS were separated and analysed by gas chromatography–mass spectrometry (GC–MS) on an Agilent Technologies 6890N Network. The GC was equipped with a 400-1HT, dimethylpolysiloxan capillary column (25 m length, 0.22 mm of internal diameter, and 0.05 µm of film thickness). The GC was connected to an Agilent 5973 mass quadrupole selective detector operating with an electron impact mode at 10 and 70 eV, scanning the *m/z* range 50–800 at 2 scans s^{-1}, in a full scan mode acquisition (Simões *et al*. 2011).

4.2.3. Determination of the presence of starch

In order to evaluate the presence of starch in XOS mixture, 19 mg of sample were treated with α-amylase from *Bacillus subtilis* (EC 3.2.1.1, 62 U/mg, Sigma-Aldrich; St. Louis, United States) at 37 °C overnight, at pH 7. Amylose and cellulose were used as positive and negative controls, respectively. The sugars released by enzymatic hydrolysis were analysed after the derivatization to alditol acetates.

4.2.4. Oligosaccharide ethanol fractionation and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry analysis

The XOS sample was suspended in absolute ethanol and the ethanol insoluble material was separated from the supernatant by centrifugation, giving origin to an ethanol soluble fraction and an ethanol insoluble fraction (Coelho *et al.* 2014). The ethanol insoluble fraction was dissolved in water and absolute ethanol was added until reaching 50 % (v/v) of ethanol, allowing separation by centrifugation of the material insoluble in 50% ethanol (Et50). The material remaining dissolved in the 50% ethanol solution was then precipitated adding more ethanol, giving origin to fractions Et75 and Et80. The material remaining soluble in the ethanol 80% solution was named EtSn. The fractions Et50 and Et75 were pooled, suspended in water, and re-precipitated in 50% ethanol. The XOS sample and the precipitated ethanol fractions were dissolved in ultrapure water (1 mg/mL) and 10 µL of each sample was added to 10 µL of 2,5-dihydroxybenzoic acid (DHB) matrix (15 mg/mL in methanol). From this mixture, 0.5 µL were deposited on top of a layer of crystals formed by deposition of 0.5 µL of DHB solution on the MALDI plate and letting it dry at ambient

conditions. MALDI-MS spectra were acquired using a MALDI-TOF/TOF Applied Biosystems 4800 Proteomics Analyzer (Applied Biosystems, Framingham, MA) instrument equipped with a nitrogen laser emitting at 337 nm and operating in a reflectron mode. Full-scan mass spectra ranging from m/z 500 to 4000 were acquired in the positive mode.

4.2.5. Diets

Table 4.1. Ingredient composition and calculated analysis of the wheat-based feed of experiments 1 and 2.

Ingredients	**%**
Wheat	60.00
Soybean meal 47%	29.00
Corn	4.17
Soybean oil	2.75
Sodium chloride	0.30
Calcium carbonate	1.43
Dicalcium phosphate 18%	1.80
Lysine	0.15
DL-Methionine	0.20
Mineral and vitamin premix [1]	0.20
Calculated nutrient content	
Energy (Kcal/Kg)	3009
Crude Protein (%)	19.90
Ether extract (%)	4.70
Crude cellulose (%)	2.93
Ash (%)	6.55

[1] Mineral-vitamin premix provided the following per kilogram of diet: vitamin A, 9,000 IU; vitamin D_3, 2,100 IU; vitamin E, 20 mg; nicotinic acid, 30 mg; vitamin B_{12}, 0.12 mg; calcium pantothenate, 10 mg; vitamin K_3, 2 mg; thiamin, 1 mg; riboflavin, 4.2 mg; vitamin B_6, 1.7 mg; folic acid, 0.5 mg; biotin, 0.5 mg; Fe, 80 mg; Cu, 10 mg; Mn, 100mg; Zn, 80 mg; Co, 0.2 mg; I, 1.0 mg; Se, 0.3 mg; monensin, 100 ppm.

The composition of the basal wheat (Experiments 1 and 2) and corn (Experiment 3) based diets used in this study, which were formulated to contain adequate nutrient levels as defined by the NRC (1994), is presented in Table 4.1 and Table 4.5, respectively. The wheat used to prepare diets used in experiment 1 and 2 were from different sources. In experiment 1 the wheat basal diet was provided with no supplement (treatment C-), or with the xylanase mixture (0.0625 g/kg incorporation, treatment XYL), the XOS mixture (0.06 g/kg incorporation, treatment XOS) or xylose (0.06 g/kg incorporation, treatment OSE). In

experiment 2 the wheat basal diet was also provided with no supplement (treatment C-), with the xylanase mixture (0.0625 g/kg incorporation, treatment XYL) or the XOS mixture at two levels of incorporation, 0.1 g/kg (treatment XOS1) or 1 g/kg (treatment XOS2). In experiment 3 the corn basal diet was provided with no supplement (treatment C-) or was supplemented with the XOS mixture at three levels of incorporation, 0.1 g/kg (treatment XOS1), 1 g/kg (treatment XOS2) or 10 g/kg (treatment XOS3). Diets were provided in the pelleted form after pelletizing at approximately 90 °C. The mineral and vitamin premix was previously mixed with the enzyme/XOS preparations and used to supplement the feed just before administration to the animals.

4.2.6. Animals and Management

For each of the three experiments 360 1-d-old Ross 308 male broiler birds were individually weighed at the beginning of the experiment and were randomly divided into 24 pens of 15 birds each. Chicks were wing-banded for individual identification. The 24 pens were randomly allocated to each of the four dietary treatments, leading to 6 pens per treatment. Throughout the experiment, chicks were given free access to water and feed, which were provided with drinking nipples and hanging feeders, respectively. Broilers were raised in floor pens that were located in an environmentally controlled room adjusted daily to the recommended temperatures and ventilation rates, according to standard brooding practices, and were exposed to constant light for the duration of the trial. While experiment 1 lasted for only 28 days, in experiments 2 and 3 animals were raised until 42 days. Weekly feed consumption and individual body weights were recorded. Feed conversion ratios were calculated by dividing the weight gain per pen, per week and at the end of the experiment, including the weight gain of any dead birds, by the total feed consumed during the respective period. Bird mortality was recorded daily. At the end of the experiments, when animals were 28 (Experiment 1) or 42 (Experiments 2 and 3) days of age, two birds per pen were slaughtered by cervical dislocation. The size of the various empty GI compartments were measured or weighed and digesta samples were collected and stored at –20 °C for subsequent analysis. Levels of β-glucanase activity in the GI tract were measured and microbial gDNA extracted from caecum samples as described below. Animal experiments were conducted in accordance with the Ethics Committee of CIISA, Faculdade de Medicina Veterinária, and approved by the Animal Care Committee of the National Veterinary Authority (Direção Geral de Veterinária, Lisboa, Portugal), following the appropriate European Union guidelines (Council Directive 2010/63/UE).

4.2.7. Analytical Procedures

Digesta samples were centrifuged and the supernatant recovered and frozen at -80 °C for subsequent analysis of 1,4-β-xylanase activity. Initially, qualitative analysis of 1,4-β-xylanase activity in the digesta samples recovered from the various GI compartments was assessed in agar plates, using wheat arabinoxylan (Megazyme, Ireland) at 0.1% (w/v) final concentration, in 10 mM Tris-HCl pH 7.0. Catalytic activity was detected after 16 h of incubation at 37 °C through the Congo Red assay plate, as described by Ponte *et al.* (2004) and Mourão *et al.* (2006). For measuring the viscosity of the small intestine contents, samples were centrifuged for 10 min at 7,500 × g and the viscosity of the supernatant was measured using a Brookfield viscometer (Model LVDVCP-II, Brookfield Engineering Laboratories, Middleboro, MA) with a cup maintained at 24 °C. Analyses for dry matter (DM; method 934.01), crude fat (920.39), crude protein (954.01), NDF (2002.04) and ADF/ADL (973.18) were performed according to the methods specified by Association of Official Analytical Chemists (1980).

4.2.8. Microbiome profiling through 16s rRNA sequencing

Caecum samples of three animals from each treatment (12 extractions in total) were subjected to genomic DNA extraction through Qiagen™ DNA extraction kit for stool samples, following the manufacturer instructions. DNA from caecum samples were then amplified for the region V3-V4 of the 16s rRNA as described previously (Vaz-Moreira *et al*. 2013), using the following primers: V3_titF ACTCCTACGGGAGGCAG and V4_titR TACNVRRGTHTCTAATYC. Resulting PCR products were sequenced at Biocant (Cantanhede, Portugal), in the 454 sequencing platform (Roche), according to manufacturers' instructions. Data analysis was carried out as previously described (Pinto *et al.* 2014) at Biocant (Cantanhede, Portugal). The raw pyrosequencing reads were assigned to the appropriate samples based on the respective barcode. Reads were quality filtered by elimination of sequence reads with <120 bp and that contained more than two undetermined nucleotides. Sequences in which the reverse primer was reached were additionally cut. Finally, sequences with more than 50% of low complexity regions, determined by DustMasker (Sogin *et al.* 2006), and chimera sequences, identified by UChime (Edgar *et al*. 2011), were discarded. The sequences were grouped by USearch (Edgar 2010) according to a phylogenetic distance of 3%, creating the Operational Taxonomic Units (OTU). Richness of population (rarefaction curves) and the diversity indices (Chao1) were calculated using the Mothur package (Schloss *et al.* 2009). The taxonomy of each OTU was identified through a BLAST search against the nt@ncbi/SILVA database. The best hits were selected and subjected to further quality control. All sequences with an alignment of less than 40% as well as those with an E-value greater than $1\times e^{-50}$ were rejected. Additionally, a bootstrap test was

applied to the OTUs to identify the least common taxonomy level. Only the sequences with a bootstrap greater than 70% after 100 replicates, as obtained by seqBoot from Phylip package (Felsenstein 1989), were kept. The taxonomic assignment of the OTUs was completed with the attribution of the NCBI taxonomy identification number, which allowed the complete taxonomy construction of all identified organisms. Finally, for each taxon identified in the sample, the total number of sequences was summed up, providing the abundance of all identified organisms, for population statistics analysis.

4.2.9. Statistical Analysis

Statistical analysis of data concerning birds' performance was conducted by analysis of variance, using the General Linear Models procedure of SAS (SAS Inst. Inc., Cary, NC). Means with a significant F ratio were separated by the least significant difference test. The experimental unit was the pen of fifteen birds (n=6 per treatment). Unless otherwise stated, differences were considered significant when $P<0.05$. A Principal Component Analysis (PCA) was established using the microbiota data following standard procedures (Edgar 2010).

4.3. Results and discussion

4.3.1. Characterization of xylo-oligosaccharides (XOS)

The XOS sample used to supplement the diets included 85% of xylose residues, determined as alditol acetates after hydrolysis, of which 5% were monosaccharides. The analysis of XOS by GC-MS showed that xylobiose was the most abundant oligosaccharide (55-58%), followed by xylotetraose (26-33%) and xylotriose (12-16%). These results are in accordance with the high solubility of the sample in ethanol solutions, as the material that remained soluble in 80% ethanol accounted for 90% of the XOS. In order to characterize the 10% of XOS material with lower solubility in ethanol, which should present higher degree of polymerization, a MALDI-MS analysis was performed. The MALDI spectrum of the XOS mixture (Figure 4.1a) showed sodium adduct ions ($[M+Na]^+$) attributed to pentose oligosaccharides ($Pent_n$), from n= 4 to 12. These data, together with GC-MS data suggest the presence of XOS in the sample with degree of polymerization from 2 to 12. Also, MALDI spectrum of XOS mixture showed $[M+Na]^+$ ions assigned to hexose oligosaccharides (Hex_n), from n=3 to 13. In order to evaluate if the glucose present came from starch, the XOS mixture was submitted to an extensive hydrolysis with α-amylase, releasing only 24% of the initial glucose that accounts for residual starch. MALDI-MS spectrum of fraction Et50 (Figure 4.1b) showed the presence of $Pent_3Hex$, $Pent_2Hex_2$, $Pent_4Hex$, and $Pent_6Hex_m$ from m=1 to 8. In addition, XOS present in Et50 showed polymerized structures until 19 pentoses (Figure 4.1b), which allows to conclude for the occurrence of hexoses linked to XOS. These glucose residues, not affected by α-amylase treatment, should not contribute to any glycogenic effect.

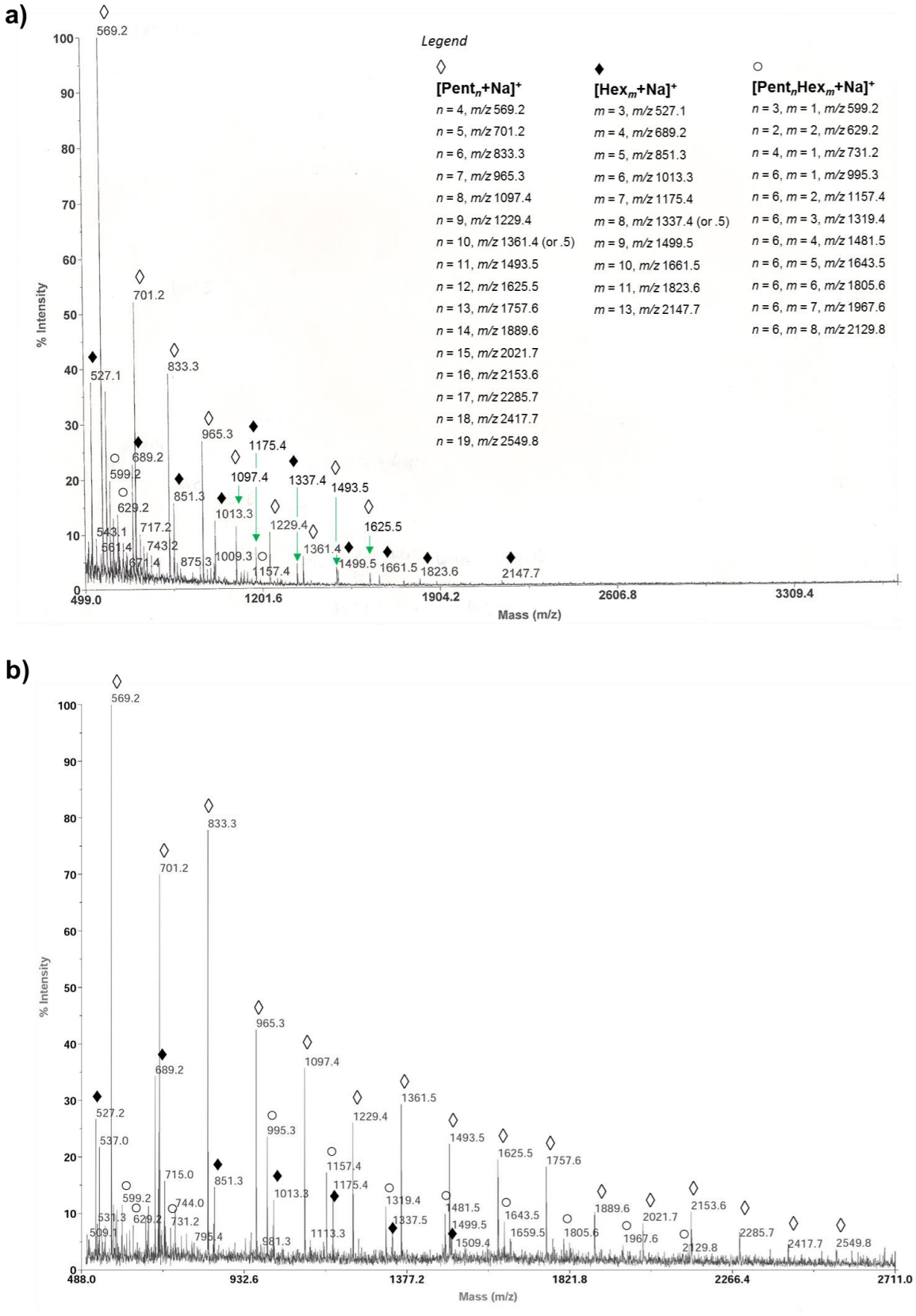


Figure 4.1. MALDI spectra.

a) XOS and **b)** the XOS insoluble in solutions of 50% of ethanol.

4.3.2. Inclusion of exogenous xylanases or XOS in wheat-based diets promote broiler performance

In experiment 1, the capacity of a microbial 1,4-β-xylanase, a XOS preparation and xylose to modulate the performance of broilers grown for 28 days was investigated. The data, presented in Table 4.2, revealed that birds fed on the wheat-based diet supplemented with the XOS mixture reached a higher final body weight (P=0.08) and presented a higher 0-28d weight gain (P=0.036) when compared with non-supplemented animals. Under identical circumstances, both the microbial xylanase and xylose were unable to affect bird's final body weight and weight gain when compared with the negative control group. No differences were found in feed intake although feed conversion ratio of animals receiving the oligosaccharide mixture was significantly lower when compared with the other groups. In addition, birds fed on the four different diets presented no differences in the weight and length of the various GI compartments (Table S4.1). Addition of xylose, XOS or the microbial enzymes had no impact in the viscosity of the intestinal contents, which were low (Table S4.1). Taken together, the results suggest that XOS used to supplement a wheat-based diet have a positive effect on performance, possibly through a prebiotic effect. However, the short duration of the trial (28 days), the majority of which was conducted when the birds still have an under-developed gut microflora, compromises the ability of the prebiotic XOS to promote animal performance. Similar data were reported by Courtin *et al.* (2008) in a study using arabinoxylo-oligosaccharides (AXOS) to supplement wheat and corn-based diets for broilers. In animal trials that lasted 21 days, the authors reported an improvement in feed utilization while body weights were unchanged (Courtin *et al.* 2008). The lack of response to enzyme supplementation observed in the present study was not due to an inactivation of the microbial xylanase through the gastrointestinal tract, as samples collected in the different digestive compartments revealed capacity to degrade arabinoxylans (data not shown). It is possible that the lack of response to the addition of the exogenous enzyme results from the lower levels of NSPs present in the diet, which is supported by the lower viscosity of the digesta (see above). Lower NSPs would result in a limited release of oligosaccharides upon polysaccharide hydrolysis, thus reducing the prebiotic effect. Finally, the end product of xylan hydrolysis, xylose, had no impact on animal performance, suggesting that this sugar has no major effect in gut microbiota.

Table 4.2. Performance of broilers fed on the wheat-based diet (Experiment 1).

Diet supplemented with a 1,4-β-xylanase mixture (XYL), a xylo-oligosaccharide preparation (XOS) and xylose (OSE). A fourth group of birds was fed on a basal non-supplemented diet (C-).

	C-	XYL	XOS	OSE	SEM	p(*F*)
Body Weight (g)						
0d	37,9	37,9	37,9	37,9	0,179	0,991
7d	152	146	149	148	4,278	0,823
14d	406	401	402	397	9,208	0,933
21d	876	881	880	875	15,10	0,991
28d	1508	1503	1558	1483	20,5	0,082
Weight Gain (g)						
0-7d	114	108	112	110	4,296	0,822
7-14d	248	250	243	244	8,416	0,901
14-21d	465	481	480	478	9,958	0,630
21-28d	616	619	631	605	13,82	0,617
0-28d	1470^{b}	1465^{ab}	1534^{a}	1445^{ab}	21,75	0,036
Feed Intake (g)						
0-7d	131	125	125	127	4,598	0,765
7-14d	348	348	350	338	8,027	0,935
14-21d	676	681	693	690	13,31	0,779
21-28d	955	994	955	955	23,33	0,566
0-28d	2081	2143	2123	2110	36,04	0,664
Feed Conversion Ratio						
0-7d	1,15	1,16	1,13	1,16	0,029	0,842
7-14d	1,40	1,41	1,45	1,39	0,033	0,667
14-21d	1,46	1,42	1,45	1,44	0,015	0,284
21-28d	1,56	1,58	1,55	1,44	0,041	0,523
0-28d	1,42^{a}	1,46^{a}	1,39^{b}	1,46^{a}	0,021	0,052

Means in the same line with different letter superscripts (a, b, c) are significantly different (P<0.05) or tend to be significantly different (P<0.1).

A similar experiment was performed using a diet with identical composition to experiment 1 (Table 4.3) but excluding xylose as a supplement. This allowed testing the effect of XOS at two different incorporation rates (0.1 and 1 g/kg). These incorporation rates were higher than those of Experiment 1 when XOS were incorporated at a 0.06 g/kg rate. A different batch of wheat was used to prepare the diets for Experiment 2; all other remaining components of the diet were of the same origin as in Experiment 1 (Table 4.1). The trial lasted for 42 days. Bird body weight, weight gain, feed intake and feed conversion ratio of

Experiment 2 are summarized in Table 4.3. The data revealed that final body weight of birds fed the basal diet supplemented with XOS, at the two levels of incorporation, or the commercial enzyme was significantly higher than those of birds fed on the non-supplemented diet. Differences in body weight were visible as soon as d 7 and remained significant for the duration of the experiment. However, although differences in weight gain were observed in the first two weeks of the experiment, they were particularly acute in the last two weeks of the trial, from days 28 to 42, when caeca fermentation was presumably far more extant than in younger birds. There was no difference in performance between the two levels of XOS incorporation. In addition, all of the supplemented diets resulted in markedly higher feed intake than the control diet in the last week of the trial, although when the entire period of the experiment was considered no differences between groups were observed. Finally, there were no differences in feed conversion ratios although in the initial two weeks of the experiment birds supplemented with both the enzyme and lower dose of XOS converted feed more effectively. This may indicate that the direct result of XOS incorporation is through the improvement of feed intake rather than from an improvement in feed efficiency. Taken together the results revealed that, as it has been extensively described previously, the exogenous xylanase was able to improve the nutritive value of the wheat-based diet of experiment 2. The fact that animals consuming diets supplemented with XOS and the microbial enzyme achieved similar performance suggests that the exogenous biocatalysts may mediate their effects through mechanisms that do not involve a decrease in the concentration of the anti-nutritive arabinoxylans or the release on cell-wall trapped nutrients, but rather involves the generation of XOS that are used as prebiotics by gut microbiota.

It has been extensively observed that diets presenting high levels of soluble NSP induce a considerable enlargement of some portions of the GI tract (Brenes *et al.* 1993) and stimulate an increase in protein turnover rates (Danicke *et al*. 2000). Generally, enzyme supplementation leads to a reduction in the relative weight of the digestive tract compartments, leading to an increase in carcass yield (Van Der Klis *et al*. 1993). Reduction in the size of GI tract of animals receiving exogenous xylanases results from a decrease in digesta viscosity (Fuente *et al*. 1998). Thus, the effects of dietary treatments in the relative length and weight of GI tract compartments of broiler chickens of experiment 2 were evaluated. The data, presented in Table 4.4 suggest that none of supplements had an effect on crop, gizzard, and liver relative weights or in the length of the small and large intestines. The beneficial effects that result from the inclusion of XOS in poultry diets seem to be mediated through the microflora route (see below) and thus it is surprising to observe that the size of the caeca remains unchanged in animals receiving the oligosaccharide supplement. As observed in experiment 1, these data suggest that the wheat lot used in experiment 2 was

not viscous enough to affect gastrointestinal tract dimensions as has been noted in the past (Fontes *et al*. 2004).

Table 4.3. Performance of broilers fed on the wheat-based diet (Experiment 2).

Diet supplemented with a 1,4-β-xylanase mixture (XYL), and a XOS preparation provided at two different incorporation rates, 0.1 g/kg (XOS1) or 1 g/kg (XOS2). A fourth group of birds was fed on a basal non-supplemented diet (C-).

	C-	XOS1	XOS2	XYL	SEM	p(*F*)
Body Weight (g)						
0d	45,5	46,3	46,5	45,5	0,391	0,203
7d	160^{b}	168^{a}	168^{a}	172^{a}	2,163	0,012
14d	419^{b}	450^{a}	441^{a}	451^{a}	6,598	0,004
21d	864^{b}	928^{a}	903^{ab}	916^{a}	15,250	0,029
28d	1526	1614	1566	1602	26,223	0,081
35d	2266^{b}	2388^{a}	2348^{ab}	2370^{a}	30,407	0,032
42d	2854^{b}	3048^{a}	2994^{a}	3021^{a}	35,386	0,003
Weight Gain (g)						
0-7d	115^{b}	121^{ab}	122^{a}	126^{a}	2,136	0,013
7-14d	258^{9b}	283^{a}	273^{ab}	279^{a}	5,275	0,019
14-21d	446	478	462	465	9,579	0,169
21-28d	661	686	663	686	14,663	0,403
28-35d	741	774	782	768	11,476	0,061
35-42d	588^{b}	660^{a}	647^{a}	651^{a}	20,165	0,053
0-42d	2808^{b}	3001^{a}	2948^{a}	2975^{a}	35,291	0,003
Feed Intake (g)						
0-7d	207	206	218	215	5,367	0,111
7-14d	469	439	467	443	17,484	0,298
14-21d	832	797	779	870	29,939	0,132
21-28d	1091	1062	1047	1079	42,602	0,865
28-35d	1370	1328	1355	1341	25,239	0,648
35-42d	1188^{b}	1373^{a}	1364^{a}	1333^{a}	28,213	0,001
0-42d	5156	5204	5230	5280	93,969	0,783
Feed Conversion Ratio						
0-7d	1,80^{a}	1,70^{b}	1,79^{a}	1,70^{b}	0,036	0,050
7-14d	1,82^{a}	1,56^{b}	1,72^{a}	1,59^{b}	0,083	0,033
14-21d	1,88	1,67	1,68	1,87	0,087	0,142
21-28d	1,66	1,55	1,58	1,58	0,076	0,719
28-35d	1,85^{a}	1,71^{b}	1,74^{b}	1,75^{b}	0,031	0,017
35-42d	2,03	2,09	2,11	2,05	0,044	0,461
0-42d	1,84	1,74	1,77	1,78	0,035	0,253

Means in the same line with different letter superscripts (a, b, c) are significantly different (P<0.05) or tend to be significantly different (P<0.1).

Table 4.4. Relative weight and length of the GI tract of broilers (Experiment 2).

Diet supplemented with a 1,4-β-xylanase mixture (XYL), and a XOS preparation provided at two different incorporation rates, 0.1 g/kg (XOS1) or 1 g/kg (XOS2), and birds fed on a non-supplemented diet (C-).

	C-	XOS1	XOS2	XYL	SEM	p(*F*)
Relative Weight (g/100 g BW)						
Crop	2,08	2,34	2,48	2,22	0,138	0,228
Gizzard	9,56	10,00	11,46	10,14	0,904	0,488
Liver	22,86	22,02	21,74	21,56	0,931	0,750
Duodenum	4,32	4,47	4,24	4,34	0,147	0,748
Jejunum	9,30	9,37	9,64	9,46	0,349	0,914
Ileum	7,86	7,92	8,03	7,71	0,324	0,919
Cecum	1,72	1,79	1,77	1,73	0,081	0,934
Relative Length (cm/kg BW)						
Duodenum	10,57	10,11	10,27	10,19	0,296	0,714
Jejunum	26,66	23,56	26,06	25,90	0,976	0,135
Ileum	27,64	25,35	25,95	25,51	0,725	0,114
Cecum	7,08	7,25	6,75	6,77	0,225	0,337

Data presented above suggest that in wheat-based diets exogenous xylanases may influence animal performance through the production of XOS rather than by reducing the concentration of soluble, viscous arabinoxylans or by reducing the integrity of cereal cell walls. XOS resulting from the activity of feed xylanases on arabinoxylans leads to a mixture of unsubstituted and arabinose substituted XOS, usually termed arabinoxylan-oligosaccharides (AXOS). In addition to arabinose side chains, arabinoxylans may be branched with a combination of side groups such as α-D- glucopyranosyl uronic acid or its 4-O-methyl derivative and acetyl groups (Coelho *et al.* 2016). The presence of these side groups in arabinoxylan hydrolysis products may result in branched XOS with potentially diverse biological properties. The data presented here and by other groups (De Maesschalck *et al.* 2015) show that unbranched XOS displayed beneficial effects in bird performance. Here we also reveal that this effect is similar when compared with the complex mixture of putative oligosaccharides that result from the hydrolysis of complex arabinoxylans, after hydrolysis by endo-1,4-β-xylanases. These data suggest that beneficial effect of XOS and AXOS in broiler performance is identical. It is also possible that exogenous xylanases may attack not only the soluble arabinoxylans but also the more recalcitrant and insoluble forms of the polysaccharide thus contributing to the production of large concentrations of XOS

despite the apparently reduced concentrations of soluble arabinoxylans in modern wheat varieties.

4.3.3. XOS modulate gut microbial populations

Data presented above suggest that XOS and AXOS indirectly generated by 1,4-β-xylanases used to supplement wheat-based diets, may modulate an improvement in broiler performance through a prebiotic mechanism. To test this hypothesis, we investigated how 1,4-β-xylanases and XOS modulate gut microflora populations. Thus, bacterial genomic DNA was extracted from caecal samples of twelve animals from each of the 4 treatments of experiment 2 (three animals per treatment) to allow profiling bacterial populations using second generation sequencing technology. Thus, 16S rRNA region V3-V6 was amplified through PCR using caecum nucleic acids as a template and the resulting amplification products were sequenced through a pyrosequencing method. A multivariate analysis using PCA was implemented in order to evaluate differences in bacterial profile. The data, presented in Figure 4.2, revealed that non-supplemented animals were characterized primarily by microorganisms of *Anaerotrumcus, Akkemansia* and *Faecalibacterium* genera. In contrast, samples from the supplemented birds (XOS and xylanase) were characterized by microorganisms of the *Solirubrobacter* and *Bifidobacterium* genera. Regarding the abundance of the microflora, the data revealed that samples from the non-supplemented animals can be distinguished from samples of the supplemented animals. This further suggests an overlap between the XOS and xylanase treatments, suggesting the latter is functioning in large part by generation of XOS/AXOS. Control samples were characterized by *Lactobacillus, Akkermansia, Clostridiales, Faecalibacterium*, among other bacteria, while the supplemented samples were characterized by *Bifidobacterium, Solirubrobacter* and also members from the family *Lactobacillaceae* and *Lachnospiraceae*. Overall the data suggest that XOS are the preferential substrate for the beneficial bacteria, mainly bifidobacteria, contributing to their predominance in the gastrointestinal ecosystem. Nonsubstituted XOS and AXOS are fermented more quickly than more complex structures (methylated glucuronic acid, glucuronic acid and acetyl branching) releasing higher concentrations of lactate (Kabel *et al*. 2002). Also, commercial XOS resulted in greater production of lactate and increased (Zhenping *et al*. 2013) populations (Moniz *et al.* 2016). Taken together, results suggested that XOS and also AXOS resulting from exogenous xylanase activity on arabinoxylans, can modulate the caecal microbiota profile of chickens. These results demonstrate a prebiotic effect of XOS/AXOS which leads to an improvement of animal performance as reflected by higher body weights and weight gains. It is possible that the improvement in the gut microbiota profile also contributes to an improvement in feed intake, thus promoting performance.

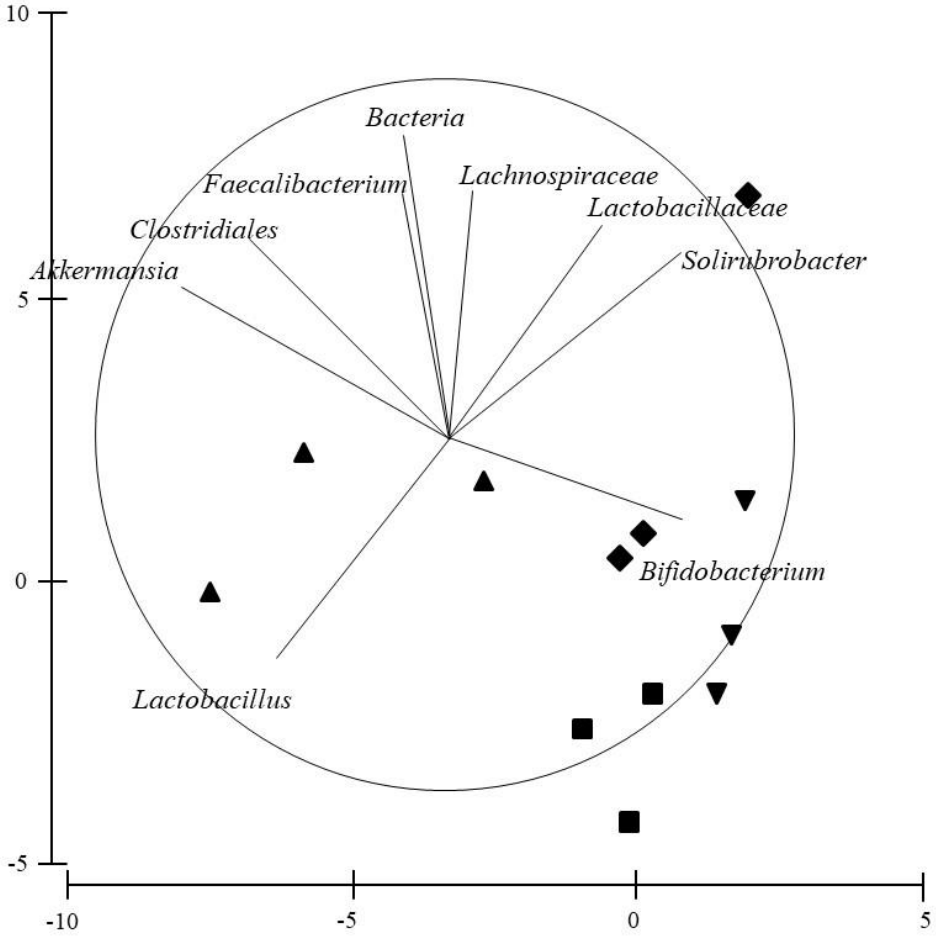


Figure 4.2. Principal Component Analysis (PCA) of bacterial microbiota in the caecum of birds fed on a wheat-based diet.

Wheat-based diet supplemented with a 1,4-β-xylanase mixture (XYL), and xylo-oligosaccharides provided at two different incorporation rates, 0.1 g/kg (XOS1) or 1 g/kg (XOS2). The fourth group of birds was fed on a basal non-supplemented diet (NC). NC (▲), XOS1 (▼), XOS2 (■), XYL (♦).

4.3.4. XOS promote the performance of broilers fed on maize-based diets

In experiment 3 the capacity of XOS to improve the performance of broilers fed a corn-based diet was investigated. A typical corn-based diet was prepared (Table 4.5) and used to feed broiler chickens throughout a 42 day trial. Three levels of XOS incorporation were compared (0.1, 1 or 10 g/kg) with a non-supplemented version of the basal diet. Bird performance in experiment 3, expressed as body weight, weight gain, feed intake and feed conversion ratios are summarized in Table 4.6. From day 14, birds receiving the corn-based feed supplemented with XOS at the lowest incorporation level (0.1 g/kg) presented a body weight higher than the non-supplemented animals. In contrast, animals that were fed on the diet containing the highest incorporation rate of XOS (10 g/kg) presented a body weight similar to birds that were not exposed to the XOS additive. From day seven the body weight

of birds supplemented with the lower levels of XOS was always higher than those of birds receiving XOS at 10 g/kg. Although birds fed on the diets containing the intermediate levels of XOS (1 g/kg) presented a final body weight similar to birds of the XOS1 treatment, at days 28 and 35 they presented a significantly lower weight. Thus, the data suggest that while XOS have a beneficial effect in broilers´ weight, the optimum is only observed at lower incorporation rates (0.1 g/kg); at higher incorporation levels XOS have a marginal or no effect on animal performance. Similar such "overdosing" effects have been demonstrated by Zhenping *et al.* (2013). A similar trend is observed for weight gain, with animals receiving XOS at the lowest incorporation's levels presenting a higher final weight gain than non-supplemented birds or birds fed on the corn diet supplemented with 10 g/kg of XOS. The addition of exogenous XOS resulted in a higher feed intake for birds receiving the oligosaccharide preparation at an intermediate level when compared with non-supplemented animals although there are no differences in the supplemented groups. In contrast, no differences in FCRs were observed between the four treatments. Taken together, the data suggest that XOS are able to improve the performance of broiler chicks although only when incorporated at the lowest rates tested (0.1 to 1 g/kg). Evaluation of body weight and weight gain evolution throughout the trial suggest that the lowest incorporation rate (0.1 g/kg), or perhaps even lower than tested here, might lead to the best results. The effects of the different dietary treatments on the relative length or weight of different organs and GI tract compartments of broiler chickens of experiment 3 were evaluated. The data, presented in Table S4.2, revealed that feed supplementation did not affect GI tract dimensions, as animals receiving the XOS preparation displayed a similar gastrointestinal size as non-supplemented animals.

Data from experiments 1, 2 and 3 suggest that the exogenous XOS used as feed supplements managed to pass through the most aggressive gastrointestinal compartments, evading gastric denaturation and attack by the endogenous repertoire of digestive enzymes which is a property required for an effective prebiotic mixture (Gibson and Roberfroid 1995; Grizard and Barthomeuf 1999; Gibson *et al*. 2004). Overall, previous data describing the biochemical properties of XOS suggest that they indeed remain intact during passage through the GI tract (Okazaki *et al*. 1990; Jaskari *et al*. 1998; Kajihara *et al*. 2000; van Laere *et al*. 2000; Crittenden *et al*. 2002; Palframan *et al*. 2003; Chung *et al*. 2007; Moura *et al*. 2007; Na and Kim 2007). Yamada *et al.* (1993) studied the resistance of XOS to low pHs and found that these molecules were only weakly hydrolysed at pH 2.0 at 100 °C, whereas sucrose and inulin were completely hydrolysed under those conditions. Courtin *et al.* (2009) mimicked gastric conditions (pH 2.0, 37 °C) *in vitro* and found less than 10% hydrolysis of XOS after up to 14 days, while more than half of the fructo-oligosaccharides were hydrolysed after 3 days. Furthermore, previous *in vitro* tests revealed that xylobiose is not hydrolysed to

xylose upon addition of either saliva, pepsin-containing artificial gastric juice, pancreatin, or intestinal mucosa homogenate (Okazaki *et al*. 1991), indicating that at least the 1,4-β-linkage in arabinoxylans resists hydrolysis in the upper GI tract. Taken together, the data presented here corroborates previous observations suggesting that XOS remain largely unabsorbed in the small intestine. The production of SCFA in *in vitro* fermentation XOS studies has been corroborated through several *in vivo* studies. Increased SCFA levels in the caeca of rats or mice were observed after feeding the animals with diets enriched with XOS (Imaizumi *et al*. 1991; J.M. Campbell *et al*. 1997; Van Craeyveld *et al*. 2008). The data presented here corroborate these reports and suggest XOS function as prebiotics. The beneficial effects of XOS on broiler performance have already shown stimulation of butyrate-producing bacteria through cross-feeding of lactate and subsequent effects of butyrate on gastrointestinal function (De Maesschalck *et al.* 2015).

Table 4.5. Ingredient composition and calculated analysis of the corn-based feed of experiment 3.

Ingredients	**%**
Corn	56,96
Soybean meal 47%	35,90
Soybean oil	3,15
Sodium chloride	0,25
Calcium carbonate	1,48
Monocalcium phosphate 22.7%	1,70
DL-Methionine	0,25
Elancoban 200	0,05
Mineral and vitamin premix [1]	0,20
Calculated nutrient content	
Energy (kcal/kg)	3000
Crude Protein (%)	20,00
Ether extract (%)	5,60
Crude cellulose (%)	4,10
Ash (%)	5,88

[1] Mineral-vitamin premix provided the following per kilogram of diet: vitamin A, 9000 IU; vitamin D3, 2100 IU; vitamin E, 20 mg; nicotinic acid, 30 mg; vitamin B12, 0,12 mg; calcium pantothenate, 10 mg; vitamin K3, 2 mg; thiamin, 1 mg; riboflavin, 4,2 mg; vitamin B6, 1,7 mg; folic acid, 0,5 mg; biotin, 0,5 mg; Fe, 80 mg; Cu, 10 mg; Mn, 100mg; Zn, 80 mg; Co, 0,2 mg; I, 1,0 mg; Se, 0,3 mg; monensin, 100 ppm.

Table 4.6. Performance of broilers fed on the corn-based diet (Experiment 3).

Diet supplemented with a XOS preparation provided at three different incorporation rates, 0.1 g/kg (XOS1), 1 g/kg (XOS2) or 10 g/kg (XOS3). A fourth group of birds was fed on a basal non-supplemented diet (C-).

	C-	XOS1	XOS2	W	SEM	p(*F*)
Body Weight (g)						
0d	46,5	46,6	46,1	46,3	0,368	0,777
7d	134^{ab}	136^{a}	137^{a}	127^{b}	2,571	0,044
14d	292^{b}	319^{a}	313^{a}	291^{b}	6,329	0,009
21d	536^{b}	646^{a}	607^{a}	566^{b}	15,774	0,0001
28d	906^{c}	1128^{a}	1036^{b}	964^{c}	27,381	0,001
35d	1480^{c}	1781^{a}	1646^{b}	1536^{c}	43,536	0,001
42d	2084^{b}	2406^{a}	2262^{a}	2136^{b}	56,732	0,003
Weight Gain (g)						
0-7d	87^{ab}	90^{a}	91^{a}	81^{b}	2,442	0,033
7-14d	158^{b}	183^{a}	176^{a}	164^{b}	4,251	0,004
14-21d	245^{c}	327^{a}	294^{b}	274^{b}	10,388	0,002
21-28d	370^{b}	482^{a}	429^{a}	398^{b}	12,640	0,0001
28-35d	574^{b}	653^{a}	615^{a}	573^{b}	18,579	0,019
35-42d	604	625	616	604	20,058	0,856
0-42d	2038^{b}	2359^{a}	2216^{a}	2090^{b}	56,696	0,003
Feed Intake (g)						
0-7d	130	133	134	126	2,747	0,210
7-14d	312^{b}	339^{a}	334^{a}	311^{b}	6,653	0,009
14-21d	539^{b}	592^{a}	585^{a}	5385^{b}	11,916	0,005
21-28d	952^{bc}	1009^{ab}	1039^{ab}	888^{c}	38,188	0,053
28-35d	1229	1234	1319	12020	36,775	0,161
35-42d	1294	1269	1370	1234	40,600	0,144
0-42d	4455^{b}	4576^{ab}	4781^{a}	4299^{b}	102,386	0,023
Feed Conversion Ratio						
0-7d	1,49^{b}	1,49^{b}	1,47^{b}	1,57^{a}	0,022	0,026
7-14d	1,98	1,86	1,91	1,90	0,048	0,418
14-21d	2,21^{a}	1,83^{b}	1,99^{b}	1,99^{b}	0,082	0,036
21-28d	2,59^{a}	2,10^{b}	2,42^{a}	2,27^{b}	0,115	0,042
28-35d	2,15	1,91	2,15	2,10	0,091	0,206
35-42d	2,17	2,04	2,25	2,09	0,111	0,571
0-42d	2,19	1,95	2,16	2,07	0,074	0,119

Means in the same line with different letter superscripts (a, b, c) are significantly different (P<0.05) or tend to be significantly different (P<0.1).

4.4. Conclusions

Data presented here revealed that exogenous xylanases used to supplement wheat-based diets for poultry may mediate their effects by modulating the gut microflora colonizing the upper portions of the birds´ GI tract. Thus, XOS resulting from the hydrolysis of wheat arabinoxylans by exogenous xylanases or used directly as a feed supplement, display a prebiotic effect in broiler chicks. Significantly, the prebiotic effect of XOS was also observed when these oligosaccharides were used to supplement corn-based diets, resulting in a significant growth promoting activity. Overall these data suggest that XOS used as feed supplements resist the passage through the upper regions of the GI tract, a condition required to their use as effective prebiotics in poultry nutrition. In addition, the bio-activity of unbranched XOS seems to be similar to the oligosaccharides resulting from the hydrolysis of feed arabinoxylans that are mostly branched with arabinose side-chains (AXOS). Finally, data of experiment 3 suggest that the incorporation rates of XOS need to be carefully fine-tuned, since XOS used at a 10g/kg level had no effect on performance. Thus, optimum incorporation rates may range between 0.1 to 1 g kg of XOS, although animals fed on corn-based diets supplemented with XOS at a 0.1 g/kg rate resulted in more than 350 grams of additional body weight gain when compared with the control group, suggesting even lower doses may be effective. This observation suggests that XOS modulate an improvement in animal performance by optimizing feed digestion, feed intake and by triggering the evolution of the microbiome to a more favorable construction. An important consideration of the mode of action of this prebiotic relates to the quantities used. At 0.1g/kg, this is equivalent to less than 0.3 Kcal/kg which is clearly so little that the XOS alone cannot be responsible for the scale of response noted in all experiments. Even if all the added XOS were converted with 100% efficiency to SCFA's this would not result in a measurable effect. Thus, we suggest that XOS are effectively a pump primers, sending a signal to encourage those bacteria that can digest feed xylan to become much more active, digest the xylan in the diet more effectively and interact with the digestive tract in such a way that efficiency of digestion as a whole is increased. Such an evolution of the microbiome would not be immediate and develops slowly which is in concert with the performance responses being greater later on in life. One proposal is that the SCFAs produced elicit a Peptide tyrosine-tyrosine (PYY) release which delays gastric emptying and thus enhances gastric digestive efficiency. If XOS and other prebiotics are acting in such a "pump priming" manner, it suggests the definition of a prebiotic may also need revisiting, as their role may simply be to guide the microbiome to more effectively utilize a fermentable substrate (i.e. in this case feed derived xylan) that is already present in the intestine rather than by quantitatively acting as a substrate for fermentation.

5. GENERATION OF A LIBRARY OF CARBOHYDRATE-ACTIVE ENZYMES FOR PLANT BIOMASS DECONSTRUCTION

V. Cardoso[1,2], J.L.A. Brás[1,2], L.M.A. Ferreira[1], L.T. Gama[1], R. Vincentelli[3,4], B. Henrissat[3,4,5] and C.M.G.A. Fontes[1,†]

[1] CIISA – Faculdade de Medicina Veterinária, Universidade de Lisboa, Pólo Universitário do Alto da Ajuda, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; [2] NZYTech LDA, Estrada do Paço do Lumiar, Campus Do Lumiar, 1649-038 Lisboa, Portugal; [3]CNRS, UMR 7257, Université Aix-Marseille, 13288 Marseille, France; [4]Institut National de la Recherche Agronomique, USC 1408 Architecture et Fonction des Macromolécules Biologiques, 13288 Marseille, France; [5]Department of Biological Sciences, King Abdulaziz University, 21589 Jeddah, Saudi Arabia.

Adapted from a manuscript in preparation

ABSTRACT

Plant biomass is the most abundant renewable source of organic carbon on Earth. Microbial degradation of carbohydrates is a key step in the carbon cycle. Deconstruction of plant carbohydrates is mediated by an array of enzymes, generally termed CAZymes. To cover the maximal number of families and activities, an high-throughput strategy was used to produce thousands of CAZymes with high expression levels in *E. coli.* A total of 1476 genes were selected for this study, representing a significant number of CAZy families and EC numbers. CAZymes originated from different source organisms, belonging to Archaea, Bacteria and Eukaryote domains. All genes were successfully produced by PCR (61%) or Gene Synthesis (GS) (39%) and subsequently cloned in *E. coli* expression vectors. Expression results revealed that approximately 79% of the recombinant proteins were obtained at significant levels. A significantly reduced number ($P<0,01$) of proteins were obtained for both eukaryotic and archaeal domains with a percentage of 57,7% and 53,3%, respectively, compared with 79,7% of soluble proteins with a bacterial origin. When the method of gene isolation was compared with production it was observed that a significantly reduced number ($P=0,04$) of proteins were produced with genes obtained by GS. Fusion the GFP tag influenced positively protein solubility ($P=0,05$). In contrast, no relations were observed between protein molecular mass and their solubility. Finally, an association between amino acid composition and protein solubility was observed, suggesting that protein composition in non-polar and negatively charged amino acids may be used to predict protein solubility in *E. coli*. The HTP approach presented here is a powerful tool for production of

recombinant CAZymes that can be used for future studies of PCW degradation. Successful production and expression of a high rate of soluble recombinant proteins open up new possibilities for the high-throughput production of targets from limitless sources.

5.1. Introduction

Plant biomass is the most abundant renewable source of organic carbon on Earth. Plant cell wall (PCW), its major constituent, is a highly heterogeneous and complex macromolecular structure that surrounds and protects the cell playing a central role for plant survival (Caffall and Mohnen 2009). PCW is composed of different types of recalcitrant polysaccharides, notably cellulose (40–50%), hemicellulose (25–35%) and lignin (15–20%) (Gray *et al*. 2006). Microbial degradation of plant structural carbohydrates to generate usable sugars is a key step in the carbon cycle and presents considerable biotechnological importance, in particular for the conversion of lignocellulosic biomass into ethanol, a vital objective for today´s society (Lynd *et al*. 2008). Thus, understanding the biochemistry of plant cell wall deconstruction is not only of biological importance but presents growing industrial significance (Sticklen 2008; Himmel and Bayer 2009; Hervé *et al*. 2010).

Deconstruction of PCW carbohydrates is mediated by a large repertoire of microbial enzymes, generally termed Carbohydrate-Active Enzymes (CAZymes). CAZymes comprise a diversity of enzyme classes including glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and glycosyl transferases (GTs). While PLs, CEs, and GHs carry out the breakdown of polysaccharides, GTs are mainly involved in the formation of the glycosidic bond and thus in the biosynthesis of carbohydrates (Chakraborty *et al*. 2017). Recently, the CAZy database incorporated a new category of proteins, named Auxiliary Activities (AAs), which covers redox enzymes that act cooperatively with CAZymes. The AA category of CAZymes includes lytic polysaccharide monooxygenases (LPMOs), as well as ligninolytic enzymes (Levasseur *et al*. 2013). CAZymes are classified in families according to their amino acid sequence similarities which reflect a common structural fold and catalytic mechanism.

In general, CAZymes have evolved complex molecular architectures containing one or more catalytic domains linked to generally more than one accessory non-catalytic carbohydrate-binding modules (CBMs) (Boraston *et al*. 2004; Hashimoto 2006; Guillén *et al*. 2010). CBMs play an important role in CAZyme function by promoting the hydrolysis of insoluble substrates by the adjoined catalytic domains. CBMs bind structural carbohydrates bringing the associated catalytic domains into close proximity of the recalcitrant substrates and thus targeting the enzyme to their specific substrates (Gilbert 2010; Guillén *et al*. 2010). As CAZymes, CBMs are also classified in families according to their primary sequence similarity (Lombard *et al*. 2014).

The classification of CAZymes in families has been continuously updated in the Carbohydrate-Active EnZymes database (CAZy; www.cazy.org) (Cantarel *et al*. 2009; Lombard *et al*. 2014). The CAZy database is an invaluable research tool with respect to the ever-increasing amount of genomic and metagenomic information relating to carbohydrate metabolism. Currently, the database contains 165 sequence-based families of GHs, 37 families of PLs, 15 families of CEs, 107 families of GTs, 16 families of AAs and, finally, 85 families of CBMs (data collected on July 2019). The functional diversity of CAZymes is enormous and reflects the wide multiplicity of glycan structures found in nature. Nevertheless, although the CAZy database provides a solid base to sustain carbohydrate research, the existence of a tangible library of CAZymes to promote carbohydrate bioengineering research and related applications is still lagging. Here, a high-throughput (HTP) strategy was used to produce a comprehensive library of highly diverse, soluble and pure CAZymes of known biochemical properties and to cover the maximal number of CAZy families and E.C. activities. Thus, a bioinformatic selection of characterized CAZymes was initially pursued, followed by HTP gene synthesis, cloning, expression and purification of recombinant proteins. This approach leads to the generation of a large library of recombinant CAZymes of significant scientific interest to study PCW deconstruction but also of great applicability in the industrial and bioprocessing sectors.

5.2. Materials and Methods

5.2.1. Identification and selection of CAZymes

CAZymes were drawn from the continuously updated CAZy database, from which 1955 biochemically characterized enzymes from distinct classes, families and activities described in literature were identified. The diversity of Enzyme Commission numbers (EC number) was also explored. To tune selection, a gene alignment algorithm (NZYTech genes & enzymes, Portugal) was used to exclude highly similar sequences, namely primary sequences with homology higher than 90%. Signal peptides were predicted using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/; Petersen *et al.* 2011) and removed from all candidate sequences. Enzyme domains were predicted using dbCAN, which provides pre-computed CAZyme sequence and annotation data for several bacterial genomes (Yin *et al*. 2012).

5.2.2. Polymerase chain reaction (PCR) and Gene synthesis (GS)

Genes encoding 1476 CAZymes selected from a database of 1955 sequences were obtained through PCR or GS (Table S5.1). When genomic DNA was commercially available (DSMZ-German Collection of Microorganisms and Cell Cultures GmbH., Germany) genes were obtained by PCR. DNA amplifications were performed with Supreme NZYProof DNA

Polymerase (NZYTech genes & enzymes, Portugal) using genomic DNA as template and gene-specific primers. PCR conditions followed the manufacturer's protocols. Following amplification, assembled PCR products were purified using NZYDNA Clean-up 96 well plate (NZYTech genes & enzymes, Portugal) in a Liquid Handling Robot workstation (TECAN, Freedom EVO series, Switzerland). On the other hand, the primary sequences of CAZymes, for which no template was available, were obtained by *de novo* gene synthesis. Gene sequences were designed by back-translating the protein sequences and optimizing codon usage for high levels of expression in *Escherichia coli*, using ATGenium codon optimization algorithm. In addition, genes were designed to ensure a Codon Adaptation Index (CAI) value higher than 0.8. Synthetic genes were produced in a HTP pipeline using optimized procedures, as described in Sequeira *et al.* (2017). The databank describing both DNA and amino acid sequences of the 1476 genes and enzymes, their predicted molecular mass, molecular architecture and origin is presented in Table S5.1.

5.2.3. Cloning, transformation and sequencing

Purified PCR and GS products were cloned into pHTP1 or pHTP9 expression vectors (NZYTech genes & enzymes, Portugal) using NZYEasy Cloning & Expression kits I and IX, respectively (NZYTech genes & enzymes, Portugal), according to the protocol reported in Turchetto *et al.* (2017) and Sequeira *et al.* (2017). Following the cloning reaction, recombinant plasmids were transformed using an HTP method into *E. coli* NZY5α competent cells (NZYTech genes & enzymes, Portugal). The transformed bacteria were spread on LB-agar kanamycin 24-deep-well-plates (24-DW). After an overnight incubation at 37 °C, one colony per transformation was picked and grown in liquid LB media supplemented with 50 µg/mL of kanamycin in 24-DW plates (5 mL) sealed with gas-permeable adhesive seals. The plasmids were purified from the bacterial pellets using NZYMiniprep 96 well plate kit (NZYTech genes & enzymes, Portugal) on a Tecan workstation (TECAN, Freedom EVO series, Switzerland). All constructs were completely sequenced in both directions to ensure 100% consistency with the gene sequences.

5.2.4. High-throughput protein expression, purification and quantification

Protein expression and purification steps were based in the protocol described by Turchetto *et al.* (2017) with few modifications. *E. coli* BL21 (DE3) cells (NZYTech genes & enzymes, Portugal) were used to transform the recombinant pHTP clones. Recombinant strains were cultured in 24-DW with 5 mL of Auto-Induction LB medium supplemented with 50 µg/mL of kanamycin and grown at 37 °C, for 4 h followed by 18 h at 25 °C. At the end of the culture, cells were harvested by centrifugation at 1500 ×g for 15 min at 4 °C. The cell pellets were resuspended in 1 mL of NZY Bacterial Cell Lysis Buffer (NZYTech genes &

enzymes, Portugal) with 4 µg/mL of DNAse and 100 µg/mL of Lysozyme. The proteins were then purified by immobilized metal ion-affinity chromatography (IMAC) and all the steps were automated on a Tecan workstation (TECAN, Freedom EVO series, Switzerland) containing a vacuum manifold. Briefly, 1 mL of crude cell lysates were incubated with 200 µL Ni^{2+} Sepharose 6 Fast flow resin (GE Healthcare, 17-5318-02) with bound Nickel and then transferred into 96-well filter plates (20 µm) (Macherey–Nagel). The wells were washed twice with 1000 µL buffer A (50 mM NaHepes, 1 M NaCl, 10 mM Imidazole, 5 mM $CaCl_2$, pH 7,5) followed by one wash with 1000 µL buffer B (NaHepes, 1 M NaCl, 35 mM Imidazole, 5 mM $CaCl_2$, pH 7,5) and, finally, the proteins were eluted with 300 µL of elution buffer (NaHepes, 1 M NaCl, 300 mM Imidazole, 5 mM $CaCl_2$, pH 7,5). Integrity of purified recombinant proteins, in terms of solubility, purity and molecular weight, was assessed by sodium dodecyl sulphate 14% polyacrylamide gel electrophoresis (SDS-PAGE) and protein concentration was measured using the Bradford protein assay (NZYTech, Portugal) against a standard curve of bovine serum albumin (BSA). Recombinant proteins molecular weight and extinction coefficient were calculated using ExPASy– ProtParam tool (http://www.expasy.ch/tools/protparam.html).

5.2.5. Statistical Analyses

The effect of different organism of origin, gene production strategies, vector with GFP tag, as well as molecular mass was analysed with chi-squared test. Data related with primary sequence analysis were subjected to ANOVA according to the general linear models' procedures. All statistical analysis were performed with a 95 % of confidence interval and conducted with SAS (2004).

5.3. Results and Discussion

5.3.1. Identification and selection of CAZymes

Initially, a bank of 1955 genes encoding biochemically characterized CAZymes was constructed. All gene sequences were analysed using a gene alignment algorithm, in order to exclude highly similar sequences and to explore CAZy family's diversity to include a broad range of EC numbers. As a result, 1476 enzymes of several classes, families and EC numbers were selected. The list of proteins included enzymes from 412 different source organisms of the three distinct domains: Bacteria (968 targets), Eukaryote (21 targets) and Archaea (9 targets). The 1476 enzymes comprise 787 GHs, 77 PLs, 522 CEs, 15 AAs, 336 CBMs and 36 others (containing different catalytic domains) (Figure 5.1). Some sequences were designed to solely include one single module while in other cases sequences were not edited in order to preserve all modules together as they are natively expressed. All signal peptides were predicted and removed from target sequences. Overall, the main goal of this

initial stage was to constitute a wide CAZy library of soluble proteins for deconstruction of PCW polysaccharides.

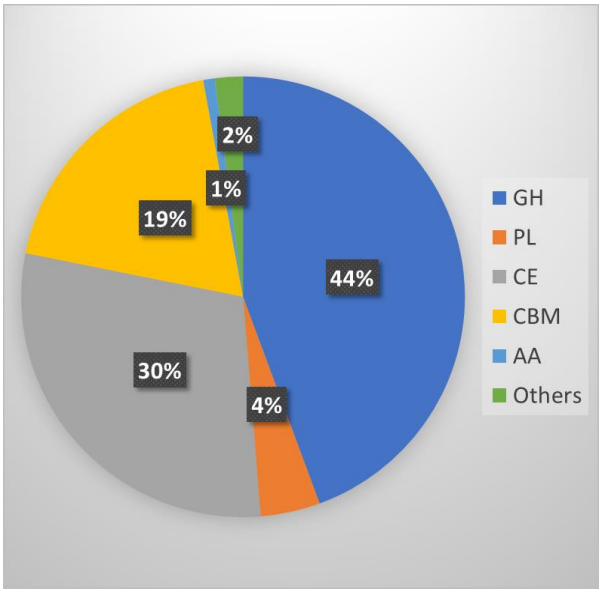


Figure 5.1. The distribution of the 1476 enzymes produced in this study by the five CAZy classes.

5.3.2. Genes production and cloning

All 1476 genes were successfully produced by amplification from genomic DNA (PCR) or by gene synthesis (GS). In total, 61 % of the genes were isolated through PCR amplification using the corresponding genomic DNA as template and the remaining 39 % of genes were codon optimized and synthesized. Both PCR and GS products were cloned in pHTP expression vectors, encoding an N-terminal Histidine tag for protein purification. A total of 1350 genes, including all genes-encoding enzymes from GHs, CEs, PLs, AAs families, as well as some CBMs, were cloned in pHTP1 expression vector. The remaining 126 genes, encoding for CBMs were cloned in pHTP9 expression vector, which means that these recombinant proteins contain a N-terminal GFP tag.

5.3.3. Protein expression and putative factors affecting solubility

The 1476 recombinant enzymes were expressed in *E. coli* BL21(DE3) and purified using an HTP automated workstation through IMAC. The molecular integrity of recombinant proteins was evaluated by SDS-PAGE. The data revealed that, in general, molecular weight (Mw) of the purified enzymes was highly similar to the calculated theorical value (Figure 5.2). In addition, 79% of the proteins were obtained in the soluble form (>20 mg of purified protein per litre of media). Considering that all the 1476 enzymes were previously biochemical characterized it was anticipated that a higher number of targets could have been obtained at

higher levels. However, considering that only one condition was tested for the 1476 targets (same expression host, growth medium, induction conditions and purification method) and that the threshold for solubility (>20 mg/L) was significantly high it is reasonable to accept that a proportion of ~20% of the enzymes required significant optimization of the expression conditions to allow recovering sufficient protein for biochemical characterization. Nevertheless, the effect of protein origin, gene production strategy, fusion with solubility tags, protein molecular weight and amino acid composition, on protein expression and solubility is discussed below.

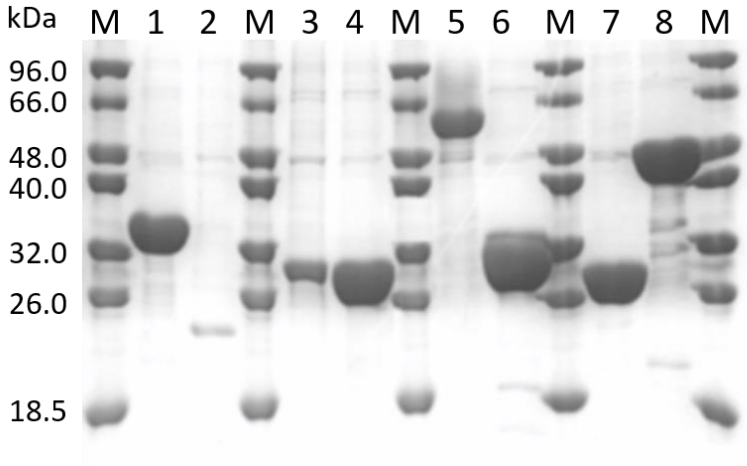


Figure 5.2. Protein expression and purity analysis of CAZymes exemplified by a SDS-PAGE gel.

5.3.4. The effect of protein origin on recombinant production

Data presented in Figure 5.3-A revealed that a significantly reduced number (P<0,01) of enzymes originating in eukaryotic or archaeal domains were obtained during the course of these studies. Comparison of recombinant proteins solubility across the three domains of life shows that 79,7 % (1143 out of 1435) of proteins with bacterial origin were obtained, but when analysing the production of eukaryotic and archaeal targets this percentage dropped to 57,7% (15 out of 26) and 53,3% (8 out of 15), respectively (Figure 5.3-A). The higher production rate found for bacterial targets is not surprising, considering that the expression system selected was *E. coli* (Camilo and Polikarpov 2014). For example, it is widely accepted that presence of native disulphide bonds in eukaryotic proteins represents an obstacle for proper protein folding and soluble expression in *E. coli* (Nguyen *et al*. 2011; Camilo and Polikarpov 2014), resulting in the formation of insoluble inclusion bodies (Fischer *et al*. 1993). In addition, archaeal targets are usually poorly expressed in *E. coli* because of their origin in extreme environments and differences in host codon usage. A variety of archaeal targets originated in hyperthermophiles microorganisms which gene sequences usually contain a high proportion of rare codons that are poorly processed in *E. coli* (Kim and

Lee 2006). Thus, it is expected that expression of archaeal targets in *E. coli* results in significant levels of insoluble inclusion bodies (Kim and Lee 2008).

5.3.5. Influence of gene production strategy, gene synthesis or PCR, in protein production

The effect of gene production strategy in production of recombinant proteins was analysed. The data, presented in Figure 5.3-B, revealed that lower numbers of targets were obtained from genes synthesised artificially (P=0,04). Several studies have demonstrated that codon-optimization for heterologous production in *E. coli* can potentially increase recombinant protein expression, especially in sequences of eukaryotic and archaeal origin (Lakey *et al*. 2000; Korepanova *et al*. 2005; Piubelli *et al*. 2013; Lai *et al*. 2014; Devi *et al*. 2016; Fang *et al*. 2016; Mirzaei *et al*. 2016; Xue *et al*. 2016). However, other studies revealed that *E. coli* codon-optimization may lead to overexpression of recombinant proteins, thus resulting in a higher tendency to the generation of insoluble inclusion bodies (Fahnert *et al*. 2004; Zhao *et al*. 2015). Formation of inclusion bodies in recombinant expression systems occurs as a result of an erroneous equilibrium between *in vivo* protein folding and aggregation (Mannall *et al*. 2007; de Groot *et al*. 2008; Eiberle and Jungbauer 2010). Data presented here revealed that although the number of obtained proteins was reduced when genes were obtained synthetically, a significant proportion of non-viable targets revealed high levels of expression in the form of inclusion bodies (data not shown). Therefore, we can conclude that *E. coli* codon-optimization increased levels of recombinant protein expression. However, for some proteins higher levels of expression are beyond *E. coli* capacity to provide a proper folding and led to the accumulation of the recombinant proteins in the form of insoluble aggregates.

5.3.6. Does the GFP tag promotes protein production?

Many studies have focused on optimising the production processes for recombinant proteins with the aim of reducing the accumulation of insoluble inclusion bodies. In contrast, solubilisation and refolding of protein from inclusion bodies is a common strategy although it requires denaturing conditions as well as a subsequent renaturing step usually resulting in poor soluble protein recoveries (Fischer *et al*. 1993; Sahdev *et al*. 2008). Different approaches have been developed to prevent the accumulation of inclusion bodies in *E. coli,* such as the optimization of culture conditions, co-expression with molecular chaperons (Mogk *et al*. 2002; Voulgaridou *et al*. 2013), lowering growth temperature during gene induction (Schein and Noteborn 1988; Vera *et al*. 2007), induction expression in early-log phase culture (San-Miguel *et al*. 2013) and induction with lower levels of inducer concentration, such as Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Huyen *et al*. 2014). In

addition, several fusion tags have been developed to increase the solubility of overexpressed proteins, but with variable degrees of success. Currently available fusion systems include maltose-binding polypeptide (MBP), glutathione S-transferase (GST), ubiquitin, and thioredoxin (TrxA) and Green Fluorescent Protein (GFP) (Zhang *et al*. 1998). The use of a fusion partner may increase the solubility of the folding and also significantly contribute to increase the expression yields (Nguyen *et al*. 2011). This is consistent with the results presented here with higher number (P=0,05) of proteins obtained when fused with GFP tag (pHTP9) (Figure 5.3-C), comparatively with proteins that only include a histidine tag (pHTP1).

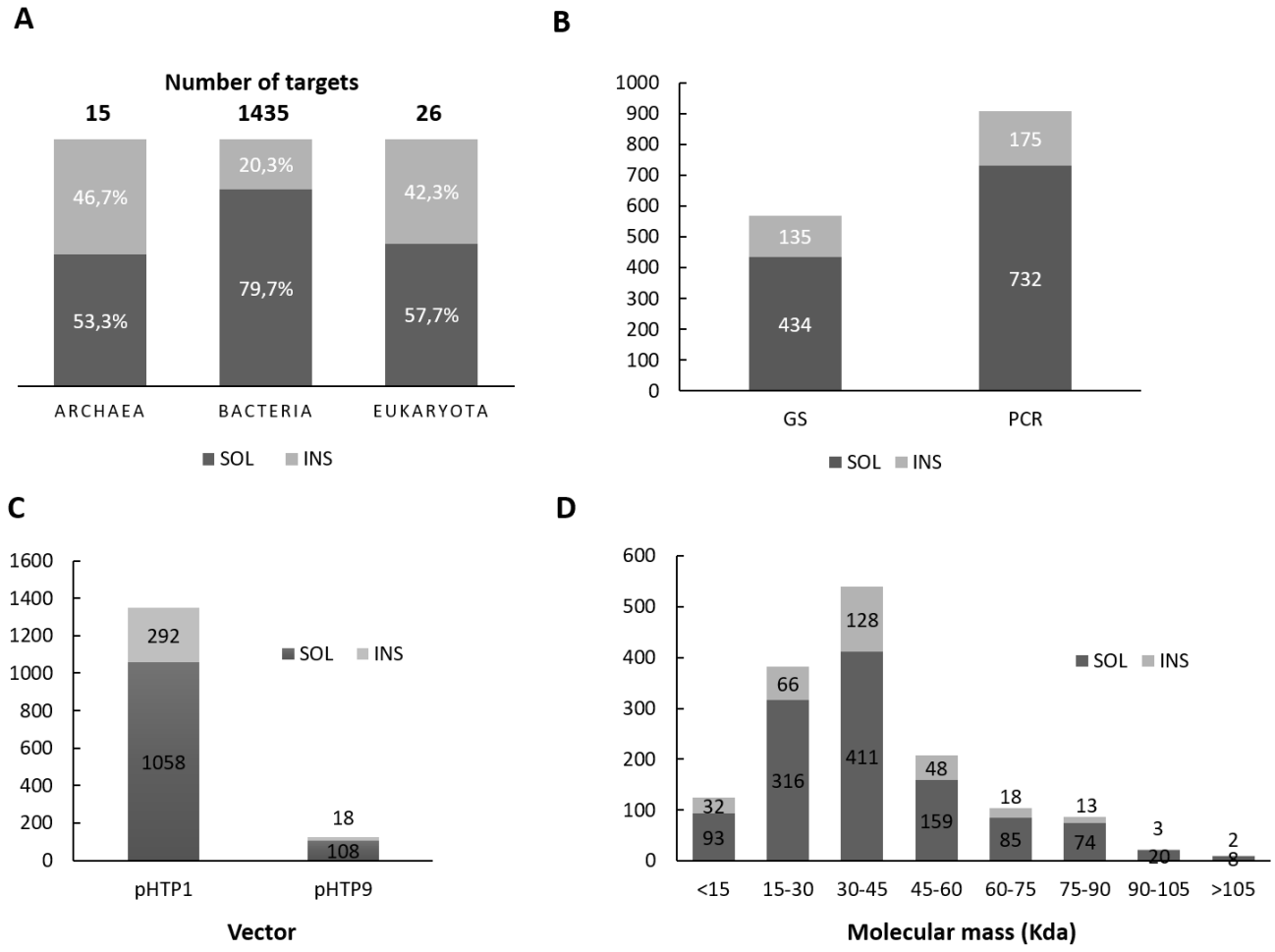


Figure 5.3. Efficacy of the recombinant production of CAZymes.

A. The three domains of life (proportion of recombinant proteins obtained among eukaryotic, bacteria and archaea targets); **B.** Production strategy, *(GS,* genes synthetized with codon-optimization for *E. coli*; *PCR,* genes obtained from genomic DNA; **C.** Expression vectors: pHTP1, expression vector encoding an N-terminal Histidine tag; pHTP9, vector containing a N-terminal GFP tag; **D.** Protein's molecular mass distribution. SOL: Soluble protein, INS: Insoluble protein.

5.3.7. Influence of protein molecular weight in protein production

There is no relation between the recombinant proteins obtained and protein molecular weight (Figure 5.3-D). An absence of relation between these two factors is not surprising. Usually, proteins with molecular mass below 100 kDa are well tolerated in *E. coli* and are therefore expressed to significant levels. In contrast, proteins above 100 kDa are usually not properly processed in the bacterium and hence are degraded or subjected to premature termination (Kaur *et al*. 2018). Palacios *et al*. (2001) and Sauvonnet *et al*. (1995) also suggested that proteins size can affect secretion performance as large cytoplasmic proteins are physically impossible to translocate (Bianchi and Baneyx 1999; Feilmeier *et al*. 2000). Since only 10 proteins in this study (Figure 5.3-D) presented a molecular mass higher than 105 kDa it is not surprising that protein molecular weight had no effect on production. In contrast, smaller proteins, below 10 kDa, are difficult to express stably in *E. coli* because of their improper folding and higher tendency to proteolytic degradation. These proteins can be stabilized when expressed in fusion to large proteins such as MBP, GST and GFP which allow them to fold properly (Kaur *et al*. 2018). Considering smaller CBMs were directly cloned into pHTP9 expression vector (121 out of 304 CBMs), which contains an N-terminal GFP tag, this difficulty was attenuated.

5.3.8. Influence of primary sequence composition on protein production

Previous studies have shown a relationship between amino acids composition and recombinant protein production. Therefore, the amino acids frequency of proteins expressed in *E. coli* using the described HTP approach was analysed. Comparison of amino acids groups frequency for proteins obtained or not obtained at significant levels revealed differences for composition on non-polar amino acids (P=0,002), for negatively charged amino acids (P<0,001) and for the content of specific amino acids (see Table 5.1). The non-polar amino acids group is composed by Glycine, Cysteine, Alanine, Leucine, Isoleucine, Valine, Methionine, Proline, Phenylalanine and Tryptophan (G, C, A, L, I, V, M, P, F and W). Negatively charged amino acids group, also named acidic amino acids, is constituted by Aspartic acid and Glutamic acid (D, E). The data suggest that, in *E. coli*, a lower frequency of non-polar amino-acids and higher percentage of acidic amino acids increases the probability of a protein to be soluble. This data are in agreement with Bertone *et al.* (2001) and Christendat *et al.* (2000), which verified that protein solubility was significantly influenced by the frequencies of the acidic amino acids, basic amino acids and non-polar amino acids. More recently, Niu *et al.* (2013) confirmed the importance of negatively charge residues (D, E). Besides that, Niu *et al.* (2013) observed that dipeptides composed by the acidic amino acids (D, E) and basic amino acids (K, R and H), especially the dipeptides composed by the acidic amino acids (D, E) were the major determinants of the protein solubility. The relative

content of negatively charged residues (D, E) seems to be the strongest determinant of protein's solubility as it was selected as key attribute in several studies (Wilkinson and Harrison 1991; Davis *et al*. 1999; Christendat *et al*. 2000; Bertone *et al*. 2001; Goh *et al*. 2004).

Table 5.1. Analysis of the amino acid composition in primary sequences of soluble and insoluble CAZymes produced in this study.

	INS	**SOL**	**SEM**	***P-value***
(n)	310	1166		
Amino Acids groups (%)				
Non-polar amino acids[1]	51,4[a]	50,2[b]	0,31	0,002
Polar neutral amino acids[2]	26,0	26,0	0,38	0,976
Negatively charged amino acids[3]	11,2[b]	12,2[a]	0,17	0,001
Positively charged amino acids[4]	11,4	11,5	0,17	0,420
Amino Acids content (%)				
Isoleucine (I)	5,0	5,2	0,11	0,051
Leucine (L)	7,6	7,4	0,15	0,213
Lysine (K)	4,4[b]	5,0[a]	0,15	0,001
Methionine (M)	1,9	2,0	0,06	0,138
Phenylalanine (F)	4,1	4,1	0,08	0,828
Threonine (T)	6,2	6,0	0,14	0,231
Tryptophan (W)	2,3[a]	2,2[b]	0,07	0,042
Valine (V)	6,5	6,5	0,10	0,610
Arginine (R)	4,7[a]	4,3[b]	0,12	0,005
Histidine (H)	2,3	2,3	0,07	0,659
Alanine (A)	9,0[a]	8,5[b]	0,20	0,014
Asparagine (N)	5,1	5,2	0,15	0,511
Aspartic acid (D)	6,0[b]	6,5[a]	0,10	0,001
Cysteine (C)	1,0	1,0	0,07	0,784
Glutamic acid (E)	5,2[b]	5,6[a]	0,12	0,002
Glutamine (Q)	3,6	3,4	0,09	0,287
Glycine (G)	8,9[a]	8,6[b]	0,13	0,016
Proline (P)	5,0	4,9	0,10	0,418
Serine (S)	6,6	6,5	0,15	0,246
Tyrosine (Y)	4,5[b]	4,9[a]	0,11	0,001

Means in the same line with different letter superscripts (a, b, c) are significantly different (P<0.05) or tend to be significantly different (P<0.1). [1] Non-polar amino acids group is composed by Glycine (G), Cysteine (C), Alanine (A), Leucine (L), Isoleucine (I), Valine (V), Methionine (M), Proline (P), Phenylalanine (F) and Tryptophan (W); [2] Polar neutral amino acids grou pis composed by Asparagine (N), Glutamine (Q), Serine (S), Threonine (T) and Tyrosine (Y); [3]Negatively charged amino acids group, also named acidic amino acids, is constituted by Aspartic acid (D) and Glutamic acid (E); 4 Positively charged amino acids, also named basic amino acids, is constituted by Arginine (R), Histidine (H) and Lysine (K).

Analysis of amino acids frequencies in proteins obtained and not obtained with success revealed that Lysine (K), Tryptophan (W), Arginine (R), Alanine (A), Aspartic acid

(D), Glutamic acid (E), Glycine (G) and Tyrosine (Y) (Table 5.1) are important to discriminate insoluble and soluble proteins. In addition, the effect of aspartic acid, glutamic acid and glutamine residues (D, E, and Q) in proteins' solubility was previously described by Bertone *et al*. (2001). However, the present analysis reveals that other amino acids, in particular the relative content of Lysine (K), Tyrosine (Y) and Arginine (R) seems to be related to the propensity of a protein to be obtained, with a confirmed statistically significant effect ($p<0,002$).

Over the past two decades, different studies have explored the relation between protein solubility and primary sequences composition. Solubility prediction can increase the overall success rate of experiments by avoiding the potentially insoluble targets and choosing the promising candidates. Therefore, further analysis of the dipeptide content should be performed in order to develop a bioinformatic tool for CAZymes solubility prediction, based on primary sequence analysis. An accurate theoretical prediction of solubility from sequence is instrumental for setting priorities on targets in large-scale proteomics projects.

5.3.9. Final outputs

In the present study, we started with 1955 putative candidates, from 486 different organisms, contemplating 62% and 66% of CAZy families and EC numbers described in CAZy database, respectively (Figure 5.4-A and Figure 5.4-B) (www.cazy.org, data collected on July 2019). Following sequence analysis, 1476 enzymes, comprising 60% of families and EC numbers described in CAZy database, were selected for production (Figure 5.4-C). From these CAZymes, 732 soluble proteins, belonging to 206 different organisms, were selected (Figure 5.4-D), to build a large CAZyme library to develop fundamental and applied research (Figure 5.4-E). Together with a previously implemented bank, the present library includes more than 1000 CAZymes (Figure 5.4-E), belonging to 219 different organisms with different origins: 968 bacterial, 21 eukaryotic and 9 archaeal. These enzymes are distributed in five CAZy classes and comprise a large number of families, namely 629 GHs, 62 PLs, 138 CEs, 8 AAs and 163 CBMs and 2 modular enzymes (Figure 5.5). Therefore, approximately 60% of all families described in CAZy database are covered by the present recombinant library that constitutes a valuable resource the exploit the biological and biotechnological potential of carbohydrates.

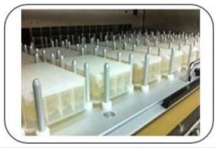
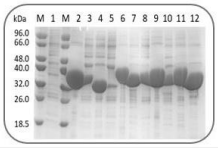
	A. CAZY Database	B. Original selection	C. Protein production	D. Final selection	E. Nzytech's Library
					
	Number of families				
GH	165	116	112	101	109
PL	37	19	18	16	19
CE	15	12	12	12	13
AA	16	5	5	3	4
CBM	85	45	45	40	48
TOTAL	**318**	**197**	**192**	**172**	**193**
	Number of ECs				
GH	186	130	120	108	113
PL	19	17	15	15	15
CE	17	9	8	8	8
AA	23	6	5	4	4
TOTAL	**245**	**162**	**148**	**135**	**140**

Figure 5.4. Number of CAZY families and ECs obtained in this study.

A. CAZy database numbers (www.cazy.org), **B.** Biochemically characterized enzymes selected, **C**. Produced enzymes, **D**. Enzymes elected to be included in Nzytech' library, and **E**. Final Nzytech' CAZy commercial library.

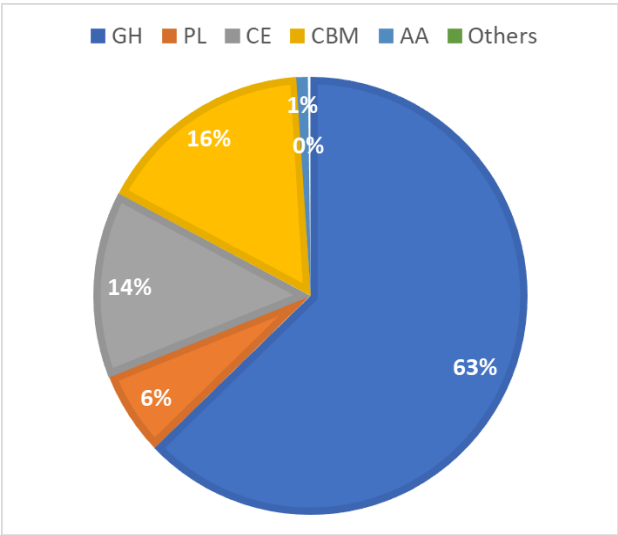


Figure 5.5. Number of CAZy classes and families represented in the current library.

5.4. Conclusions

The main drawback of biomass use is related to the complexity of macromolecular polysaccharides composition which requires a wide plethora of CAZYmes presenting a broad-spectrum of specificities. The strategy used in this study combined bioinformatic tools for enzyme selection and an optimized HTP platform for production, cloning, expression and purification of recombinant proteins with a wide diversity of enzyme specificities. This approach resulted in a library comprising more than 1000 CAZymes involved in carbohydrate degradation, making possible its applicability in both industrial and bioprocessing sectors. Data presented here reinforce the observations that different factors such as, production strategies and protein primary sequence, play a major role in determining the expression and solubility of a protein heterologously produced in *E. coli*. Further analysis on amino acid composition of the proteins selected for this study must be performed in order to develop innovative strategies to predict protein solubility. Nevertheless, the effectiveness of the HTP approach described here is illustrated by the high number of soluble targets obtained, most of them with high solubility rates. Taken together these studies reveal that the implemented strategy described here is a powerful tool not only for CAZymes production but also to generate large libraries of recombinant proteins that will allow exploring the biological function of the extensive genomic and meta-genomic information available from various sources.

6. STRUCTURAL DETERMINANTS OF SUBSTRATE SPECIFICITY IN FAMILY 15 CARBOHYDRATE ESTERASES

V. Cardoso[1,2], J.L.A. Brás[2], A.L Carvalho[3], L.M.A. Ferreira[1,2], V.D. Alves[1], C.M.G.A. Fontes[1,2] and P. Bule[1,†]

[1] CIISA - Centro de Investigação Interdisciplinar em Sanidade Animal, Faculdade de Medicina Veterinária, Universidade de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; [2] NZYTech genes & enzymes, Estrada do Paço do Lumiar, Campus do Lumiar, Ed. E - R/C, 1639-048 Lisboa, Portugal; [3] UCIBIO, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal.

Adapted from a manuscript in preparation

Abstract

Glucuronoyl esterases (EC 3.1.1.-) are microbial enzymes that cleave the ester bonds between glucuronic acid residues that decorate xylans and lignin alcohols. By untangling the lignin-polysaccharide network, glucuronoyl esterases facilitate the access of cellulases and hemicellulases to their target substrates thus promoting biomass recycling in nature. Although several glucuronoyl esterases have recently been characterized the mechanism of substrate recognition by these highly specific carbohydrate esterases remains relatively unexplored. Here we describe the isolation of 20 bacterial and fungal glucuronoyl esterases and their recombinant expression in *Escherichia coli*. Eleven enzymes were efficiently produced and purified and revealed the expected glucuronoyl esterase activity. The structure of the glucuronoyl esterase displaying the highest catalytic activity, *Tt*GE15A from *Teredinibacter turnerae*, was solved. *Tt*GE15A revealed the classical α/β hydrolase fold observed in several esterases and typical of the family 15 Carbohydrate Esterase family. The enzyme contains a peculiar and unique active site that is highly occluded by two large insertions that will probably limit the access to large polymeric lignin-carbohydrate substrates. This possibility is supported by the high activity displayed by *Tt*GE15A to small artificial substrates that mimic the structure of small xylo-oligosaccharides esterified with lignin alcohols. Structural comparisons with other CE15 enzymes reveal different active site topologies in glucuronoyl esterases, suggesting that some enzymes, such as *Tt*GE15A from *T. turnerae*, act on esterified xylo-oligosaccharides, while others may target high molecular mass lignin-carbohydrate complexes (LCCs).

6.1. Introduction

The plant cell wall is the most abundant reservoir of carbon in the biosphere. A diversity of complex molecules, including cellulose, hemicellulose and lignin, are intricately organized in this dynamic barrier to form a structurally diverse and complex construction. The intrinsic rigidity and recalcitrance of plant cell walls is, in part, due to the covalent cross-linkage between lignin and hemicellulose. These lignin-carbohydrate complexes (LCCs) are robust obstacles to enzymatic hydrolysis as they significantly contribute to the occlusion of plant cell wall substrates (Du *et al*. 2013; Huang *et al*. 2015). LCCs are formed during lignin biosynthesis and exist as three main types: esters, benzyl ethers and phenyl glycosides (Watanabe 1995; Balakshin *et al*. 2011). The ester bonds formed between hydroxyl groups of lignin alcohols and 4-O-methyl-D-glucuronic acid (MeGlcA) residues of xylans are susceptible to enzymatic degradation by glucuronoyl esterases (GEs), a recently described class of highly specific enzymes (Watanabe and Koshijima 1988; Balakshin *et al*. 2011; Bååth *et al*. 2016). By cleaving LCCs, GEs improve the overall accessibility of plant cell wall carbohydrates and lignin to enzymatic hydrolysis, a process that is not only biologically relevant but has considerable biotechnological importance, in particular in industrial applications such as bioethanol production.

The first GE to be discovered, *Sc*GE1 from *Schizophyllum commune*, was characterized in 2006 after seminal work developed by Biely and colleagues while studying the stereochemistry of a GH67 α-Glucuronidase (Špániková and Biely 2006; Li *et al*. 2007). After the discovery of this enzyme activity, homologous enzymes were identified in the proteomes of several filamentous fungi and bacteria and presently, more than 24 different GEs were characterized. All glucuronoyl esterases (EC 3.1.1.-) identified so far are organized in Carbohydrate Esterase family 15 (CE15), in the constantly updated Carbohydrate-Active enZYmes Database (CAZy, www.cazy.org). Members of CE15 are mainly of bacterial origin (>255 bacterial enzymes), although fungal enzymes are also significantly represented in this family (~30 members) with a single member from an Archaea (*Halorhabdus utahensis*). Nevertheless, unlike most CAZy families, only a fraction of CE15 enzymes have been characterized, all exhibiting GE activity.

Structurally characterized CE15 members present a typical α/β-hydrolase fold with the active site generally positioned at the protein surface. In general, the structures of bacterial enzymes reveal a more elaborated architecture in the vicinity of the active site, which results in a less exposed catalytic triad when compared with fungal enzymes. Notably, the serine-histidine-glutamate catalytic triad is not absolutely conserved across all CE15 enzymes, with the absence of the glutamate in several cases suggesting that other residues in the vicinity of the catalytic machinery can assume the role of the Brønsted acid (De Santi *et al*. 2016). The structure of *Thermothelomyces thermophilus'* GE, *Tt*GE15D, in complex

with 4-O-methyl-β-D-glucopyranuronate (4g4j), remains as the only evidence for the mechanism of substrate recognition by GEs. Glucuronic acid (GlcA) is tightly bound to the enzyme through a large hydrogen-bound network organized by a constellation of residues that are generally conserved in GEs. Although GEs are very specific towards the GlcA moiety, the mechanism for recognizing the anomeric carbon linked carbohydrate region of the substrate or the alcohol (lignin) part of the ester remains to be clarified. Generally, it is believed that GEs do not discriminate between different types of carbohydrates linked to GlcA although there is some evidence that GEs display higher activity for substrates with bulkier alcohol groups of the urinate ester (Katsimpouras *et al*. 2014; D'Errico *et al*. 2015; D'Errico *et al*. 2016; Hüttner *et al*. 2017). Here we attempt to clarify CE15's mechanism for substrate recognition by characterizing several bacterial GEs. The structure of the enzyme presenting the highest catalytic activity was solved in attempt to clarify the molecular determinants of substrate specificity in family 15 Carbohydrate Esterases.

6.2. Materials and Methods

6.2.1. GEs selection and Phylogenetic analysis

Starting from a database of 291 characterized and putative enzymes, 20 family CE15 enzymes, 5 fungal and 15 bacterial, covering a wide sequence diversity were selected using the BLASTP tool. Signal peptides were predicted using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/; Petersen *et al.* 2011) and removed from all candidate sequences. Modular architecture determination and domain assignment were performed with dbCAN (Yin *et al.* 2012). Molecular weights and molar attenuation coefficients were calculated by ExPASy ProtParam tool (http://www.expasy.ch/tools/protparam.html). Construction of the phylogenetic tree and multiple sequence alignments were performed with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

6.2.2. Chemicals

Culture medium and competent cells were purchased from NZYTech (Portugal). Similarly, all molecular biology kits, including NZYGelpure, NZYEasy Cloning & Expression and NZYMiniprep kits were provided by NZYTech (Portugal). Benzyl D-glucuronate was provided by Carbosynth (Compton, UK). All other chemicals were purchased from Sigma-Aldrich (Sigma– Aldrich, St. Louis, MO).

6.2.3. Strains, media and vector

The *Escherichia coli* strains used in this study were NZY5α for cloning and BL21(DE3) for bacterial protein expression (NZYTech, Portugal). The two *E. coli* strains were

cultured in Luria broth or Auto-Induction LB medium (NZYTech, Portugal), respectively, supplemented with 50 µg/mL kanamycin. All bacterial genes were cloned in pHTP1 expression vector (NZYTech, Portugal).

6.2.4. Polymerase chain reaction (PCR) and Gene synthesis (GS)

Genes encoding 20 GEs from 17 different microorganisms were obtained by PCR or GS. Gene isolation through PCR was performed using the respective genomic DNA as template, Supreme NZYProof DNA Polymerase (NZYTech, Portugal) and the gene-specific primers described in Table S6.1. PCR reactions were performed according to manufacturer' protocol. After DNA amplification, PCR products were purified using NZYDNA Clean-up 96 well plate in Liquid Handling Robot (TECAN, Freedom EVO series, Switzerland). In contrast, genes from GS were designed and optimized for expression in *E. coli*, using the ATGEnium algorithm of codon optimization software (NZYTech, Portugal). The databank describing both DNA and amino acid sequences of the 20 genes and enzymes, their predicted molecular mass, molecular architecture and origin is presented in Table S6.2 and Table S6.3.

6.2.5. Cloning, transformation and sequencing

Purified PCR products and GS products were cloned into pHTP1 expression vector using the NZYEasy Cloning & Expression kit I, according to the protocol reported in Turchetto *et al.* (2017) and Sequeira *et al.* (2017). Following the cloning reaction, recombinant plasmids were transformed using a high-throughput method into *E. coli* NZY5α competent cells. The plasmids were purified from the bacterial pellets using NZYMiniprep 96 well plate and subsequently sequenced. Gene sequence integrity was confirmed by Sanger sequencing.

6.2.6. High-throughput protein expression and purification by nickel affinity chromatography

All steps were based in the protocol described by Turchetto *et al.* (2017), with few modifications. *E. coli* BL21 (DE3) cells were transformed with the 20 recombinant pHTP1 GEs. Recombinant strains were cultured in DW24 plates with 5 mL of Auto-Induction LB medium, supplemented with 50 µg/mL kanamycin and grown over 24 h at 25 °C. At the end of the culture, cells were harvested by centrifugation at 1500 ×g for 15 min at 4 °C. Cell pellets were resuspended in 1 mL of NZY Bacterial Cell Lysis Buffer (NZYTech, Portugal) supplemented with 4 µg/mL of DNAse and 100 µg/mL of Lysozyme. The proteins were then purified by immobilized metal ion-affinity chromatography (IMAC), using an automated Liquid Handling Robot containing a vacuum manifold (TECAN, Freedom EVO series, Switzerland). Briefly, 1 mL of crude cell lysates was incubated with 200 µL Ni^{2+} Sepharose 6 Fast flow

resin (GE Healthcare, 17-5318-02) with bound Nickel and then transferred into 96-well filter plates (20 µm) (Macherey–Nagel). The beads were washed twice with 1 mL of buffer A (50 mM NaHepes, 1 M NaCl, 10 mM Imidazole, 5 mM $CaCl_2$, pH 7,5) followed by one wash with 1 mL of buffer B (NaHepes, 1 M NaCl, 35 mM Imidazole, 5 mM $CaCl_2$, pH 7,5). Finally, the recombinant proteins were eluted with 300 µL of elution buffer (NaHepes, 1 M NaCl, 300 mM Imidazole, 5 mM $CaCl_2$, pH 7,5). Integrity of purified recombinant proteins was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and protein concentration was measured using a µDrop Plate (Multiskan, Thermo Fisher Scientific).

6.2.7. Glucuronoyl Esterases Activity

Activity of the recombinant GEs towards benzyl D-glucuronate was performed in 100 µL reaction mixtures in a protocol adapted from Sunner *et al.* (2015). The reaction mixture contained 50 µl of 100 mM sodium phosphate buffer, pH 6.0, 2 µl of 2 mM Benzyl D-glucuronate, 38 µl MiliQ water and 10 µl of appropriately diluted enzyme, at 30 °C for 30 min. Detection of glucuronic acid release was performed by using D-Glucuronic and D-Galacturonic Acid, UV method (NZYTech, Portugal) according to the manufacturer's recommendation. One unit of enzyme activity was defined as the amount of enzyme required to de-esterify 1 µmol of benzyl-D-glucuronic acid ester per minute under the specified assay conditions.

6.2.8. Determination of optimal pH and temperature

Optimum pH was determined by measuring the activity of the recombinant GEs towards benzyl D-glucuronate in 50 mM MES buffer (pH 4 to 6), 50 mM sodium phosphate buffer (pH 6 to 8), or 50 mM Tris HCl buffer (pH 8 to 10), in a final reaction volume of 100 µL at 45 °C. To determine the optimal temperature, recombinant GEs were incubated with benzyl D-glucuronate, 50 mM sodium phosphate buffer, pH 6, at different temperatures, between 20-70 °C for 30 min. Detection of glucuronic acid release was performed by using D-Glucuronic and D-Galacturonic Acid, UV method (NZYTech, Portugal) according to the manufacturer's recommendation.

6.2.9. Crystallization of *Tt*CE15

*Tt*GE15A was initially purified through IMAC followed by size exclusion chromatography according to established protocols (Bule *et al.* 2014). Purified *Tt*GE15A was buffer exchanged to 50 mM Tris HCl, pH 8.0, 200 mM NaCl, 5 mM TCEP using PD10 desalting columns (GE Healthcare). The enzyme was concentrated to 20 mg/ml and tested against a range of commercial crystallization screens using the sitting drop diffusion method. The drops were performed using the Oryx8 robotic nanodrop dispensing system and were

composed of 0.8 µL of protein at either 20 or 10 mg/mL and 0.8 µL of well solution. Fully formed, well diffracting single crystals were detected after 5 days in 0.2 M potassium nitrate, 20% w/v PEG 3350, at 293 K. Crystals were cryo-protected in well solution with 30% glycerol and flash cooled in liquid nitrogen. Protein concentration was determined with a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific) using and a molar attenuation coefficient of 84 910 M^{-1} cm^{-1}.

6.2.10. 3D structure solution

*Tt*GE15A crystal structure was solved by molecular replacement using the published structure with PDB code 6GS0 as the input model. Data was collected at beamline BioMAX of MAX IV synchrotron, using a wavelength of 0.9184 Å. It was then processed with DIALS (Clabbers *et al*. 2018), reduced with Aimless (Evans and Murshudov 2013) and phased with Phaser MR (McCoy *et al.* 2007). The initial model was subsequently subjected to alternating rounds of manual building and refinement with Coot (Emsley *et al*. 2010) and REFMAC5 (Murshudov *et al.* 2011). The final round of refinement was performed using the TLS/restrained refinement procedure using each copy of the trimer as a single group, giving the final model. Data collection and refinement statistics are shown in Table 6.1. 3D structure figures were generated using UCSF Chimera (Pettersen *et al.* 2004).

6.3. Results and Discussion

6.3.1. GEs selection and sequence analysis

To identify candidate GEs, a genome mining strategy was conducted with the BLASTP tool using characterized and putative GEs. From a database of more than 291 family 15 carbohydrate esterases, 20 enzymes were selected, covering a wide sequence diversity. The 20 enzymes included 15 bacterial and 5 fungal proteins (Table S6.2). In order to have a reference with which to compare the newly described biocatalysts, enzymes with well establish biochemical properties were included in this panel. As such, all five fungal enzymes and one of the bacterial representatives had been previously characterized (Aurilia *et al*. 2000; Pokkuluri *et al*. 2011; Tsai *et al*. 2012; Wong *et al*. 2012; Charavgi *et al*. 2013; Katsimpouras *et al*. 2014) . The panel was composed by members originating from 17 different organisms, including three from the bacteria *Teredinibacter turnerae* and two from *Cellvibrio japonicus*. It is well established that several bacterial and fungi genomes encode more than one GE genes, with several bacteria containing up to four members (www.cazy.org/CE15). Although the significance of such a large diversity of GEs expressed by a single organism remains obscure, it suggests intrinsic differences in enzyme substrate specificities.

Table 6.1. Data collection and refinement statistics of *Tt*GE15A.

Data collection statistics were obtained from the aimless log while refinement statistics were calculated using Phoenix Table 1 utility. Statistics for the highest resolution shell are between parentheses.

	***Tt*GE15A**
Data collection	
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	70.44, 224.81, 188.65
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	49.52 – 1.99 (2.06 – 1.99)
R_{merge}	0.053 (1.232)
R_{pim}	0.031 (0.703)
CC(1/2)	0.999 (0.610)
$I / \sigma(I)$	17.2 (1.4)
Completeness (%)	100 (100)
Redundancy	7.6 (7.7)
Refinement	
No. reflections	102741 (10179)
R_{work} / R_{free}	0.175 / 0.209
No. non-H atoms	10264
Protein	1274
Ligand/ion	28
Solvent	503
B-factors (Å^2)	37.98
Protein	37.15
Ligand/ion	93.23
Solvent	50.84
R.m.s. deviations	
Bond lengths (Å)	0.018
Bond angles (°)	2.04
Ramachandran plot residues	
In most favourable regions(%)	96.60
In allowed regions (%)	2.76
Outliers	0.63

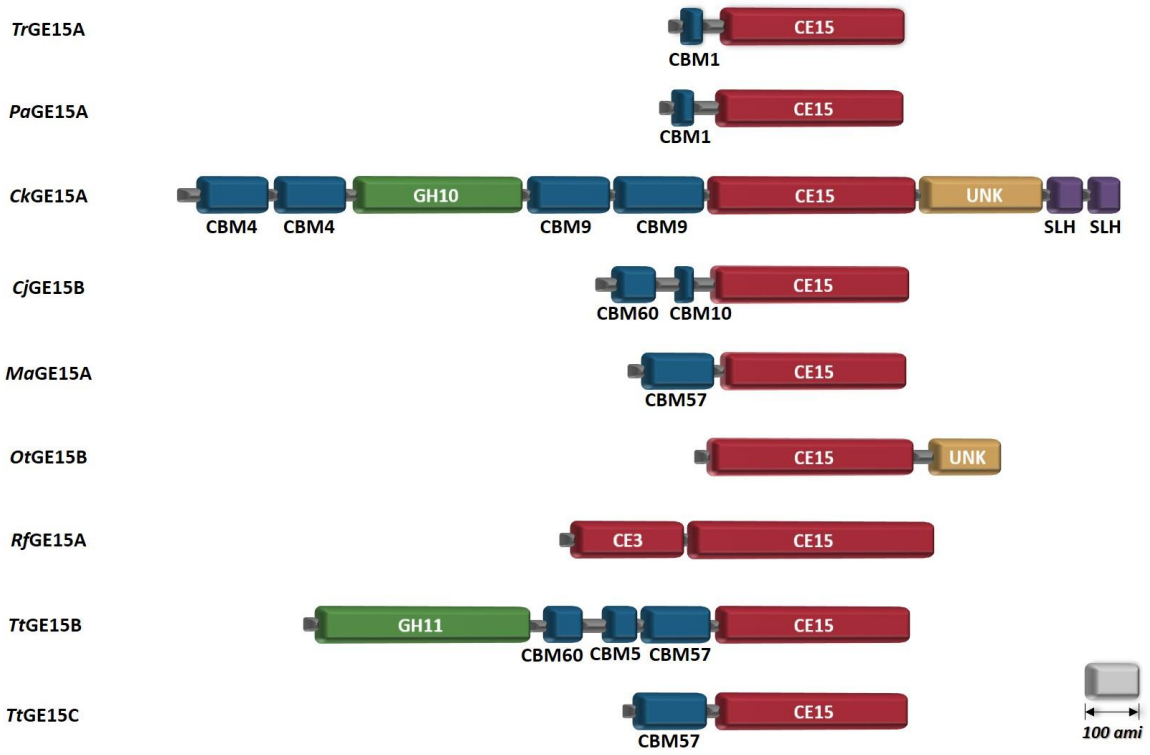


Figure 6.1. Schematic representation of modular GEs.

In addition to the GE catalytic domain, nine GEs display an architecture containing several other modules with different functions, such as GHs and CBMs.

The majority of the selected enzymes (18 out of 20) contain typical signal peptides suggesting a predominant activity in the extracellular milieu. Notably, only nine display a modular architecture containing, in addition to the GE catalytic domain, several other modules with different functions (Figure 6.1). Three of the enzymes have typical bi-functional activity containing N-terminal xylanase catalytic domains (GH10 for *Ck*GE15A and GH11 for *Tt*GE15B) or family three Carbohydrate Esterase domains (*Rf*Axe3A-GE15A) fused to the C-terminal CE15s (Figure 6.1). This type of modular organization has been previously observed in several cellulases and hemicellulases where the different domains are believed to act cooperatively upon the hydrolysis of complex polysaccharides (Cantarel *et al.* 2009). In addition to other catalytic modules, the selected modular GEs are appended to Carbohydrate-Binding Modules (CBMs) of families 1, 4, 5, 9, 10, 57 and 60 (Figure 6.1). Some of these CBMs, such as CBM1, identified at the N-terminus of fungal enzymes, are general anchors for crystalline cellulose. Although CBM10 could also target crystalline cellulose, the other CBMs are typical of type B and C and predominantly target non-crystalline ligands. The role of these CBMs in GEs function remains, however, to be characterized. In two cases, *Ot*GE15B and *Ck*GE15A, GEs are fused to modules of unknown function. Interestingly, *Ck*GE15A contains a C-terminal SLH repeat which suggests that it is

located in the bacterial cell wall. In the case of *T. turnerae* and *C. japonicas,* they encode GEs with multiple modular architectures. Both bacteria express single domain GEs, *Tt*GE15A and *Cj*GE15A, respectively, and modular GE forms appended to CBMs, namely CBM57 in *Tt*GE15C and CBM10 and CBM60 in *Cj*GE15B (Figure 6.1). In addition, *T. turnerae* contains a bi-functional enzyme containing an N-terminal GH11 xylanase catalytic domain and internal family 5, 57 and 60 CBMs (Figure 6.1). For the purpose of this study only the catalytic domains of the 20 proteins were cloned for expression.

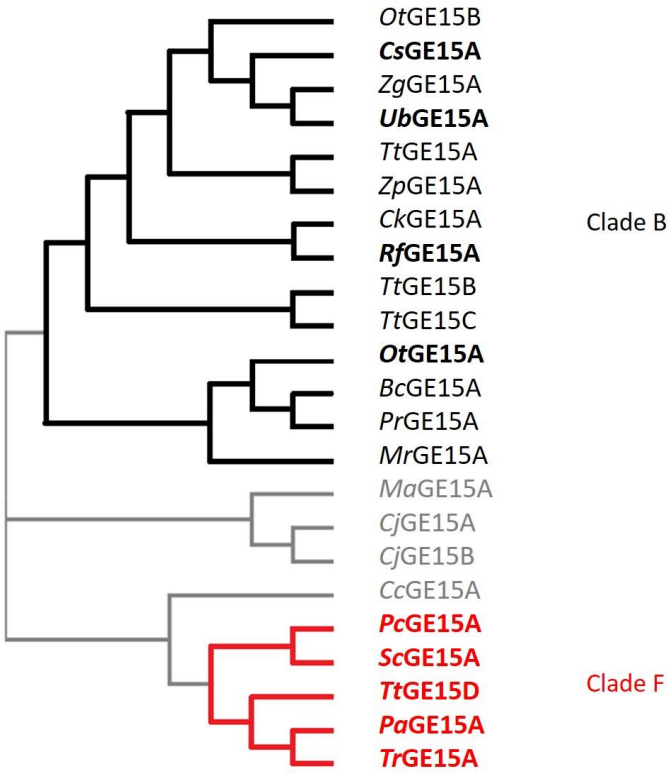


Figure 6.2. Phylogenetic relationships between the 20 GEs used in this study and 3 bacterial GEs which 3D structure were recently determined.

The characterized GEs are shown in bold; red font indicates fungal proteins; black font indicates bacterial proteins; grey font indicates bacterial proteins whose sequence diverge from either B and F clades.

The phylogenetic relation between the 20 enzymes included in this study was analysed (Figure 6.2). The analysis also incorporated three bacterial sequences for which 3D structures were recently determined (two of the fungal sequences of the database selected in this study also have a 3D structure). As reported in previous studies (De Santi *et al.* 2016; Hüttner *et al.* 2017), bacterial and fungal CE15 sequences are clustered together in separated clades, F and B, for fungal and bacterial, respectively. The two *C. japonicus* GEs, *Cj*GE15A and *Cj*GE15B, form a separate clade with *Marinimicrobium agarilyticum*

*Ma*GE15A, which is located in an intermediary position between the B and F clades. Interestingly, as reported by Monrad *et al*. (2018), the bacterial *Clostridium cellobioparum* *Cc*GE15A, is clustered together with the fungal clade, a strong evidence confirming the existence of horizontal gene transfer between fungi and bacteria. Altogether, 5 different sub-clades of bacterial enzymes were identified suggesting a larger diversity in bacterial GEs when compared with their fungal counterparts.

6.3.2. Recombinant enzyme production

The 20 putative GE-encoding genes were heterologous expressed in *E. coli* BL21 (DE3) and purified through IMAC. Out of the 20 genes, 11 recombinant GEs expressed with good yields (>9 mg/L in the eluted sample). When the crude extracts were evaluated by SDS-PAGE, nine were produced in the form of inclusion bodies (Figure 6.3). For these, sufficient levels of soluble protein were not recovered after IMAC (<5 mg of purified protein per liter of media). In addition, the molecular mass of the 11 recombinant soluble proteins was in accordance with the theoretical molecular weight (Mw) as evaluated by SDS-PAGE (Figure 6.3). The production levels for the 11 expressing recombinant GEs varied from 9-196 mg/L, and four enzymes were produced at levels above 100 mg/L.

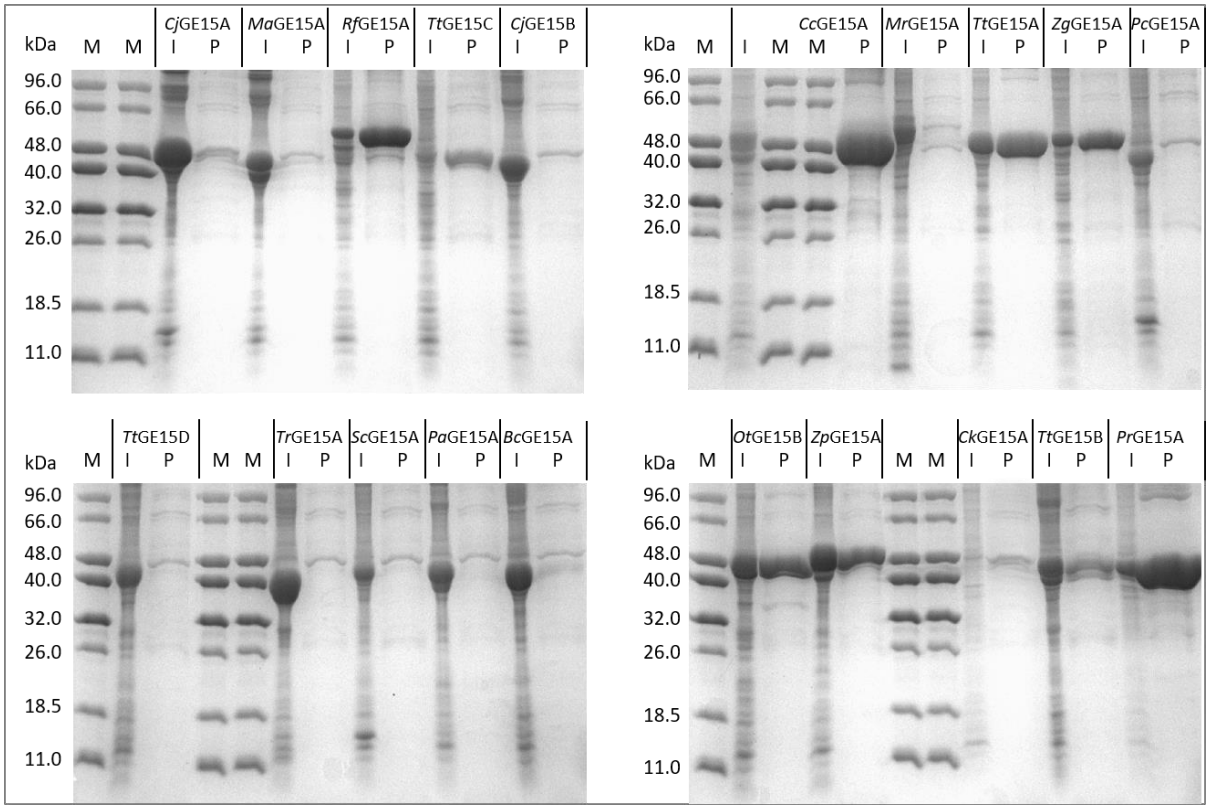


Figure 6.3. Protein expression and purity analysis of GEs by SDS-PAGE.

M, LMW Protein Marker II; I, Insoluble fraction; P, Purified fraction.

6.3.3. Biochemical properties of bacterial GEs

The catalytic activity of the recombinant GEs was determined towards benzyl D-glucuronate, by detecting the generated product using a D-Glucuronic Acid detection kit (NZYTech, Portugal). The 11 soluble GEs that were produced in *E. coli* BL21 (DE3) cells were active towards benzyl-D-glucuronic acid ester (Table 6.2). The highest specific activity (106 U/mg) was detected for *Tt*GE15A with *Rf*GE15A also displaying high activity (~95 U/mg). *Pr*GE15A and *Tt*GE15B presented medium activity (15 and 11 U/mg, respectively) and the remaining 7 enzymes displayed lower levels of activity against the artificial substrate (<10 U/mg). Benzyl D-glucuronate was also used to determine the optimal pH and temperature for catalysis for all 11 enzymes. The data, presented in Table 6.2 and Figure 6.4, revealed that all 11 GEs have an optimal pH of 6. On the other hand, seven GEs show higher activity at 30 °C (*Tt*GE15A, *Zg*GE15A, *Ot*GE15B, *Zp*GE15A, *Ck*GE15A, *Tt*GE15B and *Pr*GE15A), while the other four have an optimal temperature for catalysis of 35 °C (*Cj*GE15A and *Tt*GE15C) and 40 °C (*Rf*GE15A and *Cc*GE15A).

Table 6.2. Organism origin, production level, specific activity and optimal conditions of characterized GEs in this study.

Enzyme	Organism	Production (mg/L media)	Specific activity (U/mg)	Optimum pH	Optimum Temperature (°C)
***Ck*GE15A**	*Caldicellulosiruptor kristjanssonii*	8,50	1,61	6	30
***Zp*GE15A**	*Zunongwangia profunda*	32,28	3,04	6	30
***Cj*GE15A**	*Cellvibrio japonicus*	15,21	3,50	6	35
***Ot*GE15B**	*Opitutus terrae*	69,81	3,97	6	30
***Tt*GE15C**	*Teredinibacter turnerae*	29,25	4,94	6	35
***Cc*GE15A**	*Clostridium cellobioparum*	163,71	7,08	6	40
***Zg*GE15A**	*Zobellia galactanivorans*	110,29	8,34	6	30
***Tt*GE15B**	*Teredinibacter turnerae*	28,10	11,36	6	30
***Pr*GE15A**	*Prevotella ruminicola*	195,50	14,50	6	30
***Rf*GE15A**	*Ruminococcus flavefaciens*	64,90	94,30	6	40
***Tt*GE15A**	*Teredinibacter turnerae*	126,91	105,82	6	30

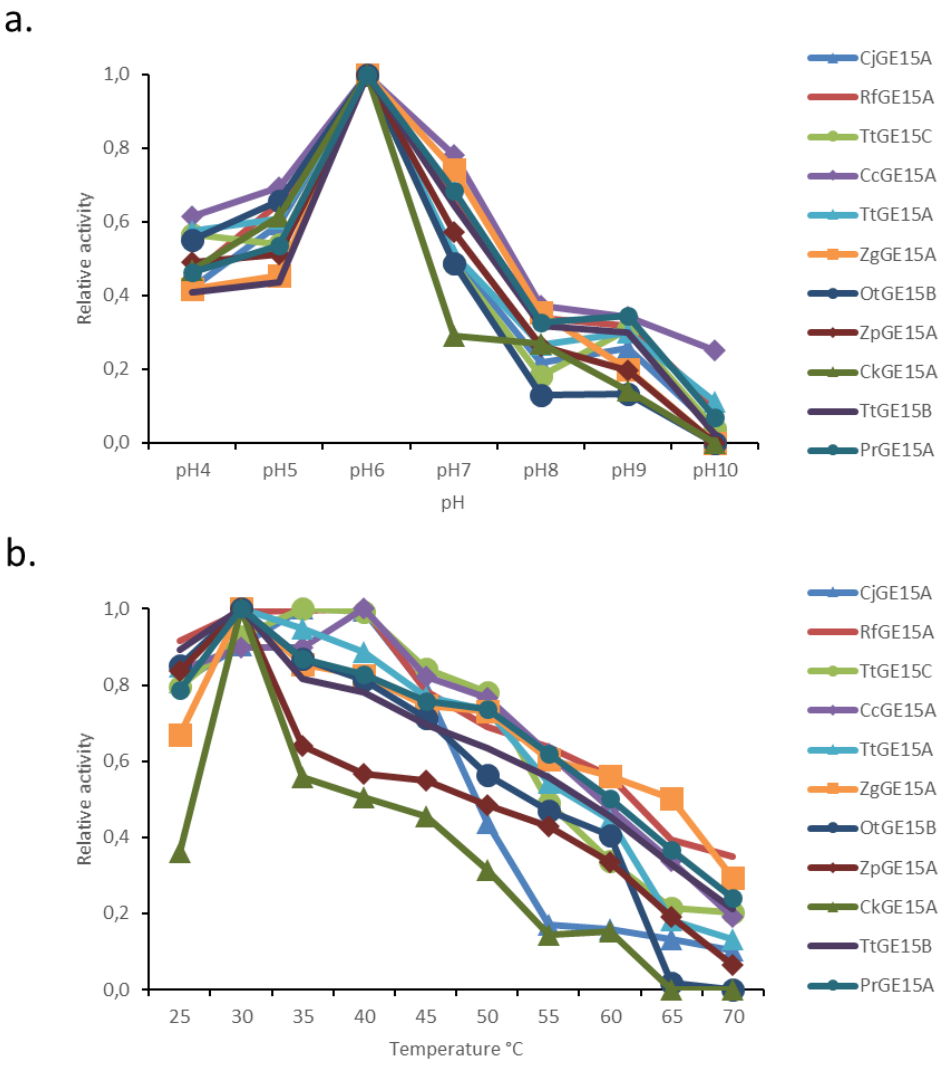


Figure 6.4. pH (a.) and temperature (b.) dependence of the tested GEs' catalytic activity. Detection of glucuronic acid release was performed by using a D-hexuronic acid detection kit at 30 °C, according to supplier's instructions. Values are presented as the ratio between the highest absorbance measured for each enzyme (at 340nm) and the absorbance measured at each pH and temperature after 30 min of incubation with benzyl D-glucuronate, followed by inactivation and uronate dehydrogenase reaction.

6.3.4. Crystal Structure of *Tt*GE15A

The structure of the most catalytically efficient GE identified in this study, *Tt*GE15A from *T. turnerae*, was solved at a resolution of 1.99Å by molecular replacement, using PDB code 6GS0 as an input model. During refinement, the structure of *Tt*GE15A was reported in the literature (Bååth *et al*. 2019) and released in PDB (code 6hsw). Three-dimensional structural comparison using the Dali Server (Holm and Rosenström 2010) revealed that the two structures are highly similar or virtually identical with a Z score of 69.0, r.m.d.s. of 0.9 Å

over 422 aligned residues. The relatively high r.m.s.d between the two structures, which share over 98% identity, can be explained by a highly flexible region between residues 175-197 and the different geometry of the two crystal forms. While 6hsw structure originated from a crystal in the trigonal $P3_121$ space group, the crystals generated in this study belonged to the orthorhombic $C222_1$. Nonetheless, using the secondary structure matching tool in coot (Emsley *et al.* 2010), two chains were overlayed with an r.m.s.d of 0.66Å, over 417 aligned residues.

According to Dali server, *Tt*GE15A/6hsw closest structural homologues are the bacterial *Ot*GE15A from *Opitutus terrae* (PDB code 6gs0), *Ub*GE15A from an uncultured bacterium (PDB code 6ehn) and *Cs*GE15A from a *Candidatus Solibacter usitatus* (PDB code 6gry) with Z scores varying from 45.1 to 43.8 and r.m.d.s. between 2.0 and 2.3 Å, over ~350 aligned residues. *Tt*GE15A/6hsw also display significant structural homology with the fungal homologues *Tr*GE15A from *Trichoderma reesei* (PDB code 3pic) and *Tt*GE15D from *Thermothelomyces thermophiles* (PDB code 4g4i) with Z scores of 33.6 and 32.5, r.m.d.s. of 2.5 and 2.6 Å over 305 aligned residues, respectively. Remote structural homologies were observed with acetyl xylan esterases, tannases and deacetylases from other CE families.

*Tt*GE15A displays the typical α/β hydrolase fold reported for members of the CE15 family (Figure 6.5a). Three *Tt*GE15A molecules were found in the asymmetric unit, with a Cα root mean square deviation below 0.2 Å. An assembly analysis with PISA calculated an interaction surface area of 509, 314 and 1167 Å between chains A-B, A-C and B-C respectively, with low hydrophobic P-values and predicted solvation energy gains (ΔG) of -3.3, -3.9 and -6.4 kcal/mol, suggesting that the biological assembly of *Tt*GE15 is likely to be monomeric. Previous structural studies on bacterial CE15s (De Santi *et al*. 2016; Bååth *et al*. 2018) identified 3 insertions, termed Regions 1, 2 and 3 (Reg1, Reg2 and Reg3, respectively), when compared with their fungal counterparts (Bååth *et al*. 2019). Reg1 and Reg3, highlighted in magenta and green in Figure 6.5b, respectively, are sequence and structurally similar in all bacterial enzymes. It is predicted that these two regions should have no implication in substrate recognition as they are distantly positioned in relation to the catalytic triad. In contrast, Reg2 of *Tt*GE15A (highlighted in blue in Figure 6.5b) is composed of two helices that diverge, both at the tertiary and primary structure levels, from the Reg2s of the other bacterial enzymes (*Ot*GE15A from *O. terrae*, *Ub*GE15A from an *uncultured bacterium* and *Cs*GE15A from a *C. S. usitatus*). Furthermore, Reg2 is located in close proximity to the putative lignin-binding site (Monrad *et al.* 2018). The observation that Reg2 displays three structural conformations - no residue insertion in fungal enzymes, an open form that is relatively more distant from the catalytic site in three bacterial enzymes and a bulkier insertion in closer proximity to the catalytic triad in *Tt*GE15A - suggests the existence of at least three different mechanisms of lignin recognition in CE15 GEs. In *Tt*GE15A Reg2

the two helices insertion is in immediate vicinity of the catalytic histidine, resulting in a more buried catalytic cleft. Putative *Tt*GE15A residues that may modulate lignin recognition are Phe174, Trp179, Phe182, Arg185, Arg190, Ile178 and Ile194. At the predicted carbohydrate-recognition site, located in the opposite face of Reg2, *Tt*GE15A displays a structural insertion not observed in any other CE15 enzyme (highlighted in yellow in Figure 6.5). This insertion is preceded by a long helix that extends the catalytic cleft and, together with Reg2 helices, contributes to further occlude the catalytic triad. The role this structural peculiarity of *Tt*GE15A has in carbohydrate recognition remains to be clarified, but it is clear that less room is available to accommodate a carbohydrate xylan backbone. In addition, the absence of surface aromatic side chains and the prevalence of several hydrophilic residues, such as Asn50 and Asp52, suggest that the biological role of this N-terminal insertion could be related with limiting the interaction with polymeric, promoting specificity for esterified xylo-oligosaccharides. This hypothesis remains, however, to be clarified.

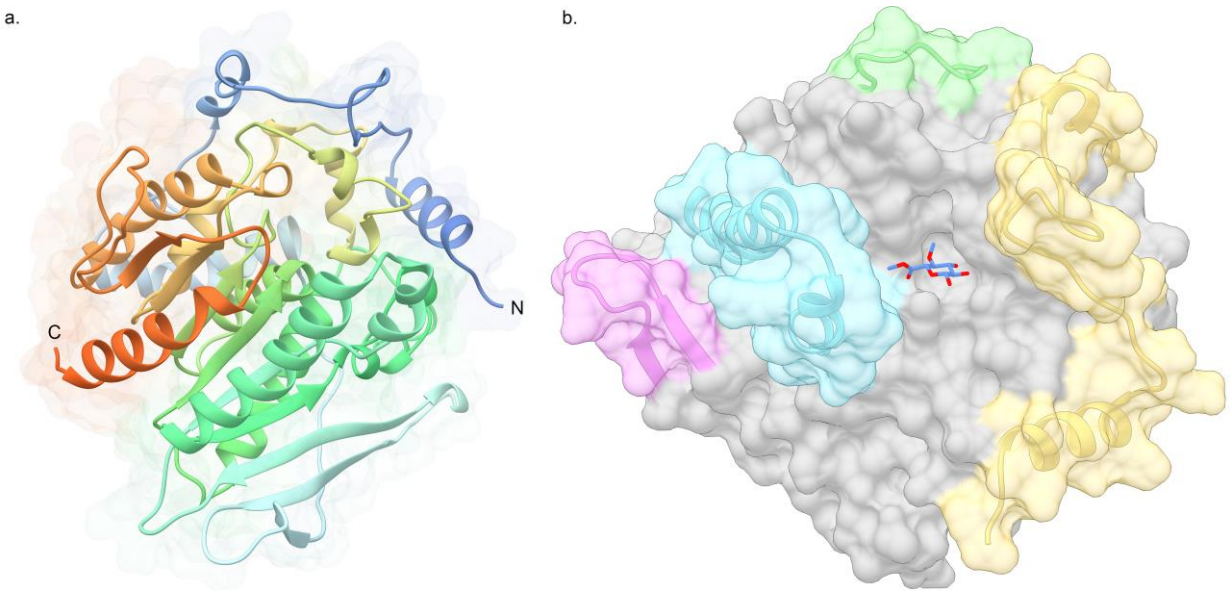


Figure 6.5. Representation of *Tt*CE15A's structure.

a. Ribbon representation of *Tt*CE15A's structure with rainbow colouring (N-terminus in blue, C-terminus in red) showing a typical (β/α)-hydrolase fold. **b.** Van der Walls surface of *Tt*CE15A coloured in grey, with the bacterial CE15 inserts highlighted in purple (Reg1), Blue (Reg2), Green (Reg3) and yellow (*Tt*CE15 specific insert). A cornflower blue coloured molecule of 4-O-methyl-β-D-glucopyranuronate in stick representation was placed in the putative active centre by overlaying with the crystal structure of PDB 4g4j, a fungal CE15 in complex with this ligand.

6.3.5. *Tt*GE15A Catalytic triad and recognition of Glucuronic Acid (GlcA)

It was previously observed that the catalytic triad is not conserved in CE15 members (Monrad *et al.* 2018). Although the catalytic serine and histidine are usually conserved across the family, the canonical catalytic glutamate is absent in several enzymes, including

*Ub*GE15A and also in *Tt*GE15A. In these two cases, Asp332 (*Ub*GE15A) and Glu374 (*Tt*GE15A) display a conserved position in close proximity to the catalytic serine and histidine residues and could eventually act as the general catalytic acid. Recent work by Baath *et al.* (2019) confirmed Glu374 as the non-canonical catalytic acid of *Tt*GE15A triad, while the conserved His427 serves as the general base and Ser281 as the nucleophile (Figure 6.6a). A constellation of residues at *Tt*GE15A catalytic site orchestrates the firm recognition of the glucuronic acid moiety. The superposition of *Tt*GE15A structure with the structure of *Tt*GE15D from *T. thermophiles* (PDB code 4g4j), in complex with the methyl esterified 4-O-methyl glucopyranuronate ligand, suggests that *Tt*GE15A Arg282 coordinates the glucuronoyl carboxylate moiety (Figure 6.6b). In addition, Lys285 binds both the 3-OH and the oxygen of the 4-O-methyl group (Figure 6.6b). The 4-O-methyl group is held in a hydrophobic environment created by the side chain of Val377. Trp376 makes a productive hydrogen bond to the glucuronoyl 2-OH, that also binds with Glu319, which makes an additional bond with 3-OH (Figure 6.6b). Finally, Glu329 firmly hydrogen bonds the glucuronoyl 1-OH and 2-OH (Figure 6.6b). The above-mentioned amino acids are highly conserved across the family, strongly suggesting that CE15s have evolved a conserved mechanism to specifically recognize glucuronic acid (Figure 6.7). In addition, there is strong evidence that Trp376 plays an important role in anchoring the xylose residue 1,2-α-linked to the O-methyl-glucuronic acid. This residue is completely conserved across the CE15 family although its role in binding the xylose backbone remains to be confirmed.

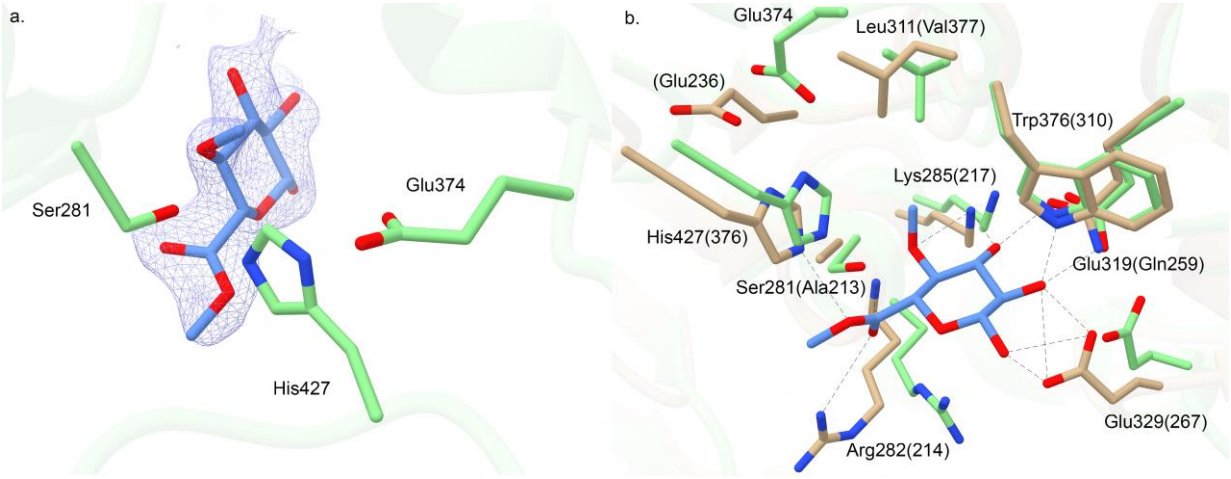


Figure 6.6. Putative catalytic triad of *Tt*CE15A's.

a. Putative catalytic triad seen near a 4-O-methyl-β-D-glucopyranuronate molecule that was placed by overlaying *Tt*CE15A structure with a fungal CE15 complex (PDB code 4g4j). The ligand is surrounded by a mesh representation of the Refmac5 maximum-likelihood σA–weighted 2Fo−Fc electron density map contoured at 1σ (0.46 electrons/A3). **b.** Detailed view of *Tt*CE15A (green) active centre overlaid with 4g4j. Important residues for ligand recognition are shown in stick representation, with the hydrogen-bond contacts between the ligand and the protein represented in black dashed lines.

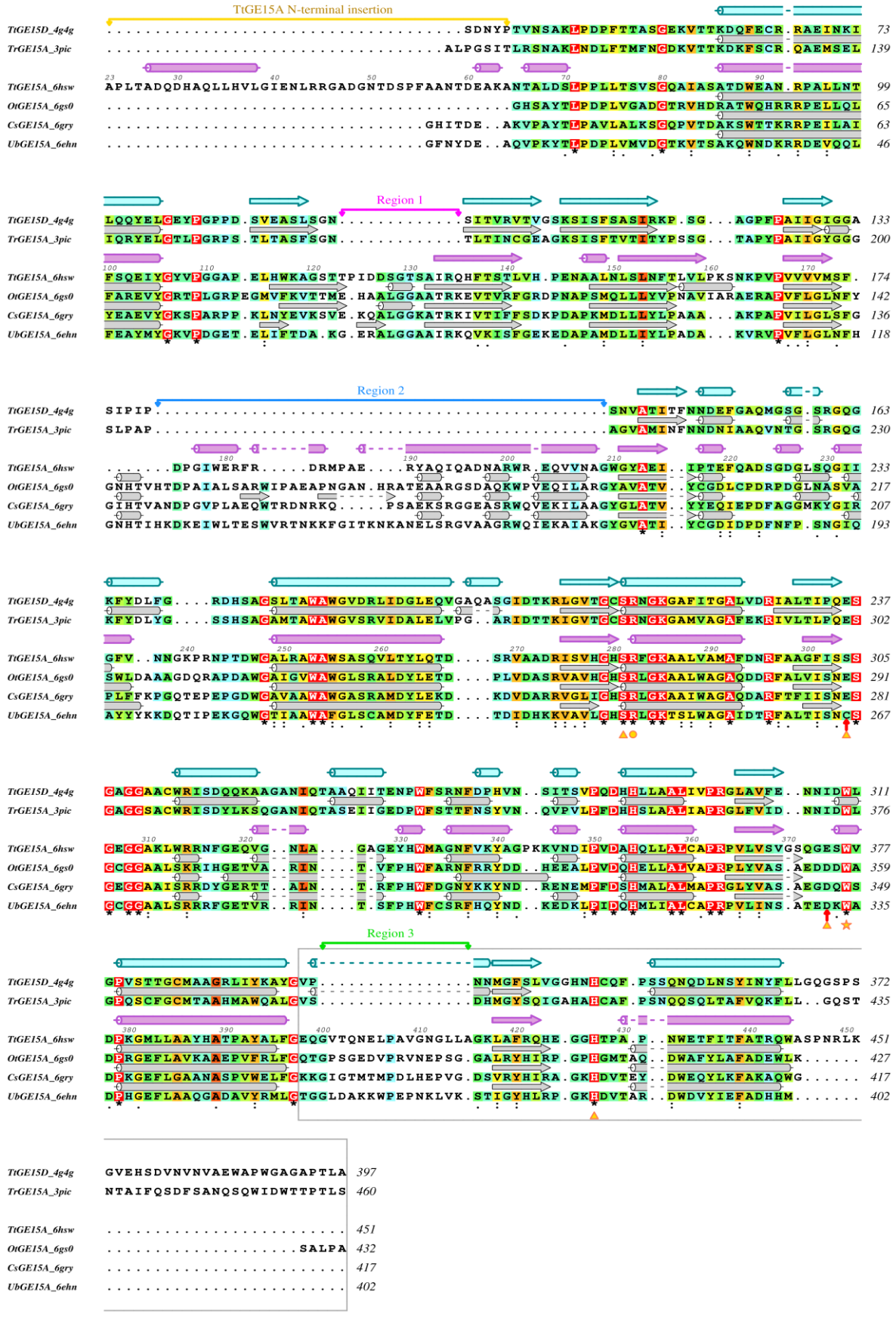


Figure 6.7. Multiple sequence alignment of GE15 members with known structures.

All sequence numbering is based on the respective PDB entries and depicted at the right end-side, while *Tt*GE15A residues are also numbered above its own sequence. Above each sequence a cartoon representation of the secondary structure is shown. Regions 1, 2 and 3, respectively coloured magenta, blue and green, are delimited, as is the N-terminal insertion sequence (coloured yellow) of *Tt*GE15A (colour coded in accordance with Figure 6.5). A yellow triangle identifies the catalytic residues, while a red arrow shows the acidic ones in the two possible locations depending on the enzyme. A yellow circle indicates the catalytic site stabilizing Arginine. A yellow star shows a relevant conserved aromatic residue for ligand interaction. The C-terminal end of the alignment, enclosed in a grey box, was based on a structural sequence alignment. The primary sequence background is coloured according to the ALSCRIPT Calcons convention, implemented in ALINE (Bond and Schüttelkopf 2009): red, identical residues; orange to blue, lowering colour-ramped scale of conservation. Below the alignment, the ClustalO consensus symbols represent the position conservation status.

6.4. Conclusions

In the present study, we have confirmed the ability of the genome mining strategy to identify novel bacterial GE encoding genes, by demonstrating that 11 out of 20 putative bacterial GEs possessed GE activity towards benzyl D-glucuronate. The remaining nine enzymes did not properly express in *E. coli*, which was not a complete surprise, at least for the fungal enzymes that were part of the initial library. *Tt*GE15A from *T. turnerae* revealed the highest levels of activity against the model substrate benzyl D-glucuronate. The structure of *Tt*GE15A revealed the classical α/β hydrolase fold observed in several esterases and typical of the CE15 family. *Tt*GE15A contains a peculiar and unique active site that is bordered by two large insertions that could limit the access to large polymeric substrates. This possibility is supported by the high activity displayed for the small artificial substrate used in this study that mimics the structure of small xylo-oligosaccharides esterified with lignin alcohols. Future work is ongoing to clarify the molecular modulators of CE15 GEs specificity.

7. SUBTLE CHANGES AT THE CATALYTIC SITE OF FERULOYL ESTERASES MODULATE ENZYME SPECIFICITY TO HYDROXYCINNAMIC ACIDS

V. Cardoso[1,2], J.L.A. Brás[2], A.L Carvalho[3], V.D. Alves[1], R.P. de Vries[4], V. Puchart[5], P. Biely[5], L.M.A. Ferreira[1,2], C.M.G.A. Fontes[1,2] and P. Bule[1,†]

[1] CIISA - Centro de Investigação Interdisciplinar em Sanidade Animal, Faculdade de Medicina Veterinária, Universidade de Lisboa, Pólo Universitário do Alto da Ajuda, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; [2] NZYTech LDA, Estrada do Paço do Lumiar, Campus Do Lumiar, 1649-038 Lisboa, Portugal;
[3] UCIBIO, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal; [4] Fungal Physiology, CBS-KNAW Fungal Biodiversity Centre & Fungal Molecular Physiology, Utrecht University, Uppsalalaan 8, 3584 CT, Utrecht, The Netherlands.; [5] Institute of Chemistry, Slovak Academy of Sciences, Dúbravská cesta 9, SK-845 38 Bratislava, Slovak Republic.

Adapted from a manuscript submitted for publication

Abstract

Feruloyl esterases (EC 3.1.1.73) are a class of highly diverse enzymes that play a major role in the hydrolysis of a multiplicity of important molecules, including plant phenols and lignocellulosic biomass. Here, from a database of >4000 putative bacterial feruloyl esterases, a library of 480 representative enzymes was produced/purified in *Escherichia coli*. The 202 soluble enzymes revealed different levels of activity against model substrates for evaluating feruloyl esterases activity, namely pNP-ferulate and four methyl esterified hydroxycinnamic acids. A detailed homology study established with the 202 enzymes revealed that substrate specificity against the artificial substrates does not correlate with phylogenetic proximity. For example, in a subset of 5 feruloyl esterases presenting a primary sequence homology greater than 70%, differences in substrate specificity against hydroxycinnamic acids were observed. The structure of *Lactobacillus crispatus* feruloyl esterase 1A (*Lc*FAE1A) was solved and revealed, similarly to the highly homologous enzyme from *Lactobacillus johnsonii*, the presence of an α/β subdomain insertion that creates an enclosed pocket for the recognition of the aromatic moiety of the substrate. Structure-function studies revealed that a single Gln-Thr change limits the enzyme capacity to recognize methyl p-coumaric and caffeic acid substrates. This work provides a clear indication of how subtle changes at the catalytic site of feruloyl esterases modulate the specificity against hydroxycinnamic acids. In addition, data presented here indicates that

feruloyl esterases developed a highly versatile catalytic machinery that may rapidly evolve to cleave ester bonds in the most diverse biological contexts.

7.1. Introduction

The plant cell wall is a complex structure composed by several polymers, including cellulose, hemicellulose, pectin and lignin, whose biodegradation is a key component in the global recycling of photosynthetically fixed carbon (Kroon *et al*. 2000). Xylan, the most abundant hemicellulose in graminaceous plants, consists of repeating 1,4-β xylose subunits that are often decorated with 1,2-α- and 1,3-α-arabinosyl residues, acetate and methyl glucuronic acid. It is now well established that hydroxycinnamic acids, primarily ferulic acid (FA), are usually esterified to the arabinose side chains of arabinoxylan promoting the crosslink between the polysaccharide and lignin through ester-ether bonds (arabinoxylan-ferulate-lignin). In addition, through the formation of diferulic acid dimers (predominantly 5,5´), ferulic acid side chains contribute to the covalent cross-linking between hetero-xylan chains (Ishii 1997; Saulnier and Thibault 1999). The formation of this inter-molecular covalent network leads to a dramatic increase in the mechanical strength of plant cell walls and acts as a barrier that limits the access of hydrolytic enzymes (Iiyama *et al*. 1994; Ishii 1997; Buanafina 2009; Oliveira *et al*. 2015). The biotechnological exploitation of lignocellulose in the context of bioethanol production or to improve the nutritive value of animal feed, requires a clear understanding of the diverse arsenal of carbohydrate- and lignin-active enzymes required to cooperatively deconstruct the plant cell wall (Wong 2006).

Feruloyl esterases (FAEs, EC 3.1.1.73), a subclass of the carboxylic acid esterases (EC 3.1.1.-) also known as ferulic acid esterases, cinnamoyl esterases or cinnamic acid esterases, are enzymes that cleave the ester bonds between hydroxycinnamic acids and carbohydrates (Williamson *et al*. 1998; Faulds 2010). Besides playing an important role in lignocellulose breakdown, these enzymes are key animal and dietary components involved in the release of absorbable bioactive monophenols, in particular ferulic, caffeic, coumaric and sinapic acids, from plant polyphenols such as chlorogenic and rosmarinic acids (Topakas *et al*. 2007; Guglielmetti *et al*. 2008). The health benefits associated with the intake of hydroxycinnamic acids reflect their ability to work as free radical scavengers, and their anti-inflammatory, anti-cancer and immune-stimulant properties (Wong 2006; Benoit *et al*. 2008; Faulds 2010; Topakas *et al*. 2010; Gopalan *et al*. 2015). FAEs adopt the α/β hydrolase fold shown by the majority of esterases and share a catalytic mechanism with serine proteases, including a Ser-His-Asp catalytic triad. In the initial steps of the mechanism, the catalytic serine is deprotonated by the conserved histidine and acts as a typical nucleophile. The nucleophilic attack by the Ser generates a tetrahedral acyl-enzyme oxyanion intermediate.

Hydrolysis of the generated intermediate is promoted by a nucleophilic water that is activated by the catalytic His (Prates *et al*. 2001).

FAEs are grouped in carbohydrate esterase family 1 (CE1) of the CAZy database (www.cazy.org). Initially FAEs were classified into four types (A, B, C and D) based on substrate specificity towards four model substrates (methyl ferulate, sinapate, caffeate, and p-coumarate) and the ability to release diferulic acid (Crepin *et al*. 2004). Briefly, FAEs classified as type A prefer ferulic and sinapic acids but are active towards methyl p-coumarate. Type B enzymes act preferentially on substrates such as p-coumaric or caffeic acids and are not active against methyl sinapate. Type C enzymes generally exhibit broad substrate specificity and are active toward all four model substrates. Finally, type D enzymes also exhibits broad substrate specificity but exhibit particularly high activity toward substrates containing acetyl residues such as 4-nitrophenyl acetate (Crepin *et al*. 2004). Although it was anticipated that differences in substrate specificity would reflect major differences in the architecture of FAEs, thus providing a simple method for classification, the ABCD classification failed to organize the growing number of FAEs in a reliable manner. More recently, Dilokpimol *et al*. (2016) suggested a new classification for fungal FAEs which were distributed across 13 subfamilies based on a new phylogenetic analysis (SF1-13). Here we have developed an extensive study aiming to characterize and classify bacterial FAEs. The work reveals why the ABCD classification is unappropriated to classify FAEs and provides a solid base to identify novel specificities in these highly diverse group of enzymes.

7.2. Materials and Methods

7.2.1. FAEs selection and Phylogenetic analysis

480 enzymes originating from 249 different microorganisms, were selected from a database of 4510 putative feruloyl esterases using the BLASTP tool, covering the widest possible sequence diversity. Signal peptides were predicted using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/; Petersen *et al*. 2011) and removed from all candidate sequences. Modular architecture determination and domain assignment were performed with dbCAN (Yin *et al*. 2012). Molecular weights and molar attenuation coefficients were calculated by ExPASy ProtParam tool (http://www.expasy.ch/tools/protparam.html). Construction of the phylogenetic tree and multiple sequence alignments were performed with MEGA software (Tamura *et al*. 2013) and Multiple Sequence Comparison by Log-Expectation (Clustal Omega) (https://www.ebi.ac.uk/Tools/msa/clustalo/), respectively.

7.2.2. Chemicals

Culture medium and competent cells were purchased from NZYTech (Portugal). Similarly, all molecular biology kits, including NZYGelpure, NZYEasy Cloning & Expression and NZYMiniprep kits were provided by NZYTech (Portugal). Methyl esterified substrates (MFA, MCA, MpCA and MSA) and 4-nitrophenol palmitate (pNPP) were purchased from Carbosynth (Compton, UK). P-nitrophenyl ferulate (pNP-Fe) was provided by Taros Chemicals (Dortmund, Germany). Methyl gallate and all remaining chemicals were purchased from Sigma–Aldrich (Sigma– Aldrich, St. Louis, MO).

7.2.3. Strains, media and vector

Escherichia coli strain NZY5α was for plasmid isolation while strain BL21(DE3) was used for bacterial protein expression (NZYTech, Portugal). The two *E. coli* strains were cultured in Luria broth and Auto-Induction medium (NZYTech, Portugal), respectively, supplemented with 50 µg/mL kanamycin. All bacterial genes were cloned in pHTP1 expression vector (NZYTech, Portugal).

7.2.4. Polymerase chain reactions (PCR) and Gene synthesis (GS)

Genes encoding 480 FAEs from 249 different microorganisms were obtained by PCR or through gene synthesis (GS). Genes isolated through PCR resulted from the amplification of the nucleic acids with the KOD Hot Start DNA Polymerase (Novagen, United States) using bacterial genomic DNA as template and the primers shown in Table S7.1. PCR reactions were performed according to the manufacturer's protocol. After amplification, the assembled PCR products were purified using NZYGelpure 96 well plate kit in a Tecan workstation (Switzerland). Genes obtained by GS were initially designed for optimised expression in *E. coli*, using the ATGenium codon optimization algorithm (NZYTech, Portugal). The 339 synthetic genes were produced by NZYTech following established protocols (Sequeira *et al*. 2017). The databank describing both DNA and amino acid sequences of the 480 FAE genes and respective enzymes, their predicted molecular mass, molecular architecture and origin is presented in Table S7.2.

7.2.5. Cloning, transformation and sequencing

Purified PCR and gene synthesis products were directly cloned into pHTP1 expression vector using the NZYEasy Cloning kit, according the protocol reported in Turchetto *et al*. (2017) and Sequeira *et al*. (2017). Following the cloning reaction, recombinant plasmids were used to transform *E. coli* NZY5α competent cells using a high-throughput method. The plasmids were purified from the bacterial pellets using NZYMiniprep

96 well plate kit and subsequently sequenced. Sequences were determined by Sanger Sequencing to confirm that no mutations accumulated during the amplification/synthesis.

7.2.6. High-throughput protein expression and purification by nickel affinity chromatography

All steps were based in the protocol described by Turchetto *et al*. (2017), with few modifications. The 96 recombinant pHTP1 FAEs were used to transform *E. coli* BL21 (DE3) cells. Recombinant strains were cultured in DW24 with 5 mL of Auto-Induction LB medium and grown over 24 h at 25 °C. Cells were harvested by centrifugation at 1500 ×g for 15 min at 4 °C. The cell pellets were resuspended in 1 mL of NZY Bacterial Cell Lysis Buffer (NZYTech, Portugal) with 4 µg/mL of DNAse and 100 µg/mL of Lysozyme. The proteins were then purified by immobilized metal ion-affinity chromatography (IMAC). Briefly, 1 mL of crude cell lysates were incubated with 200 µL Ni^{2+} Sepharose 6 Fast flow resin (GE Healthcare, 17-5318-02) with bound Nickel and then transferred into 96-well filter plates (20 µm) (Macherey–Nagel). The wells were washed twice with 1000 µL buffer A (50 mM NaHepes, 1 M NaCl, 10 mM Imidazole, 5 mM $CaCl_2$, pH 7,5) followed by one wash with 1000 µL buffer B (NaHepes, 1 M NaCl, 35 mM Imidazole, 5 mM $CaCl_2$, pH 7,5) and, finally, the proteins were eluted with 300 µL of elution buffer (NaHepes, 1 M NaCl, 300 mM Imidazole, 5 mM $CaCl_2$, pH 7,5). Purified recombinant proteins were analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (14% acrylamide) according to the general instructions. Total protein concentration was determined using a Bradford protein assay (NZYTech, Portugal) against a standard curve of bovine serum albumin (BSA) (NZYTech, Portugal).

7.2.7. High-throughput enzyme activity assays

The different activity assays were performed as described below. Unless otherwise stated, one unit of enzyme was defined as the amount of enzyme that produced 1 mmol of product per minute under the assay conditions.

7.2.7.1. p-Nitrophenyl Ferulate (pNP-Fe)

Feruloyl-esterase activity was measured by determining the rate of hydrolysis of p-Nitrophenyl-ferulate (Taros Chemicals, Germany) to p-Nitrophenol (pNP) at 37 °C for 22 min. The reaction mixture contained 245 µl of 100 mM sodium phosphate buffer, pH 8, 2.5 µl of 6 mM pNP-ferulate, and 2.5 µl of appropriately diluted enzyme. The increased absorbance at 410 nm was measured with a recording spectrophotometer (Multiskan Go, Thermo Fisher Scientific, Finland). The extinction coefficient for pNP under these assay conditions was 13.7×10^3 $M^{-1}.cm^{-1}$.

7.2.7.2. Methyl esterified substrates (MFA, MCA, MSA or MpCA).

The activity of the several FAEs on esterified substrates was determined by incubating the enzymes with methyl ferulate (MFA), caffeate (MCA), p-coumarate (MpCA), sinapate (MSA) and continuously measuring the mixtures absorbance at 340 nm. The typical reaction mixtures contained 242.5 µl of 100 mM sodium phosphate buffer pH 8, 0,15 mM methyl esterified substrates and 5 µl of appropriately diluted enzyme. The absorption spectra of the methyl esters and their hydrolysis products were monitored for 22 min, at 37 °C. The extinction coefficient for methyl substrates using 96 well plate was 10.65 $mM^{-1}.cm^{-1}$.

7.2.7.3. Tannase Activity

Tannase activity was assayed using the protocol described by Sharma *et al*. (2000) and Srivastav and Kar (2009), with some adaptations. The reaction mixture containing 6,25 µL of 100 mM methyl gallate, 31,25 µl 100 mM phosphate buffer, pH 6.0 and 25 µL of enzyme was incubated at 30 °C for 10 min. 75 µL of methanolic rhodanine (0.667 % w/v) was then added to the mixture. After 5 min incubation at room temperature, 100 µL of 0.5 M KOH was added. A control was performed by adding enzyme after the KOH. Finally, the reaction mixture was diluted with 1.0 mL distilled water and incubated at 30 °C for 10 min. Absorbance was recorded at 520 nm.

7.2.7.4. Lipase Activity

Lipase activity was determined by incubating the enzymes with p-Nitrophenol Palmitate (pNPP) at 30ºC according to the protocol described by Qamsari *et al*. (2011). The reaction mixture was composed of 175 µl pNPP solution and 75 µl of lipase solution. The pNPP solution was prepared by adding the solution A (0.001 g pNPP in 1 mL 2-propanol) into solution B (0.01 g gum arabic, 0.02 g Sodium deoxycholate, 50 µl Triton X-100 and 9 mL of 50 mM Tris-HCl buffer, pH 8) with stirring until all was dissolved. The absorbance of the mixture was continuously measured at 410 nm over the course of the first 2 min of reaction.

7.2.8. Crystallization of *Lc*FAE1A

*Lc*FAE1A was initially purified through IMAC followed by size exclusion chromatography according to established protocols (Bule *et al*. 2014). Purified *Lc*FAE1A was buffer exchanged to 50 mM Tris pH 8.5 with 200 mM NaCl and 1mM DTT using PD10 desalting columns (GE Healthcare). The enzyme was concentrated to 47.5 mg/ml and tested against a range of commercial crystallization screens using the sitting drop diffusion method. The drops were performed using the Oryx8 robotic nanodrop dispensing system and were composed of 1 µL of protein at either 20 or 10 mg/mL and 1 µL of well solution. Fully formed, well diffracting single crystals were detected after 5 days in 2.0 M Ammonium sulfate 0.1 M

BIS-Tris pH 5.5 at 293 K. Crystals were cryo-protected in well solution with 30% glycerol and flash cooled in liquid nitrogen. Protein concentration was determined with a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific) using and a molar attenuation coefficient of 16390 M^{-1} cm^{-1}.

7.2.9. 3D structure solution

*Lc*FAE1A crystal structure was solved by molecular replacement using the published structure with PDB code 3PF8 as the input model. Data was collected to completeness in house with an X-ray diffractometer (IμS 3.0 microfocus D8 Venture with copper K_α radiation of 1.5418 Å wavelength), coupled to a CMOS Photon II detector, cooled at 110 K with a gaseous nitrogen stream (Oxford Cryosystems 600 Series). The crystal diffracted to a maximum resolution of 2.3 Å. Intensities were processed, integrated and scaled using PROTEUM3 software pipeline (Bruker AXS 2015), converted to observed structure factors using SCALEPACK2MTZ and TRUNCATE from the CCP4 suite (CCP4 1994) and phased with Phaser MR (McCoy *et al*. 2007). The initial model was subsequently subjected to alternating rounds of manual building and refinement with Coot (Emsley *et al*. 2010) and REFMAC5 (Murshudov *et al*. 2011). The final round of refinement was performed using the TLS/restrained refinement procedure using each copy of the trimer as a single group, giving the final model. Data collection and refinement statistics are shown in Table 7.1. *Lc*FAE1A 3D structure figures were generated using UCSF Chimera (Pettersen *et al*. 2004).

7.3. Results and Discussion

7.3.1. Discovery of novel bacterial FAEs

In order to identify novel FAE candidates, a genome mining strategy was performed through a BLAST analysis using characterized FAEs as search inputs against several bacterial genomes. Recovered duplicated and truncated sequences were removed from the analysis. Over 4000 putative bacterial FAE sequences were identified and 480 enzymes (~10%) covering the widest primary sequence diversity were selected for biochemical characterization. Globally, the selected enzymes originated from 249 different bacterial organisms. The majority of the selected enzymes contained typical signal peptides suggesting a predominant activity in the extracellular milieu. Signal peptides and other functional appended domains, i.e. carbohydrate-binding modules and xylanase/esterase catalytic domains, were removed from the candidate sequences in order to produce exclusively the FAE catalytic module of each enzyme.

Table 7.1. Data collection and refinement statistics of *Lc*FAE1A.

Data collection statistics were obtained from the aimless log while refinement statistics were calculated using Phoenix Table 1 utility. Statistics for the highest resolution shell are between parentheses.

	***Lc*FAE1A**
Data collection and Processing	
Source	IμS 3.0 (Bruker)
Detector	CMOS Photon II (Bruker)
Wavelength (Å)	1.5418
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	96.93, 126.26 138.98
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	24.07 – 2.35 (2.45-2.35)
R_{sigma}	0.093 (0.491)
I / σ*(I)*	9.30 (1.77)
No. of observed reflections	35812 (4053)
Completeness (%)	99.6 (97.9)
Redundancy	9.47 (5.3)
Model building and Refinement	
No. reflections	35678 (3429)
R_{work} / R_{free}	0.21/0.28
No. non-H atoms	6121
Protein	5916
Ligand/ion	17
Solvent	188
B-factors (Å^2)	16.09
Protein	15.88
Ligand/ion	36.35
Solvent	21.02
R.m.s. deviations	0.017
Bond lengths (Å)	0.017
Bond angles (°)	2.17
Ramachandran plot residues	
In most favourable regions (%)	95.47

7.3.2. Recombinant enzyme production

Genes encoding 480 bacterial FAEs were isolated through PCR or gene synthesis and cloned into the bacterial expression vector pHTP1. The resulting recombinant plasmids were used to transform *E. coli* BL21 (DE3) and, after expression the respective recombinant enzymes, were purified in a high-throughput automated platform through IMAC. The molecular integrity of purified proteins was analysed through SDS-PAGE and protein yield was quantified through the Bradford assay. Significant yields (>10 mg of pure protein per litre of media) were obtained with 372 enzymes. In general, the molecular mass of the 372 soluble proteins was in accordance with the theoretical molecular weight (Mw) as evaluated by SDS-PAGE (Figure 7.1). In contrast, 108 enzymes were produced at residual levels in *E. coli* (<10 mg of pure protein/L). Further analysis of the crude extracts of the recombinant *E. coli* cells that failed to express the recombinant proteins at significant levels confirmed that the majority of these 108 proteins were produced in the form of inclusion bodies (data not shown). In general, expression levels of bacterial FAEs in *E. coli* were high; 140 enzymes were produced with yields of 10-100 mg/L, 137 enzymes were obtained at 100-300 mg/L and 95 displayed a protein yield greater than 300 mg/L. The enzyme that displayed the best expression was HA045 from *Thermobacillus xylanilyticus* with a protein yield of ~700 mg/L. All data concerning the expression and production of the 480 bacterial FAEs is resumed in Table S7.2.

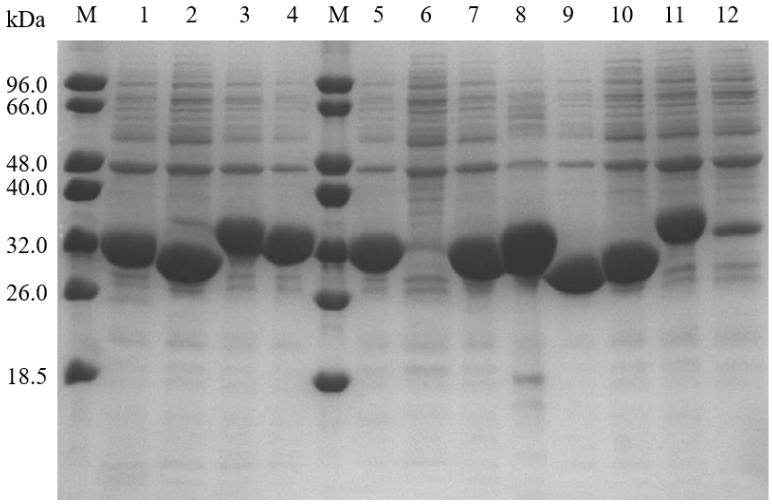


Figure 7.1. Exemplification of molecular integrity and degree of purity of recombinant FAEs produced in this study.

The figure represents the SDS-PAGE analysis of 12 representative recombinant FAEs purified following a high-through put protocol. Lane M, LMW protein marker; Lanes 1-12: LA528, LA625, HA067, HA113, HA592, IN183, HA530, LA645, HA547, HA542, HA206 and HA041.

7.3.3. Biochemical properties of bacterial FAEs

The catalytic activity of the 372 soluble FAEs was initially determined towards pNP-Ferulate and the methyl esterified substrates MFA, MCA, MSA and MpCA, under standard assaying conditions. Out of the 372 enzymes analysed, 202 revealed activity against at least one of the five substrates tested, suggesting that 170 enzymes were either inactive or should display another enzyme activity unrelated with FAE (Table S7.2). Esterases are quite a diverse group of enzymes and this observation is not completely surprising considering the limited number of substrates used in these initial experiments. Taking into consideration the phylogenetic proximity between FAEs and lipases and tannases, these two enzyme activities were tested. Thus, out of the 170 enzymes that were inactive for the hydroxycinnamic acid substrate derivatives, 14 revealed lipase activity while 6 displayed properties typical of tannases, revealing that indeed part of these proteins are active biocatalysts. Nevertheless, inspection of the primary sequence of the initial 170 inactive enzymes revealed that several of those do not possess conserved catalytic triad residues characteristic of bacterial esterases, suggesting the absence of catalytic activity (data not shown). The biological rational for the microbial expression of non-active FAEs remains to be elucidated.

Within the 202 active FAEs, 199 displayed activity on pNP-ferulate while the remaining three enzymes were exclusively active on methyl esterified substrates (Table S7.2). Among the methyl ester substrate specific enzymes MT248 is active on p-coumaric and caffeic acids, MT502 displays activity on ferulic acid and MT630 on caffeic and ferulic acids. Of the 199 enzymes that display activity for pNP-ferulate, 94 are not active on any of the methyl esterified substrates tested. The FAE activity of these 94 enzymes was relatively modest (average 1,8 U/mg measured on pNP-ferulate) with some also displaying lipase (26 enzymes) or tannase (7 enzymes) activities in addition to the FAE activity (Table S7.2). In order to develop a rational approach to analyse the data of such a large number of enzymes, the remaining 105 FAEs that displayed activity on pNP-ferulate and at least one of the methyl ester substrates were classified in two groups. The high activity (HA) group contained all enzymes showing activity on pNP-ferulate above 1 U/mg of protein, while the low activity (LA) group contained members displaying an activity below 1 U/mg on the same substrate. A small portion of the enzymes initially organized into the LA group were re-classified as HA as they demonstrated high activity (>15 U/mg of protein) on at least one of the methyl ester substrates. The 74 enzymes classified into the HA group displayed an average activity on pNP-ferulate of ~10 U/mg, while the 31 enzymes of the LA group had an average activity of ~0,45 U/mg (Table S7.2).

7.3.4. Phylogenetic tree of active FAEs

To clarify the primary sequence similarities of the soluble and active FAEs identified in this study, a phylogenetic tree including the 202 enzymes was constructed (Figure 7.2). The primary sequences of the 3 FAEs that are exclusively active on methyl ester substrates (termed MT enzymes), the 94 enzymes that attack pNP-ferulate but not the methyl substrates used in this study (termed PP enzymes) and the 31 and 74 LA and HA FAEs, respectively, totalizing 202 proteins, were aligned using Clustal Omega. Phylogenetic analysis was performed using the neighbour-joining method by the Molecular Evolutionary Genetics Analysis software (Kaur *et al*. 2018). The sequences of bacterial FAEs with reported three dimensional structures (PDBs 3QM1, 1JT2, 1GKL, 5CXX, 2WTN, 5VOL and 5YAE) were also included in the analysis. In addition, fungal FAEs representing the 13 sub-families described by Dilokpimol *et al*. (2016) were also incorporated to reveal the relationships of bacterial and fungal enzymes. The phylogenetic analysis developed in this study organized the 202 active bacterial FAEs into 3 major clades, termed B1, B2 and B3 (Figure 7.2). Clades B1 and B2 were further divided into two sub-clades each, which were termed B11 and B12, for B1 sub-clades, and B21 and B22, for B2 sub-clades (Figure 7.2). Fungal enzymes of sub-families SF1, SF2, SF3, SF4, SF9, SF10 and SF11 were clustered together and are phylogenetic related with the bacterial clade B1 (Dilokpimol *et al*. 2016). In contrast, fungal enzymes of SF5 and SF6 grouped separately while SF11 and SF12 are integrated with the enzymes of clade B2. Members of SF7 and SF8 are dispersed within bacterial enzymes of group B3. These observations suggest that there is considerable horizontal gene transfer between bacteria and fungi and this is particularly obvious for fungal enzymes of subfamilies SF7 and SF8 and clade B3 bacterial enzymes.

7.3.5. Relation between phylogenetic origin and FAE substrate specificity

In order to extrapolate potential correlations between biochemical function and sequence homology, the substrate specificities of the 202 soluble FAEs were compared with the family classification and the phylogenetic organization described above. In the past, FAE classification was based not only on primary sequence relations but also on enzyme biochemical properties, namely in relation with enzyme capacity to attack methyl ester substrates and diferulic dimers. The ABCD FAE classification scheme has recently been questioned. The systematic biochemical analysis of a large number of representative bacterial enzymes, as developed in this study, allows probing the real validity of the system. A preliminary analysis reveals that the majority of HA enzymes are classified either in FAE groups B11 and B3 (52 enzymes out of 74) but there is no obvious clustering of LA or MT enzymes, although PP enzymes predominate in clade B2 and sub-clade B12 (Figure 7.2).

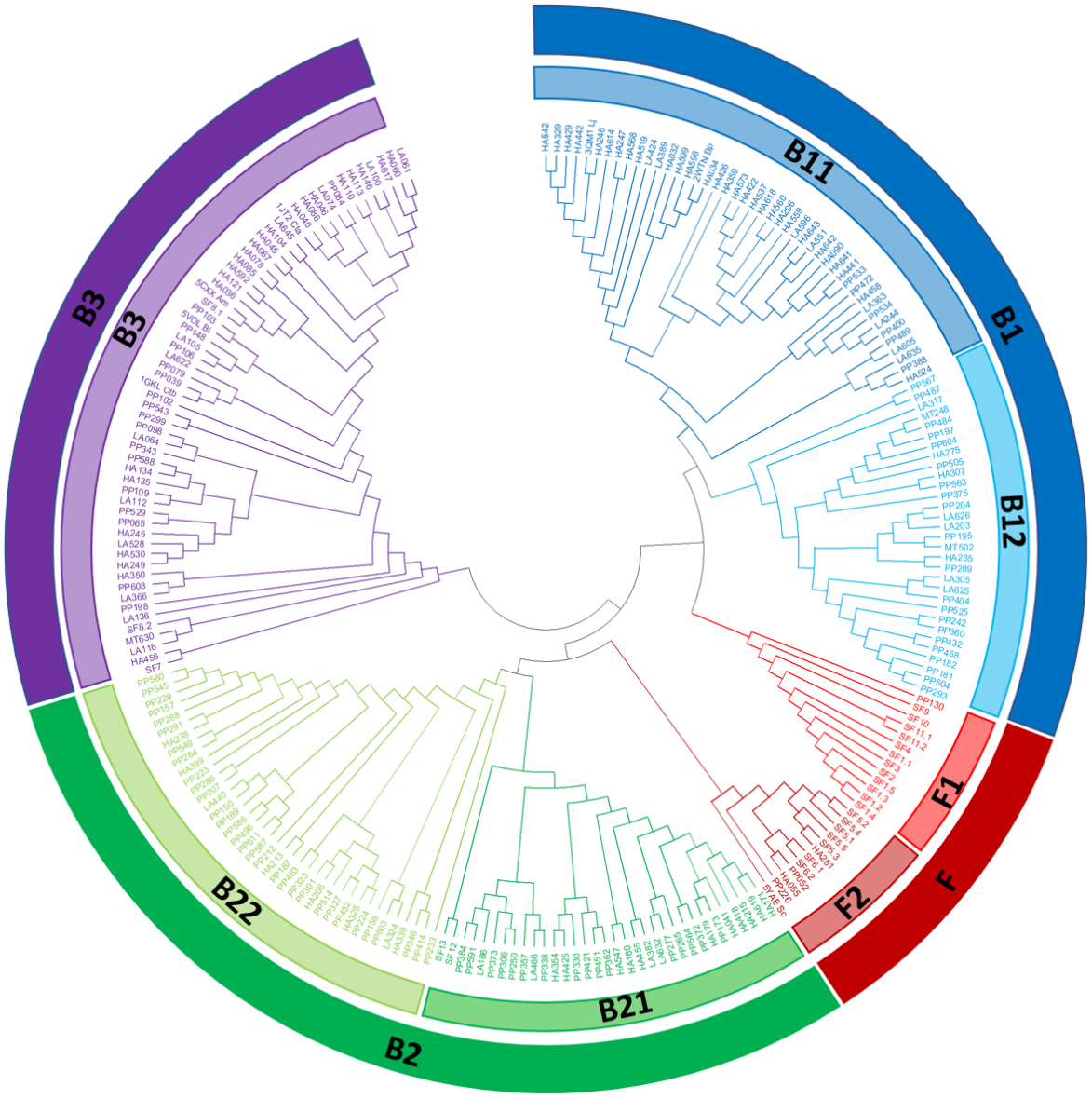


Figure 7.2. Phylogenetic relationships among the 202 active bacterial FAEs characterized in this study and representative fungal FAEs of subfamilies 1 to 13 (SF1-SF13).

For some of the fungal subfamilies more than one member were included. Enzymes with known 3D structure were also incorporated in the analysis and comprise FAEs with the following PDB codes: 3QM1, 1JT2, 1GKL, 5CXX, 2WTN, 5VOL and 5YAE. Bacterial FAEs clustered into three clades (B1 to B3), while fungal enzymes in two (F1 and F2). Clade B1 (blue); Clade F, characterized fungal FAEs (red), Clade B2 (green); Clade B3 (purple).

The substrate specificity of HA enzymes is presented in Table 7.2, which also highlights the enzyme phylogenetic interactions identified in Figure 7.2. Enzymes of the B11 group display a wide specificity for the methyl ester substrates since the majority can cleave MFA, MCA, MSA and MpCA. Notable exceptions to this general trend are the cases of HA247, HA441, HA524 and HA641, which specifically cleave MpCA (HA524), MCA (HA641)

or MFA (HA247 and HA441). Some enzymes display a preference for MFA and MSA (HA246) while others cleave exclusively the smaller substrates MpCA and MCA (HA032 and HA618). In general, within HA enzymes, those that are members of B11 group display the highest catalytic activities against the artificial substrates, suggesting a potential preference for smaller ligands (Table 7.2). B12 enzymes seem to display a preference for MpCA except for HA235 which can also cleave MFA and MSA. B21 and B22 enzymes do not display a common trend in terms of substrate specificity, although the majority of B21 enzymes can specifically cleave MpCA, MCA and MFA, while B22 FAEs HA206 and HA339 attack exclusively MCA, MFA and MSA. Finally, B3 enzymes display wide substrate specificities, cleaving either all four methyl ester substrates or exclusively MCA, MFA and MSA. When compared with the other bacterial groups, members of B3 display generally a lower catalytic activity. Taken together, the data suggest that although some common trends are possible to decipher, there is no clear relation between enzyme grouping and FAE substrate specificity.

Table 7.2. Specific activity of characterized FAEs in this study.

Pink filling demonstrates a value of activity that is above the mean of the column.

Enzyme	Sub-family	pNP Ferulate (U/mg)	MpCA (U/mg)	MCA (U/mg)	MFA (U/mg)	MSA (U/mg)
HA542	B11	24,5	26,6	29,3	32,8	6,4
HA329	B11	3,9	3,3	2,4	6,0	0,4
HA429	B11	12,8	32,5	38,3	34,7	4,9
HA442	B11	5,4	33,2	36,4	27,8	1,6
HA246	B11	41,9			67,2	6,8
HA614	B11	1,1	8,0	20,8	3,2	0,2
HA247	B11	12,7			6,3	
HA568	B11	0,9	2,4	20,2	2,9	0,3
HA519	B11	1,5	56,6	67,4	53,8	2,2
HA032	B11	20,7	41,3	8,6		
HA599	B11	2,8	2,7	1,6	18,8	7,2
HA598	B11	0,8	13,0	8,3	53,6	14,7
HA034	B11	4,0	14,8	8,6	48,3	12,4
HA426	B11	94,3	221,3	5,6	2,2	111,3
HA359	B11	30,1	3,3	9,8	41,4	22,2
HA573	B11	28,6	44,3	58,5	53,9	1,0
HA422	B11	32,1	44,2	57,0	46,4	1,4
HA537	B11	3,3	21,1	38,4	20,5	1,0
HA618	B11	1,4	4,2	6,7		
HA560	B11	0,8			34,6	
HA296	B11	3,2	37,5	58,7	60,2	3,3
HA559	B11	0,9	34,7	44,9	1,7	
HA643	B11	3,2	0,5	0,2	3,4	0,7
HA642	B11	77,9	49,0	61,4	72,6	43,5
HA090	B11	9,2	32,1	15,9	23,4	6,2
HA641	B11	2,8		3,6		
HA441	B11	20,8			2,9	
HA458	B11	1,8	4,1	0,8	11,5	

HA524	B11	1,3	0,8			
HA275	B12	3,7	6,5			
HA307	B12	0,8	17,0			
HA235	B12	3,6	19,2		6,7	1,8
HA171	B21	6,0	28,8	11,8	1,0	
HA619	B21	0,6	33,3	11,4	3,6	
HA215	B21	3,0	7,0	2,3		
HA418	B21	0,8	20,4	16,1	2,0	
HA041	B21	2,0	67,7	52,9	20,7	
HA179	B21	5,9	2,4	2,6		
HA455	B21	8,7			7,5	
HA160	B21	3,1			7,7	2,3
HA547	B21	31,0	25,0	38,3	32,4	16,6
HA425	B21	1,8	8,2	19,6	2,9	
HA354	B21	7,1	10,7	1,5	4,0	
HA339	B22	12,9		1,9	19,2	2,5
HA325	B22	1,3	12,2	6,3		
HA206	B22	2,2		6,0	12,0	3,1
HA213	B22	1,5	21,2			
HA399	B22	5,8	0,2			
HA238	B22	1,4	30,1			
HA456	B3	1,9		2,8		
HA350	B3	4,1	2,8	8,2	6,9	
HA249	B3	11,6	7,4	6,2	7,4	2,8
HA530	B3	4,7		2,4	5,3	
HA245	B3	1,9	15,5	16,5	15,2	
HA135	B3	1,5		4,4		
HA134	B3	1,1	0,7	0,4		
HA036	B3	3,1	40,4	12,9	67,2	45,3
HA121	B3	2,3	0,3	0,2	16,4	9,5
HA592	B3	3,4	0,7	0,2	10,2	5,7
HA085	B3	8,6	1,1	0,4	6,0	7,7
HA078	B3	39,9	2,3	0,4	32,7	15,0
HA067	B3	5,0			0,5	0,4
HA045	B3	50,9	18,2	1,1	29,3	27,5
HA104	B3	20,0	5,1	2,1	9,7	15,8
HA040	B3	1,8			0,7	0,4
HA086	B3	3,5		0,1	0,5	0,7
HA046	B3	1,3		0,2	1,2	1,9
HA110	B3	1,9	0,2	0,1	0,7	1,2
HA113	B3	3,2	0,2	0,1	0,6	1,3
HA146	B3	1,7		0,2	1,2	1,6
HA617	B3	4,9	0,3	0,3	1,7	4,0
HA060	B3	2,0		0,3	1,9	2,0
HA251	F2	1,6	8,5	10,7	0,8	
HA055	F2	4,0	6,8	3,0	2,1	

Enzymes that attack large polymeric recalcitrant plant cell wall carbohydrates, such as cellulases and hemicellulases, usually present a modular architecture where catalytic domains, generally one but sometimes two, are appended to one or more non-catalytic carbohydrate-binding modules (CBMs) (Guillén *et al*. 2010). Although the majority of the 480 FAEs selected for this study present a non-modular architecture, 27 display the typical

modular structure of cellulases and hemicellulases (Table S7.2). Two notably complex enzymes are the pentamodular HA121 from the rumen bacterium *Ruminococcus albus* (GH43-CBM22-CBM22-DOC-CE1) and the tetramodular HA146 from the aerobic bacterium colonizing deep sea sediments *Zunongwangia profunda* (CE6-CE1-CBM48-CE1). Interestingly, all 27 FAEs presenting a modular architecture belong to clade B3, meaning that clades B1 and B2 have exclusively non-modular members. It has previously been shown that FAEs usually act cooperatively with xylanases (predominantly of Glycoside Hydrolase families 10 and 11, GH10 and GH11, respectively) during the hydrolysis of decorated xylans. Notably, modular FAEs usually display appended GH10 and GH11 xylanase catalytic domains (see Table S7.2 for examples), revealing the natural evolution of bi-functional enzymes to target different bonds in large polymeric carbohydrates.

7.3.6. Structure of *Lactobacillus crispatus* feruloyl esterase 1A (*Lc*FAE1A)

Data presented above suggest that the substrate specificities of bacterial FAEs do not reflect a phylogenetic proximity. A notable example supporting this observation is the highly identical group of enzymes that are closely clustered in sub-clade B11 comprising HA542, HA329, HA429, HA442 and HA246 (all enzymes share >70% primary sequence identity) (Figure 7.2). Although highly homologous the five enzymes display clear differences in terms of substrate specificity (Table 7.2); HA246 and 3QM1_Lj, *Lactobacillus johnsonii Lj*FAE1A for which the structure is known (PDB code 3QM1), represent the same enzyme. Thus, while *Lj*FAE1A (HA246) is only active on MFA and MSA, the other four enzymes cleave the four methyl ester substrates with similar efficiency. To clarify the structural determinants of substrate specify in this group of highly related enzymes, the structure of *Lactobacillus crispatus* HA246, termed *Lc*FAE1A was solved. Statistics for data collection and refinement are shown in Table 7.1. The enzyme crystallized as a trimer in the asymmetric unit. *Lc*FAE1A adopts the canonical α/β-hydrolase fold (Figure 7.3a) with a central eight-stranded mixed β-sheet topology. The central β-sheet is flanked by two α-helices on one side and three α-helices on the opposite side (Figure 7.3b). The most exclusive property of *Lc*FAE1A is the insertion of an α/β subdomain between strands β6 and β11, which is absent in other FAEs. This insertion contains two short β-hairpins and three α-helices that are positioned to almost cover the enzyme catalytic site providing the required structural environment to for a pocket for the recognition of the aromatic part of the substrate (Figure 7.3c).

A DALI search with *Lc*FAE1A coordinates revealed that, as expected, the closest structural homologue is *Lactobacillus johnsonii* Lj0536 (PDB code 3PFC), here termed *Lj*FAE1A, with a Z score of 42.2 and an r.m.d.s. of 0.7 Å over ~251 aligned residues. *Lc*FAE1A also displays significant structural homology with the bacterial homologue of the B11 bacterial clade, Est1E from *Butyrivibrio proteoclasticus* (PDB code 3WTN), with a Z

score of 38.2 and a r.m.d.s. of 1.3 over 250 residues. Remote structural homologies were observed with acetyl xylan esterases, lipases and deacetylases from other CE families. The overlay of *Lc*FAE1A and *Lj*FAE1A reveals that although the two enzymes present ~70% primary sequence identity, at a structural level the two enzymes are virtually identical (Figure 7.3d). The only significant difference that might have functional implications when the structure of the two enzymes is compared is the size of the α/β subdomain insertion that is greater in *Lc*FAE1A enzyme, thus providing a wider occlusion of the enzyme catalytic site. Thus, while the α/β subdomain creates a clear pocket for the accommodation of hydroxycinnamic acid in *Lc*FAE1A and *Lj*FAE1A, the opposite end of the substrate will rest in a solvent exposed area, suggesting that it can accommodate different types of moieties. This suggests that substrate discrimination is primarily provided by the hydrophobic and hydrophilic pocket contacts to the aromatic ring and its substituents.

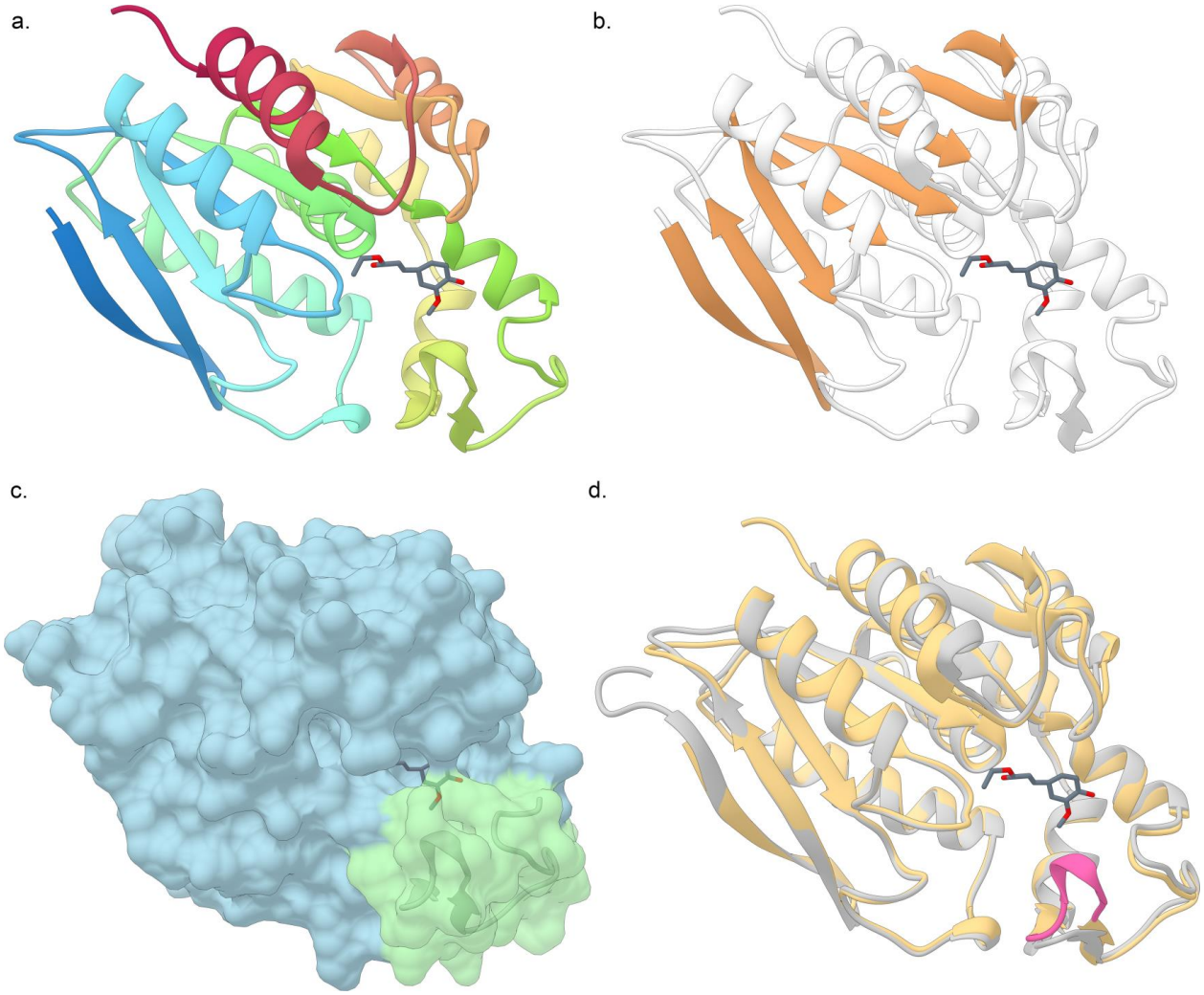


Figure 7.3. Structure of *Lactobacillus crispatus* feruloyl esterase 1A (*Lc*FAE1A).

Panel **a**, Ribbon representation of *Lc*FAE1A's structure with rainbow colouring (N-terminus in blue, C-terminus in red. Panel **b**, Ribbon representation of *Lc*FAE1A highlighting the typical (β/α)-hydrolase

fold, with the β-sheet coloured orange, between the α-helix sandwich. Panel **c**, Van der Walls surface of *Lc*FAE1A coloured in light blue, with the inserted α/β subdomain highlighted in green. Panel **d**, Overlay of *Lc*FAE1A's structure with PDB 3QM1, a *Lactobacillus johnsonii* cinnamoyl esterase in complex with ethyl ferulate. The main difference between the two structures is a small loop extension over *Lc*FAE1A's catalytic center, highlighted in magenta. All panels show a dark grey coloured molecule of ethyl ferulate in stick representation in the putative active center, resulting from the overlay with 3QM1.

7.3.7. Role of catalytic site residues for substrate recognition

Superposition of *Lc*FAE1A with *Lj*FAE1A in complex with ethyl ferulate (PDB 3QM1) and ferulic acid (PDB 3PFC) allowed visualizing the catalytic residues of *Lc*FAE1A and potential substrate specificity determinants (Figure 7.3d and Figure 7.4a). *Lc*FAE1A catalytic triad, composed Ser-125, His-248 and Asp-220, is in a completely conserved position with *Lj*FAE1A (Figure 7.4a) while *Lc*FAE1A oxyanion hole is formed by the nitrogen atoms of Phe-53 and Gln-126. Ferulate binds in the enzyme pocket formed by the α/β subdomain described above, which provides recognition elements for both the hydroxy and methoxy ring substituents. The hydroxyl group should form hydrogen bond contacts with *Lc*FAE1A Asp-157 and Tyr-163 (Figure 7.4b,c). In contrast, the methoxy ring substituent is accommodated by a highly hydrophobic cavity formed by the side chains of Phe-53 and Leu-188, which are completely conserved in the two *Lactobacillus* enzymes. Thus, with the exception of Gln-153 that is replaced by Thr-134 in *Lj*FAE1A the constellation of residues that binds the aromatic ring of the substrate and contributes for catalysis is completely conserved in the two enzymes (Figure 7.4).

In order to explore the structural differences in HA542 (*Lc*FAE1A), HA329, HA429, HA442 and HA246 (*Lj*FAE1A) enzymes that could explain the observed differences in substrate specificity, a structural alignment combining the 5 proteins was constructed (Figure 7.5). The data revealed that all enzymes which cleave the 4 methyl ester substrates, HA542 (*Lc*FAE1A), HA329, HA429 and HA442, contain a Gln in the position of Gln-153 *Lc*FAE1A (Figure 7.5). In contrast, *Lj*FAE1A (HA246) is the only enzyme that is unable to cleave MpCA and MCA and contains a Thr replacing Gln-153 of *Lc*FAE1A (Figure 7.5). This observation together with proximity of the Gln-Thr *Lc*FAE1A-*Lj*FAE1A side-chain to the aromatic ring of the substrate is a strong indication suggesting that this subtle amino acid change modulates differences in substrate specificity observed in the five enzymes. Thus, when the methoxy substituent is absent from the substrate, presence of the Gln side chain should provide extra binding interactions that are not supported by Thr-134 in *Lj*FAE1A. Although *Lj*FAE1A was soaked with chlorogenic acid, suggesting that putatively the enzyme could cleave substrates containing the caffeic acid aromatic ring, the enzyme activity was measure against a different

substrate (chlorogenic acid) suggesting that other residues could stabilize the enzyme-substrate complex (Lai *et al*. 2014). Presently, a mutagenic study comprising the dual change of Gln and Thr residues in HA542 (*Lc*FAE1A), HA329, HA429, HA442 and HA246 (*Lj*FAE1A) to their counterpart substituents is ongoing. This should provide insightful data to support the hypothesis that subtle catalytic site changes modulate the mechanism of substrate recognition in FAEs.

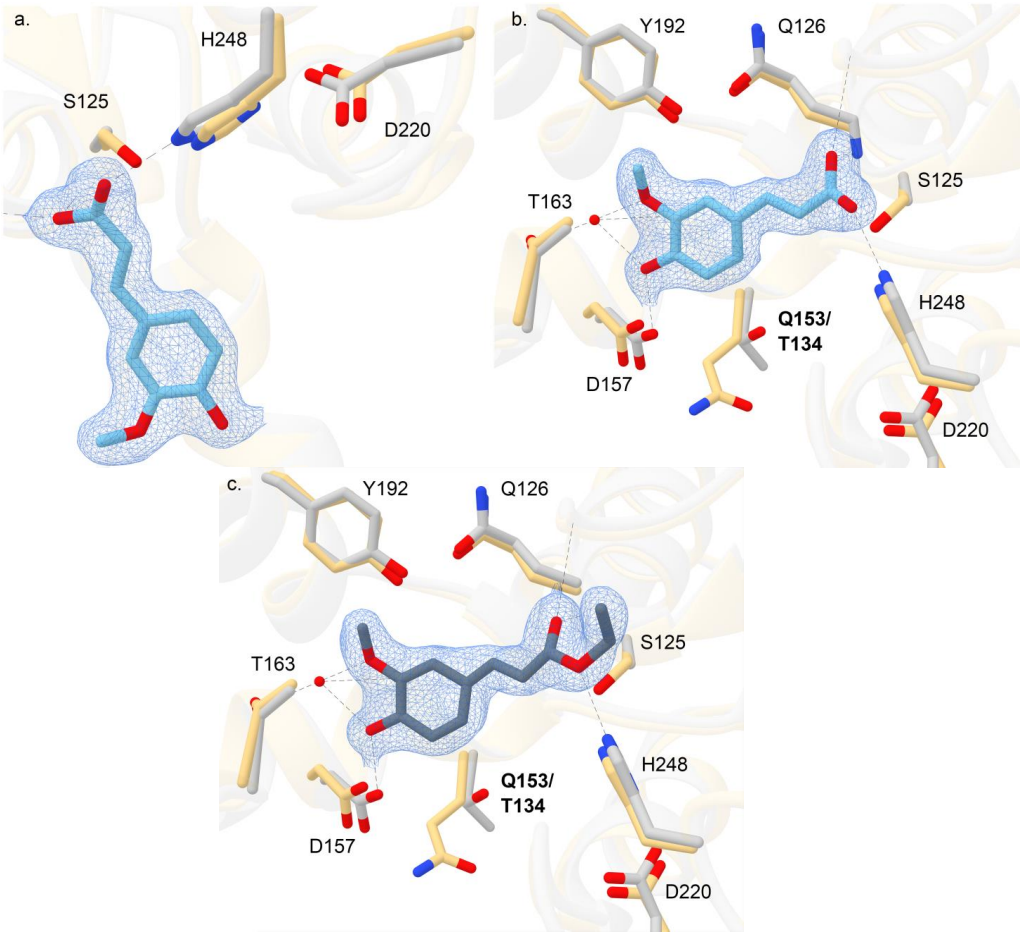


Figure 7.4. Detailed view of *Lc*FAE1A's putative active center, identified by overlaying its structure with *Lactobacillus johnsonii* cinnamoyl esterase in complex with ferulic acid. Panel **a,** shows the putative catalytic triad seen near a ferulic acid molecule that was placed by overlaying *Lc*FAE1A's structure with 3PFC. Panels **b** and **c** show a detailed view of *Lc*FAE1A's putative active center, identified by overlaying its structure with *Lactobacillus johnsonii* cinnamoyl esterase in complex with ferulic acid (3PFC) and ethyl ferulate (3QM1), respectively. Important residues for ligand recognition are shown in stick representation, with the most relevant hydrogen-bond contacts represented in black dashed lines. All ligands are surrounded by a mesh representation of the Refmac5 maximum-likelihood σA–weighted 2Fo−Fc electron density map contoured at 1σ (0.46 electrons/A3).

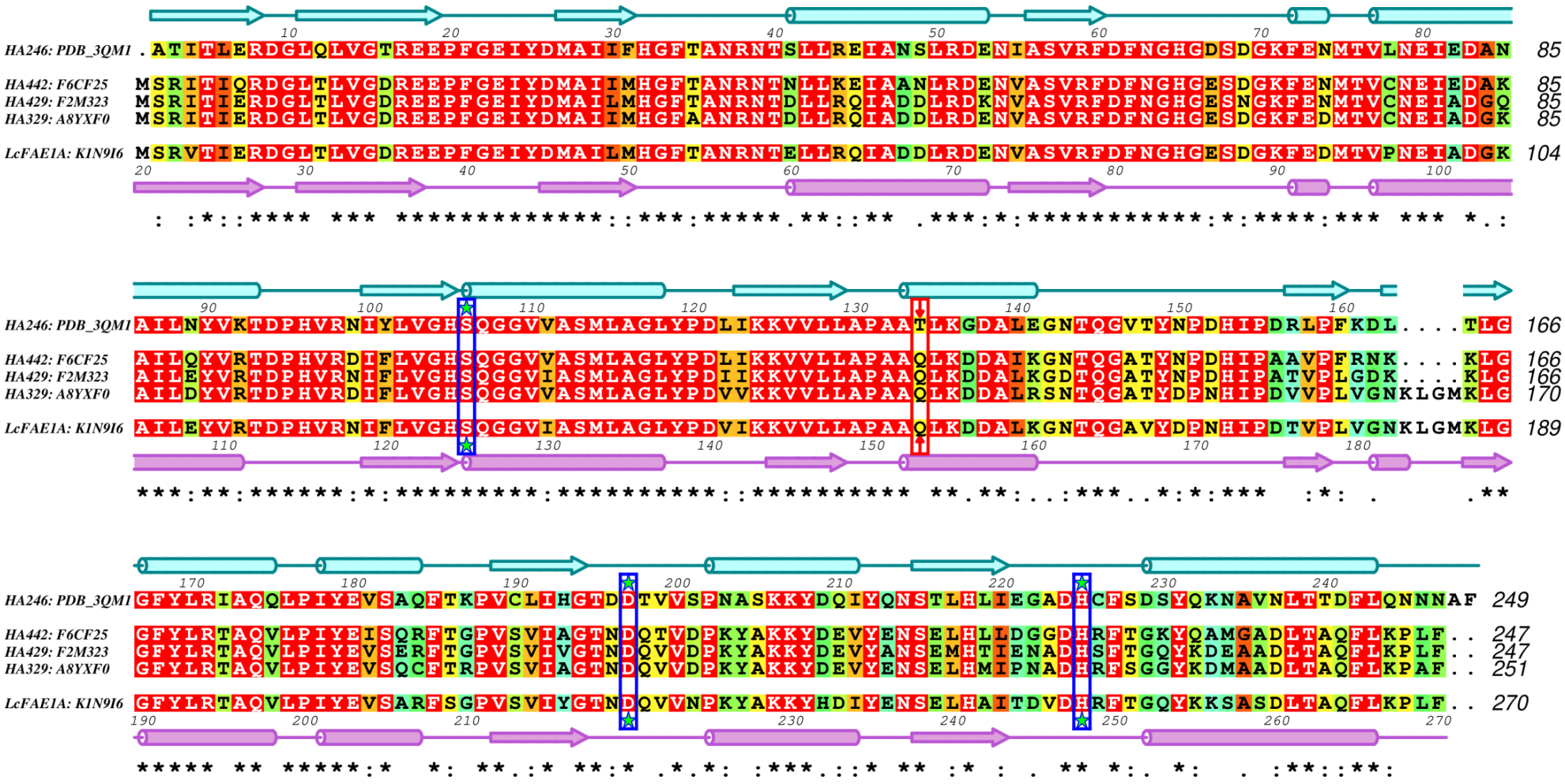


Figure 7.5. Multiple sequence alignment of *Lc*FAE1A clade members.

The secondary structure of HA246 (PDB ID 3QM1) and *Lc*FAE1A (PDB ID not yet available) is depicted above and below, respectively, of its sequence. Apart from HA246 which uses its PDB ID, all the other sequences are identified by their UniProt accession numbers. The sequences are numbered on the right-hand side of the alignment, except for HA246 (*Lj*FAE1A) and *Lc*FAE1A, which are numbered according to their PDB entries and also shown above and below their respective sequences. A blue box with stars identify the catalytic residue triad. A red box and arrows highlights the relevant residue for substrate modulation. The primary sequence background is coloured according to the ALSCRIPT Calcons convention, implemented in ALINE (Bond and Schüttelkopf 2009): red, identical residues; orange to blue, lowering colour-ramped scale of conservation. Below the alignment, the ClustalO consensus symbols represent the position conservation status.

7.4. Conclusions

Data presented here suggest that subtle changes at the catalytic site of feruloyl esterases may dramatically change the enzyme substrate specificity. Using two model *Lactobacillus* enzymes, *Lc*FAE1A and *Lj*FAE1A, we reveal that a Glutamine to Threonine change at the catalytic site restricts the capacity of the enzyme to cleave methyl p-coumaric and caffeic acids. The structures of the two enzymes also reveal the presence of a subdomain insertion, absent in the majority of known FAEs, that covers the enzyme catalytic site creating a pocket where hydroxycinnamic acids bind. Together, the data suggests that bacterial FAEs of clade B11 and possibly those of clade B1, to which the *Lactobacillus* enzymes belong, developed an exquisite specificity to hydroxycinnamic acids from plant polyphenols such as chlorogenic and rosmarinic acids. This is supported by the presence of an enclosed pocket for the specific recognition of the aromatic moiety of the substrate and the high levels of activity presented by B11 enzymes against the small substrates used in this study. In addition, it seems unlikely that B11 enzymes participate in the cooperative hydrolysis of hetero-xylans as inspection of the *Lactobacillus* proteome (www.cazy.org) revealed the presence of no xylanases of families GH10 and GH11. In conclusion, bacterial FAEs of clade B11 are particular interesting enzymes to release anti-oxidant, anti-inflammatory and anti-cancer hydroxycinnamic acids, in particular ferulic acid, from food and feed in biotechnological applications in human and animal nutrition.

8. GENERAL DISCUSSION AND FUTURE PERSPECTIVES

As referred in the previous chapters, the biodegradation of plant cell wall polysaccharides is a process of growing interest. In nature this process of outstanding biological and biotechnological significance is performed by microbial enzymes that are generally referred as Carbohydrate-Active Enzymes (CAZymes). It is now well established that monogastric animals cannot use efficiently some highly abundant ingredients (e.g. barley, oats, rice and wheat) and the dietary supplementation with exogenous CAZymes (such as cellulases and hemicellulases) leads in an improvement of the nutritive value of animals' diets which results in a better animal performance. The action mechanism of exogenous enzymes remains, however, to be understood in its entireness. For example, it remains to be established if feed enzymes which generate oligosaccharides with prebiotic activity, such as xylo-oligosaccharides, have an effect in animal performance. Therefore, this thesis describes the development of several technologies to discover and characterize novel CAZymes as well as their mechanism of action in animal nutrition.

The experiments described in Chapters 2 and 3 explored the utilization of CAZymes in poultry nutrition. In recent years, the view that CAZymes contribute to improve the nutritive value of cereal-based diets for monogastric animals was consolidated. Monogastric animals have a restricted repertoire of digestive enzymes that does not include the enzymes required to degrade PCW polysaccharides. Inclusion of microbial CAZymes in animal diets significantly reduce the levels of anti-nutritive factors of these cereals thus improving their intrinsic nutritive value (Charlton 1996). Initially, we started to investigate which enzymes, 1,3-1,4-β-glucanases or 1,4-β-glucanases, are the most important to improve the nutritive value of β-glucan-containing diets for monogastric animals. In addition, we investigated if the variation in levels of non-starch polysaccharides and endogenous endo-1,4-β-xylanases affect the nutritive value of wheat for poultry. Experiments reported in Chapter 2 were developed following the work reported by Fernandes *et al*. (2016) and, together, the data strongly suggest that cellulases/1,4-β-glucanases from CAZy families GH8 (Fernandes *et al.* 2016) and GH5 (this thesis) are not efficient to improve the nutritive value of barley-based diets for poultry. These observations indicated that eventually all cellulases are unable to effectively hydrolyse, *in vivo,* the anti-nutritive β-glucans due to their preferential substrate specificity for cellulosic carbohydrates. In contrast, enzymes such as the GH16 1,3-1,4-β-glucanase productively contributed to decrease the anti-nutritive effects of carbohydrates present in barley-based diets. Thus, effective microbial mixtures used to supplement barley-based diets for monogastric animals should contain primarily 1,3-1,4-β-glucanases but not cellulases/1,4-β-glucanases. These results established the importance of a detailed analyse

of all variable factors present in animal diets that should be considered when evaluating the efficacy of feed enzymes. In the second experiment, the variation in the endogenous levels of non-starch polysaccharides and endo-1,4-β-xylanases in different wheat lots was measured. Results revealed that wheat extract viscosity and endo-1,4-β-xylanase, although not correlated, vary widely in different wheat lots. Exogenous enzyme supplementation improved the nutritive value of the diet incorporating the wheat lot displaying higher extract viscosity and lower endogenous endo-1,4-β-xylanase activity. In opposition, when the wheat lot presented lower viscosity and higher levels of endogenous xylanase activity, broiler response to feed supplementation with exogenous microbial xylanases was clearly diminished. These data clearly indicated that, as observed previously for endogenous β-glucanase activity in barley (Ribeiro *et al.* 2011), endogenous xylanase activity may influence response to exogenous enzyme supplementation. Thus, wheat lots presenting higher levels of endogenous xylanase activity may be less responsive to the supplementation with exogenous enzymes. However, it remains to be investigated which factor, endogenous endo-1,4-β-xylanase activity or intrinsic extract viscosity, influences mostly the effective response to exogenous enzyme supplementation of broilers fed on wheat-based diets.

In another perspective, in Chapter 4 we evaluated the capacity of CAZymes to generate prebiotic xylo-oligosaccharides (XOS), resulting from the degradation of arabinoxylans by xylanases, and tested the capacity of xylo-oligosaccharides to stimulate animal performance. The data revealed that exogenous xylanases and XOS used to supplement wheat-based diets for poultry lead to a similar broiler performance. Thus, the results suggested that in wheat-based diets exogenous xylanases may influence animal performance through the production of XOS rather than by reducing the concentration of the soluble, viscous, arabinoxylans or by reducing the integrity of cereal cell walls. In addition the data indicate that XOS display a prebiotic property in broiler chicks, resulting in a significant beneficial effects in bird performance (De Maesschalck *et al*. 2015). Furthermore, the bio-activity of unbranched XOS seems to be similar to the oligosaccharides resulting from the hydrolysis of feed arabinoxylans that are mostly branched with arabinose side-chains (AXOS). Finally, data suggested that the incorporation rates of XOS need to be carefully fine-tuned, since XOS used at a 10g/kg level had no effect on performance. Similar such “overdosing” effects have been demonstrated by Zhenping *et al.* (2013). Thus, optimum incorporation rates may range between 0.1 to 1 g kg of XOS, although animals fed on corn-based diets supplemented with XOS at a 0.1 g/kg rate resulted in more than 350 grams of additional body weight gain when compared with the control group, suggesting that even lower doses may be effective. Consequently, taken together the data suggest that XOS modulate an improvement in animal performance by optimizing feed digestion, feed intake and by triggering the evolution of the microbiome to a more favourable construction.

The animal trials developed during the first part of this project indicated the role and importance of exogenous CAZymes in scientific perspective of animal nutrition. In addition, they revealed the importance of discovering and characterizing in detail CAZymes to be used in particular applications in animal nutrition. The second part of the project was dedicated to identify novel enzymes for animal nutrition applications, in particular those related with the release of hemicellulose and cellulose from lignin. These experiments are discussed in the paragraphs below.

The deconstruction of PCW polysaccharides is mediated by an array of CAZymes, that include GHs, PLs, CEs, GTs and AAs and their classification in families is continuously updated in the CAZy database (www.cazy.org) that provides an efficient mechanism to understand and systematically organize the current knowledge on the different enzymes/families (Cantarel *et al*. 2009; Lombard *et al*. 2014). Thus, the CAZy database is an invaluable research tool with respect to the ever-increasing amount of genomic information relating to the carbohydrate metabolism. Currently, the database contains 165 sequence-based families of GHs, 37 families of PLs, 15 families of CEs, 107 families of GTs, 16 families of AAs and finally, 85 families of CBMs (data collected on July 2019). The functional diversity of CAZymes is enormous and reflects the wide multiplicity of glycan structures found in nature. Nevertheless, the bottleneck in the current biotechnological application of this growing knowledge is the availability of commercially characterized CAZymes that cover a wide range of enzyme specificities. In Chapter 5, it was aimed to build a large enzyme library that can be applied to study PCW degradation and to exploit potential biotechnological applications. The work started with a major bioinformatic analysis to identify and select putative enzymes with significant interest for the deconstruction of PCW carbohydrates. This was achieved through the implementation of HTP methods for production, cloning, expression and purification of recombinant proteins. Thus, a HTP platform was used to clone 1476 genes, from 486 different microorganisms and express them in *E. coli*. The data revealed that 79% of recombinant proteins were produced in the soluble form. In other words, this approach resulted in the production of 1166 characterized enzymes leading to a considerable increase in the number of available enzymes for PCW degradation studies and possibly for their application in different industrial and bioprocessing processes. The strategy used, in addition to the soluble targets obtained, allowed to define different factors that could affect the solubility of proteins and allowed establishing parameters to predict the solubility of the proteins. Thus, the data revealed a significantly reduced number ($P<0,01$) of soluble proteins belonging to both eukaryotic and archaeal domains with a percentage of 57,7% and 53,3%, respectively, compared with 79,7% of soluble proteins with a bacterial origin.

Relatively to the gene isolation strategy, when we compared genes produced by PCR or GS, we observed a significantly reduced number (P=0,04) of soluble proteins produced by GS. Fusion GFP tag also influenced positively protein's solubility (P=0,05) (Zhang *et al*. 1998). In contrast, no relations were observed between protein molecular mass and solubility. Finally, an association between amino acid composition and protein solubility was observed, being the percentage of non-polar and negatively charged amino acids a key predictor of protein's solubility (Christendat *et al*. 2000; Bertone *et al*. 2001; Niu *et al*. 2013). Overall, the HTP approach developed in this thesis was a powerful tool for the production of recombinant CAZymes that can be used for future studies of PCW degradation.

In the last experimental chapters of this thesis, Chapters 6 and 7, 500 putative feruloyl esterases (FAEs) and glucuronoyl esterases (GEs), belonging to CE families CE1 and CE15, respectively, were selected for recombinant production and biochemical characterization. FAEs and GEs participate in the hydrolysis of the ester linkages between hemicellulosic carbohydrates and lignin (Ishii 1997; Saulnier and Thibault 1999; Balakshin *et al*. 2011; Bååth *et al*. 2016). By cleaving the carbohydrate-lignin covalent linkages in fibre, these enzymes are anticipated to dramatically contribute to improve the accessibility of cellulose and hemicellulose to carbohydrate-degrading enzymes, such as cellulases and xylanases, respectively, thus improving fibre digestion (Monrad *et al*. 2018). However, the current knowledge of FAEs and GEs, in particular in what concerns their potential use to improve the nutritive value of lignin-rich diets for herbivores, is still very limited. Therefore, the main aim of these chapters was the use of HTP approaches described in detail in Chapter 5, to create a large library of FAEs and GEs that could be screened to develop a novel class of exogenous feed enzymes for herbivores. After gene isolation, 372 FAEs and 11 GEs were produced at a concentration greater than 10 mg/L in *E. coli*. Notably, 96 enzymes were recombinantly produced at yields above 300 mg/L. The highest production level was 700 mg/L for a bacterial FAE. Using the optimized conditions for activity 199 FAEs and 11 GEs were shown to efficiently cleave lignin-carbohydrate substrates with the highest enzyme activity higher than 94 and 105 U/mg, respectively. Overall the results reveal the ability of the genome mining strategy to identify novel FAE and GE encoding genes, with over 50% of the enzymes produced retaining significant levels of activity and stability. To our knowledge this library is, worldwide, the largest currently available to identify novel feed enzymes to improve the nutritive-value of lignin-rich diets for herbivores. In addition, the work allowed developing structure-function studies in a selected number of GEs and FAEs. The clarification of the structural modulators of enzyme substrate specificity is key to develop rational methods to explore the biotechnological potential of CAZymes, in particular FAEs and GEs, in animal nutrition. Thus, the work developed in this thesis opens new avenues to develop a novel class of enzymes to animal nutrition.

In conclusion, this thesis describes the importance of exogenous CAZymes in monogastric animal nutrition and the capacity of novel HTP strategies to identify novel CAZymes, in a pool of dozens to thousands putative proteins, to supplements diets posing significant limitations to digestibility in the animal´s gastrointestinal tract. This strategy was optimized and demonstrated great levels of success. Future work should address the possibility to test some of the most promising enzymes, produced and characterized here, in animal nutrition, in particular to analyse de effect of GEs and FAEs to degrade ligno-cellulosic biomass present in animal feed. Initial studies should explore the use of FAEs and GEs to improve the nutritive-value of lignin-rich diets for ruminants/herbivores. Later these studies could be extended for monogastric animals by analysing the capacity of FAEs to release interesting bioactive molecules, such ferulic acid that present anti-oxidant, anti-cancer and anti-inflammatory properties, during the digestive process.

9. BIBLIOGRAPHIC REFERENCES

A.O.A.C. 1980. Official Methods of Analysis. In: Association of Official Analytical Chemists. 13th ed. Washington, DC. p. 376–384.

Aachary AA, Prapulla SG. 2011. Xylooligosaccharides (XOS) as an Emerging Prebiotic: Microbial Synthesis, Utilization, Structural Characterization, Bioactive Properties, and Applications. Compr Rev Food Sci Food Saf. 10(1):2–16.

Abdel-Hamid A, Solbiati J, Cann I. 2013. Insights into lignin degradation and its potential industrial applications. Adv Appl Microbiol. 82:1–28.

Alber O, Noach I, Rincon MT, Flint HJ, Shimon LJW, Lamed R, Frolow F, Bayer EA. 2009. Cohesin diversity revealed by the crystal structure of the anchoring cohesin from Ruminococcus flavefaciens. Proteins Struct Funct Bioinforma. 77(3):699–709.

Albersheim P, Bar-Peled M, Bergmann C, Darvill A, Eberhard S, Freshour G, Gelino-Albersheim I, Hahn M, Mohnen D, O'Neill MA, *et al*. 2007. Plant Cell Walls. Complex Carbohydr Res Cent - Univ Georg. [accessed 2019 Mar 20]. https://www.ccrc.uga.edu/~mao/intro/ouline.htm.

Angkanaporn K, Choct M, Bryden WL, Annison EF, Annison G. 1994. Effects of wheat pentosans on endogenous amino acid losses in chickens. J Sci Food Agric. 66(3):399–404.

Antoniou T, Marquardt RR, Cansfield PE. 1981. Isolation, Partial Characterization, and Antinutritional Activity of a Factor (Pentosans) in Rye Grain. J Agric Food Chem. 29(6):1240–1247.

Apajalahti J, Bedford MR. 1999. Improve bird performance by feeding its microflora. Worlds Poult Sci J. 15:20–23.

Artzi L, Bayer EA, Moraïs S. 2016. Cellulosomes: Bacterial nanomachines for dismantling plant polysaccharides. Nat Rev Microbiol. 15(2):83–95.

Artzi L, Dassa B, Borovok I, Shamshoum M, Lamed R, Bayer EA. 2014. Cellulosomics of the cellulolytic thermophile Clostridium clariflavum. Biotechnol Biofuels. 7(1):1–13.

Artzi L, Morag E, Barak Y, Lamed R, Bayer EA. 2015. Clostridium clariflavum: Key cellulosome players are revealed by proteomic analysis. MBio. 6(3):1–12.

Aurilia V, Martin JC, McCrae SI, Scott KP, Rincon MT, Flint HJ. 2000. Three multidomain esterases from the cellulolytic rumen anaerobe Ruminococcus flavefaciens 17 that carry divergent dockerin sequences. Microbiology. 146(6):1391–1397.

Bååth JA, Giummarella N, Klaubauf S, Lawoko M, Olsson L. 2016. A glucuronoyl esterase from Acremonium alcalophilum cleaves native lignin-carbohydrate ester bonds. FEBS Lett. 590(16):2611–2618.

Bååth JA, Mazurkewich S, Knudsen RM, Poulsen JN, Olsson L, Leggio L Lo,

Larsbrink J. 2018. Biochemical and structural features of diverse bacterial glucuronoyl esterases facilitating recalcitrant biomass conversion. Biotechnol Biofuels. 11(1):1–14.

Bååth JA, Mazurkewich S, Poulsen JCN, Olsson L, Lo Leggio L, Larsbrink J. 2019. Structure-function analyses reveal that a glucuronoyl esterase from Teredinibacter turnerae interacts with carbohydrates and aromatic compounds. J Biol Chem. 294(16):6635–6644.

Balakshin M, Capanema E, Gracz H, Chang H min, Jameel H. 2011. Quantification of lignin-carbohydrate linkages with high-resolution NMR spectroscopy. Planta. 233(6):1097–1110.

Baurhoo B, Ferket PR, Zhao X. 2009. Effects of diets containing different concentrations of mannanoligosaccharide or antibiotics on growth performance, intestinal development, cecal and litter microbial populations, and carcass parameters of broilers. Poult Sci. 88(11):2262–2272.

Baurhoo B, Letellier A, Zhao X, Ruiz-Feria CA. 2007. Cecal populations of lactobacilli and bifidobacteria and Escherichia coli populations after in vivo Escherichia coli challenge in birds fed diets with purified lignin or mannanoligosaccharides. Poult Sci. 86(12):2509–2516.

Baurhoo B, Phillip L, Ruiz-Feria CA. 2007. Effects of purified lignin and mannan oligosaccharides on intestinal integrity and microbial populations in the ceca and litter of broiler chickens. Poult Sci. 86(6):1070–1078.

Bayer EA. 2019. The Cellulosome Complex. [accessed 2019 Feb 3]. http://www.weizmann.ac.il/Biomolecular_Sciences/Bayer/.

Bayer EA, Belaich J-P, Shoham Y, Lamed R. 2004. The Cellulosomes: Multienzyme Machines for Degradation of Plant Cell Wall Polysaccharides. Annu Rev Microbiol. 58(1):521–554.

Bayer EA, Chanzyt H, Lamed R, Shoham Y. 1998. Cellulose, cellulases and cellulosomes. Curr Opin Struct Biol. 8(5):548–557.

Bayer EA, Kenig R, Lamed R. 1983. Adherence of Clostridium thermocellum to cellulose. J Bacteriol. 156(2):818–827.

Bayer EA, Lamed R, White BA, Flints HJ. 2008. From cellulosomes to cellulosomics. Chem Rec. 8(6):364–377.

Bayer EA, Morag E, Lamed R. 1994. The cellulosome - A treasure-trove for biotechnology. Trends Biotechnol. 12(9):379–386.

Bedford MR. 1996. The effect of enzymes on digestion. J Appl Poult Res. 5(4):370–378.

Bedford MR. 1997. Factors affecting response of wheat based diets to enzyme supplementation. In: Recent advances in animal nutrition in australia. Australia: University of New England. p. 1–7.

Bedford MR. 2000. Exogenous enzymes in monogastric nutrition - Their current value and future benefits. Anim Feed Sci Technol. 86(1–2):1–13.

Bedford MR. 2006. Effect of non-starch polysaccharides on avian gastrointestinal function. In: Perry GC, editor. Avian gut function in health and disease. Wallingford, UK: CAB International. p. 159–170.

Bedford MR, Classen HL. 1992. Reduction of intestinal viscosity trough manipulation of dietary rye and pentosanase concentration is effected through changes in the carbohydrate compostion of the intestinal aqueous phase and results in improved growth rate and food conversion efficienc. J Nutr. 122(3):560–569.

Bedford MR, Classen HL, Campbell GL. 1991. The effect of pelleting, salt, and pentosanase on the viscosity of intestinal contents and the performance of broilers fed rye. Poult Sci. 70(7):1571–1577.

Bedford MR, Morgan AJ. 1996. The use of enzymes in poultry diets. Worlds Poult Sci J. 52(1):61–68.

Bedford MR, Schulze H. 1998. Exogenous enzymes for pigs and poultry. Nutr Res Rev. 11(1):91–114.

Bélaich JP, Tardif C, Bélaich A, Gaudin C. 1997. The cellulolytic system of clostridium cellulolyticum. J Biotechnol. 57(1–3):3–14.

Belancic A, Scarpa J, Peirano A, Diaz R, Steiner J, Eyzaguirre J. 1995. Penicillium purpurogenum produces several xylanases: Purification and properties of two of the enzymes. J Biotechnol. 41(1):71–79.

Benoit I, Danchin EGJ, Bleichrodt RJ, De Vries RP. 2008. Biotechnological applications and potential of fungal feruloyl esterases based on prevalence, classification and biochemical diversity. Biotechnol Lett. 30(3):387–396.

Bertone P, Kluger Y, Lan N, Zheng D, Christendat D, Yee A, Edwards AM, Arrowsmith CH, Montelione GT, Gerstein M. 2001. SPINE: an integrated tracking database and data mining approach for identifying feasible targets in high-throughput structural proteomics. Nucleic Acids Res. 29(12):2884–2898.

Bhat KM, Wood TM. 1992. The cellulase of the anaerobic bacterium Clostridium thermocellum: Isolation, dissociation, and reassociation of the cellulosome. Carbohydr Res. 227:293–300.

Bianchi AA, Baneyx F. 1999. Hyperosmotic shock induces the sigma32 and sigmaE stress regulons of Escherichia coli. Mol Microbiol. 34(5):1029–38.

Biely P. 2012. Microbial carbohydrate esterases deacetylating plant polysaccharides. Biotechnol Adv. 30(6):1575–1588.

Biely P, Vršanská M, Tenkanen M, Kluepfel D. 1997. Endo-β-1,4-xylanase families: Differences in catalytic properties. J Biotechnol. 57(1–3):151–166.

Bogusławska-Tryk M, Piotrowska A, Burlikowska K. 2012. Pokarmowe fruktany i ich potencjalny korzystny wpływ na zdrowie i wyniki produkcyjne u kurcza{ogonek}t brojlerów. J Cent Eur Agric. 13(2):272–291.

Bond CS, Schüttelkopf AW. 2009. ALINE: A WYSIWYG protein-sequence alignment

editor for publication-quality alignments. Acta Crystallogr Sect D Biol Crystallogr. 65(5):510–512.

Bonnin E, Garnier C, Ralet MC. 2014. Pectin-modifying enzymes and pectin-derived materials: Applications and impacts. Appl Microbiol Biotechnol. 98(2):519–532.

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J. 382(Pt 3):769–781.

Bourbonnais R, Paice M. 1990. Oxidation of non-phenolic substrates. An expanded role for laccase in lignin biodegradation. FEBS Lett. 5267(1):99–102.

Breen A, Singleton FL. 1999. Fungi in lignocellulose breakdown and biopulping. Curr Opin Biotechnol. 10(3):252–258.

Brenes A, Smith M, Guenter W, Marquardt RR. 1993. Effect of enzyme supplementation on the performance and digestive tract sue of broiler chickens fed wheat- and barley-based diets. Poult Sci. 72(9):1731–1739.

Breslmayr E, Hanžek M, Hanrahan A, Leitner C, Kittl R, Šantek B, Oostenbrink C, Ludwig R. 2018. A fast and sensitive activity assay for lytic polysaccharide monooxygenase. Biotechnol Biofuels. 11(1):1–13.

Broekaert WF, Courtin CM, Verbeke K, van de Wiele T, Verstraete W, Delcour JA. 2011. Prebiotic and other health-related effects of cereal-derived arabinoxylans, arabinoxylan-oligosaccharides, and xylooligosaccharides. Crit Rev Food Sci Nutr. 51(2):178–194.

Brown Jr RM. 1996. The Biosynthesis of Cellulose. J Macromol Sci Part A - Chem Pure Appl Chem. 33(10):1345–1373.

Brown ME, Chang MCY. 2014. Exploring bacterial lignin degradation. Curr Opin Chem Biol. 19(1):1–7.

Buanafina MM de O. 2009. Feruloylation in grasses: Current and future perspectives. Mol Plant. 2(5):861–872.

Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. 2011. Pathways for degradation of lignin in bacteria and fungi. Nat Prod Rep. 28(12):1883–1896.

Bule P, Pires VMR, Alves VD, Carvalho AL, Prates JAM, Ferreira LMA, Smith SP, Gilbert HJ, Noach I, Bayer EA, *et al*. 2018. Higher order scaffoldin assembly in Ruminococcus flavefaciens cellulosome is coordinated by a discrete cohesin-dockerin interaction. Sci Rep. 8(1):1–14.

Bule P, Ruimy-Israeli V, Cardoso V, Bayer EA, Fontes CMGA, Najmudin S. 2014. Overexpression, crystallization and preliminary X-ray characterization of Ruminococcus flavefaciens scaffoldin C cohesin in complex with a dockerin from an uncharacterized CBM-containing protein. Acta Crystallogr Sect F Struct Biol Commun. F70(8):1061–1064.

Bulone V, Fincher GB, Stone BA. 1995. In vitro synthesis of a microfibrillar (1→3)-B-glucan by a ryegrass (Lolium multiflorum) endosperm (1→3)-B-glucan synthase enriched by product entrapment. Plant J. 8(2):213–225.

Butler MJ, Day AW. 1998. Destruction of Fungal Melanins by Ligninases of Phanerochaete chrysosporium and Other White Rot Fungi. Int J Plant Sci. 159(6):989–995.

Caffall KH, Mohnen D. 2009. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydr Res. 344(14):1879–1900.

Camilo CM, Polikarpov I. 2014. High-throughput cloning, expression and purification of glycoside hydrolases using Ligation-Independent Cloning (LIC). Protein Expr Purif. 99:35–42.

Campbell JA, Davies GJ, Bulone V, Henrissat B. 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J. 326(Pt 3):929–939.

Campbell JM, Fahey GC, Wolf BW. 1997. Nutrient Metabolism Selected Indigestible Oligosaccharides Affect Large Bowel Mass, Cecal and Fecal Short-Chain Fatty Acids, pH and Microflora in Rats. J Nutr. 127(1):130–136.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 37:233–238.

Cardoso V, Ferreira AP, Costa M, Ponte PIP, Falcão L, Freire JP, Lordelo MM, Ferreira LMA, Fontes CMGA, Ribeiro T. 2014. Temporal restriction of enzyme supplementation in barley-based diets has no effect in broiler performance. Anim Feed Sci Technol. 198:186–195.

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

CCP4. 1994. The CCP4 suite: Programs for protein crystallography. Acta Crystallogr Sect D Biol Crystallogr. 50(5):760–763.

Chakraborty S, Rani A, Dhillon A, Goyal A. 2017. Polysaccharide Lyases. In: Current Developments in Biotechnology and Bioengineering: Production, Isolation and Purification of Industrial Products. India: Elsevier B.V. p. 527–539.

Charavgi MD, Dimarogona M, Topakas E, Christakopoulos P, Chrysina ED. 2013. The structure of a novel glucuronoyl esterase from Myceliophthora thermophila gives new insights into its role as a potential biocatalyst. Acta Crystallogr Sect D Biol Crystallogr. 69(1):63–73.

Charlton P. 1996. Expanding enzyme application: higher aminoacid and energy values for vegetable proteins. In: Biotechnology in the feed industry, 12. Nottingham: Proceedings...Nottingham: Nottingham University Press. p. 317–326.

Chaubey M, Kapoor VP. 2001. Structure of a galactomannan from the seeds of Cassia angustifolia Vahl. Carbohydr Res. 332(4):439–444.

Chee SH, Iji PA, Choct M, Mikkelsen LL, Kocher A. 2010. Characterisation and response of intestinal microflora and mucins to manno-oligosaccharide and antibiotic supplementation in broiler chickens. Br Poult Sci. 51(3):368–380.

Chesson A. 1993. Feed enzymes. Anim Feed Sci Technol. 45(1):65–79.

Choct M. 1997. Feed non-starch polysaccharides: Chemical structures and nutritional significance. Feed Milling Int.(June Issue):13–26.

Choct M. 2006. Enzymes for the feed industry: past, present and future. Worlds Poult Sci J. 62(01):5–16.

Choct M, Annison G. 1990. Anti-nutritive activity of wheat pentosans in broiler diets. Br Poult Sci. 31(4):811–821.

Choct M, Annison G. 1992. Anti-nutritive effect of wheat pentosans in broiler chickens: Roles of viscosity and gut microflora. Br Poult Sci. 33(4):821–834.

Choct M, Dersjant-Li Y, McLeish J, Peisker M. 2010. Soy oligosaccharides and soluble non-starch polysaccharides: A review of digestion, nutritive and anti-nutritive effects in pigs and poultry. Asian-Australasian J Anim Sci. 23(10):1386–1398.

Choct M, Hughes RJ, Wang J, Bedford MR, Morgan AJ, Annison G. 1996. Increased small intestinal fermentation is partly responsible for the anti-nutritive activity of non-starch polysaccharides in chickens. Br Poult Sci. 37(3):609–621.

Christendat D, Yee A, Dharamsi A, Kluger Y, Savchenko A, Cort JR, Booth V, Mackereth CD, Saridakis V, Ekiel I, *et al*. 2000. Structural proteomics of an archaeon. Nat Struct Biol. 7(10):903–909.

Christov L, Prior B. 1993. Esterases of xylan-degrading microorganisms: production, properties, and significance. Enzyme Microb Technol. 15(6):460–475.

Chung YC, Hsu CK, Ko CY, Chan YC. 2007. Dietary intake of xylooligosaccharides improves the intestinal microbiota, fecal moisture, and pH value in the elderly. Nutr Res. 27(12):756–761.

Ciruela A, Gilbert HJ, Ali BRS, Hazlewood GP. 1998. Synergistic interaction of the cellulosome integrating protein (CipA) from Clostridium thermocellum with a cellulosomal endoglucanase. FEBS Lett. 422(2):221–224.

Clabbers MTB, Gruene T, Parkhurst JM, Abrahams JP, Waterman DG. 2018. Electron diffraction data processing with DIALS. Acta Crystallogr Sect D Struct Biol. 74(6):506–518.

Claus H. 2004. Laccases: Structure, reactions, distribution. Micron. 35(1–2):93–96.

Clifford CB. 2018. 6.3.3 Hemicellulases and Lignin-degrading Enzymes. Penn State's Coll Earth Miner Sci. [accessed 2019 Feb 5]. https://www.e-education.psu.edu/egee439/node/671.

Coelho E, Rocha MAM, Moreira ASP, Domingues MRM, Coimbra MA. 2016. Revisiting the structural features of arabinoxylans from brewers' spent grain. Carbohydr Polym. 139:167–176.

Collins MD, Gibson GR. 1999. Probiotics, prebiotics, and synbiotics: Approaches for modulating the microbial ecology of the gut. Am J Clin Nutr. 69(5):1052S–7S.

Colpa DI, Fraaije MW, Van Bloois E. 2014. DyP-type peroxidases: A promising and versatile class of enzymes. J Ind Microbiol Biotechnol. 41(1):1–7.

Colquhoun IJ, Ralet MC, Thibault JF, Faulds CB, Williamson G. 1994. Structure identification of feruloylated oligosaccharides from sugar-beet pulp by NMR spectroscopy. Carbohydr Res. 263(2):243–256.

Cornet P, Millet J, Béguin P, Aubert J-P. 1983. Characterization of Two Cel (Cellulose Degradation) Genes of Clostridium Thermocellum Coding for Endoglucanases. Nat Biotechnol. 1:589–594.

Correia MAS, Prates JAM, Brás J, Fontes CMGA, Newman JA, Lewis RJ, Gilbert HJ, Flint JE. 2008. Crystal Structure of a Cellulosomal Family 3 Carbohydrate Esterase from Clostridium thermocellum Provides Insights into the Mechanism of Substrate Recognition. J Mol Biol. 379(1):64–72.

Cosgrove DJ. 2001. Wall Structure and Wall Loosening. A Look Backwards and Forwards. Plant Physiol. 125(1):131–134.

Cosgrove DJ. 2005. Growth of the plant cell wall. Nat Rev Mol Cell Biol. 6(11):850–861.

Courtin CM, Broekaert WF, Swennen K, Lescroart O, Onagbesan O, Buyse J, Decuypere E, Van De Wiele T, Marzorati M, Verstraete W, *et al*. 2008. Dietary inclusion of wheat bran arabinoxylooligosaccharides induces beneficial nutritional effects in chickens. Cereal Chem. 85(5):607–613.

Courtin CM, Gys W, Gebruers K, Delcour JA. 2005. Evidence for the involvement of arabinoxylan and xylanases in refrigerated dough syruping. J Agric Food Chem. 53(19):7623–7629.

Courtin CM, Swennen K, Verjans P, Delcour JA. 2009. Heat and pH stability of prebiotic arabinoxylooligosaccharides, xylooligosaccharides and fructooligosaccharides. Food Chem. 112(4):831–837.

Coutinho PM, Deleury E, Davies GJ, Henrissat B. 2003. An evolving hierarchical family classification for glycosyltransferases. J Mol Biol. 328(2):307–317.

Couturier M, Ladevèze S, Sulzenbacher G, Ciano L, Fanuel M, Moreau C, Villares A, Cathala B, Chaspoul F, Frandsen KE, *et al*. 2018. Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. Nat Chem Biol. 14(3):306–310.

Van Craeyveld V, Swennen K, Dornez E, Van de Wiele T, Marzorati M, Verstraete W, Delaedt Y, Onagbesan O, Decuypere E, Buyse J, *et al*. 2008. Structurally Different Wheat-Derived Arabinoxylooligosaccharides Have Different Prebiotic and Fermentation Properties in Rats. J Nutr. 138(12):2348–2355.

Crepin VF, Faulds CB, Connerton IF. 2004. Functional classification of the microbial feruloyl esterases. Appl Microbiol Biotechnol. 63(6):647–652.

Crittenden R, Karppinen S, Ojanen S, Tenkanen M, Fagerstrom R, Matto J, Saarela M, Mattila-Sandholm T, Poutanen K. 2002. In vitro fermentation of cereal dietary fibre carbohydrates by probiotic and intestinal bacteria. J Sci Food Agric. 82(8):781–789.

Czjzek M, Bolam DN, Mosbah A, Allouch J, Fontes CMGA, Ferreira LMA, Bornet O, Zamboni V, Darbon H, Smith NL, *et al*. 2001. The Location of the Ligand-binding Site of Carbohydrate-binding Modules That Have Evolved from a Common Sequence Is Not Conserved. J Biol Chem. 276(51):48580–48587.

D'Errico C, Börjesson J, Ding H, Krogh KBRM, Spodsberg N, Madsen R, Monrad RN. 2016. Improved biomass degradation using fungal glucuronoyl-esterases-hydrolysis of natural corn fiber substrate. J Biotechnol. 219:117–123.

D'Errico C, Jørgensen JO, Krogh KBRM, Spodsberg N, Madsen R, Monrad RN. 2015. Enzymatic degradation of lignin-carbohydrate complexes (LCCs): Model studies using a fungal glucuronoyl esterase from Cerrena unicolor. Biotechnol Bioeng. 112(5):914–922.

D'Souza TM, Merritt CS, Reddy CA. 1999. Lignin-Modifying Enzymes of the White Rot Basidiomycete Ganoderma lucidum. 65(12):5307–5313.

Danicke S, Jeroch H, Bottcher W, Simon O. 2000. Interactions between dietary fat type and enzyme supplementation in broiler diets with high pentosan contents: effects on precaecal and total tract digestibility of fatty acids, metabolizability of gross energy, digesta viscosity and weights of small inte. Anim Feed Sci Technol. 84(3–4):279–294.

Dassa B, Borovok I, Lamed R, Henrissat B, Coutinho P, Hemme CL, Huang Y, Zhou J, Bayer EA. 2012. Genome-wide analysis of acetivibrio cellulolyticus provides a blueprint of an elaborate cellulosome system. BMC Genomics. 13(1):1.

Davies G, Henrissat B. 1995. Structures and mechanisms of glycosyl hydrolases. Structure. 3(9):853–859.

Davis GD, Elisee C, Mewham DM, Harrison RG. 1999. New fusion protein systems designed to give soluble expression in Escherichia coli. Biotechnol Bioeng. 65(4):382–388.

DeBoy RT, Mongodin EF, Fouts DE, Tailford LE, Khouri H, Emerson JB, Mohamoud Y, Watkins K, Henrissat B, Gilbert HJ, *et al*. 2008. Insights into plant cell wall degradation from the genome sequence of the soil bacterium Cellvibrio japonicus. J Bacteriol. 190(15):5455–5463.

Desch HE, Dinwoodie JM. 1996. Timber Structure, Properties, Conversion and Use. London: MacMillan Press Ltd.

Desvaux M. 2005. The cellulosome of Clostridium cellulolyticum. Enzyme Microb Technol. 37(4):373–385.

Devi N, Adivitiya, Khasa YP. 2016. A combinatorial approach of N-terminus blocking and codon optimization strategies to enhance the soluble expression of recombinant hIL-7 in E. coli fed-batch culture. Appl Microbiol Biotechnol. 100(23):9979–9994.

Dilokpimol A, Mäkelä MR, Aguilar-Pontes MV, Benoit-Gelber I, Hildén KS, De Vries RP. 2016. Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. Biotechnol Biofuels. 9(1):1–18.

Dilokpimol A, Mäkelä MR, Cerullo G, Zhou M, Varriale S, Gidijala L, Brás JLA, Jütten P, Piechot A, Verhaert R, *et al*. 2018. Fungal glucuronoyl esterases: Genome mining based enzyme discovery and biochemical characterization. N Biotechnol. 40(Pt B):282–287.

Din N, Forsythe IJ, Burtnick LD, Gilkes NR, Miller RC, Warren RAJ, Kilburn DG. 1994. The cellulose-binding domain of endoglucanase A (CenA) from Cellulomonas fimi: evidence for the involvement of tryptophan residues in binding. Mol Microbiol. 11(4):747–755.

Ding S-Y, Bayer EA, Steiner D, Shoham Y, Lamed R. 1999. A novel cellulosomal scaffoldin from Acetivibrio cellulolyticus that contains a family 9 glycosyl hydrolase. J Bacteriol. 181(21):6720–6729.

Ding S-Y, Bayer EA, Steiner D, Shoham Y, Lamed R. 2000. A scaffoldin of the Bacteroides cellulosolvens cellulosome that contains 11 type II cohesins. J Bacteriol. 182(17):4915–4925.

Ding S-Y, Rincon MT, Lamed R, Martin JC, Mccrae SI, Aurilia V, Shoham Y, Bayer EA, Flint HJ. 2001. Cellulosomal Scaffoldin-Like Proteins from Ruminococcus flavefaciens. J Bacteriol. 183(6):1945–1953.

Dinwoodie JM. 2000. Timber: Its Nature and Behaviour. 2nd Editio. London: CRC Press.

Doi RH, Kosugi A. 2004. Cellulosomes: Plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol. 2(7):541–551.

Doi RH, Kosugi A, Murashima K, Tamaru Y, Han SO. 2003. Cellulosomes from Mesophilic Bacteria. J Bacteriol. 185(20):5907–5914.

Doi RH, Tamaru Y. 2001. The Clostridium cellulovorans cellulosome: an enzyme complex with plant cell wall degrading activity. Chem Rec. 1(1):24–32.

Dornez E, Gebruers K, Delcour JA, Courtin CM. 2009. Grain-associated xylanases: occurrence, variability, and implications for cereal processing. Trends Food Sci Technol. 20(11–12):495–510.

Dornez E, Joye IJ, Gebruers K, Delcour JA, Courtin CM. 2006. Wheat-kernel-associated endoxylanases consist of a majority of microbial and a minority of wheat endogenous endoxylanases. J Agric Food Chem. 54(11):4028–4034.

Du X, Gellerstedt G, Li J. 2013. Universal fractionation of lignin–carbohydrate complexes (LCCs) from lignocellulosic biomass: an example using spruce wood. Plant J. 74(2):328–338.

Dusel G, Kluge H, Jeroch H, Simon O. 1998. Xylanase Supplementation of Wheat-Based Rations for Broilers: Influence of Wheat Characteristics. J Appl Poult Res. 7(2):119–131.

Ebringerová A. 2006. Structural diversity and application potential of hemicelluloses. Macromol Symp. 232(333):1–12.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 26(19):2460–2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 27(16):2194–2200.

Eiberle MK, Jungbauer A. 2010. Technical refolding of proteins: Do we have freedom to operate? Biotechnol J. 5(6):547–559.

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. Acta Crystallogr Sect D Biol Crystallogr. 66(4):486–501.

Engelhardt H, Peters J. 1998. Structural research on surface layers: A focus on stability, surface layer homology domains, and surface layer-cell wall interactions. J Struct Biol. 124(2–3):276–302.

Evans PR, Murshudov GN. 2013. How good are my data and what is the resolution? Acta Crystallogr Sect D Biol Crystallogr. D69(7):1204–1214.

Faber TA, Dilger RN, Iakiviak M, Hopkins AC, Price NP, Fahey GC. 2012. Ingestion of a novel galactoglucomannan oligosaccharide-arabinoxylan (GGMO-AX) complex affected growth performance and fermentative and immunological characteristics of broiler chicks challenged with Salmonella typhimurium. Poult Sci. 91(9):2241–2254.

Fahnert B, Lilie H, Neubauer P. 2004. Inclusion Bodies: Formation and Utilisation. Adv Biochem Eng Biotechnol. 89:93–142.

Faix BO. 1991. Classification of Lignins from Different Botanical Origins by FT-IR Spectroscopy. Holzforschung. 45(s1):21–28.

Fang Z, Schubert WD, Gey van Pittius NC. 2016. Expression and production of soluble Mycobacterium tuberculosis H37Rv mycosin-3. Biochem Biophys Reports. 5:448–452.

FAO/WHO. 2002. Evaluation of health and nutritional properties of probiotics in food including powder milk with live lactic acid bacteria. Jt FAO/WHO Expert Consult.

Faulds CB. 2010. What can feruloyl esterases do for us? Phytochem Rev. 9(1):121–132.

Feilmeier BJ, Iseminger G, Schroeder D, Webber H, Phillips GJ. 2000. Green Fluorescent Protein Functions as a Reporter for Protein Localization in Escherichia coli. J Bacteriol. 182(14):4068–4076.

Felsenstein J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics. 5:164–166.

Fengler AI, Marquardt RR. 1988. Water-soluble pentosans from rye. II. Effects on rate of dialysis and on the retention of nutrients by the chick. Cereal Chem. 65:298–302.

Fernandes VO, Costa M, Ribeiro T, Serrano L, Cardoso V, Santos H, Lordelo M, Ferreira LMA, Fontes CMGA. 2016. 1,3-1,4-β-Glucanases and not 1,4-β-glucanases improve the nutritive value of barley-based diets for broilers. Anim Feed Sci Technol. 211:153–163.

Figueiredo AA, Correia BA, Ribeiro T, Ponte PIP, Falcão L, Freire JP, Prates JAM, Ferreira LMA, Fontes CMGA, Lordelo MM. 2012. The effects of restricting enzyme supplementation in wheat-based diets to broilers. Anim Feed Sci Technol. 172(3–4):194–200.

Fischer B, Sumner I, Goodenough P. 1993. Isolation and renaturation of bio-active proteins expressed in Escherichia coli as inclusion bodies. Arzneimittelforschung. 41(12):3–13.

Fontes CMGA, Gilbert HJ. 2010. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu Rev Biochem. 79:655–681.

Fontes CMGA, Gilbert HJ, Hazlewood GP, Clarke JH, Prates JAM, McKie VA, Nagy T, Fernandes TH, Ferreira LMA. 2000. A novel Cellvibrio mixtus family 10 xylanase that is both intracellular and expressed under non-inducing conditions. Microbiology. 146(8):1959–1967.

Fontes CMGA, Hall J, Hirst BH, Hazlewood GP, Gilbert HJ. 1995. The resistance of cellulases and xylanases to proteolytic inactivation. Appl Microbiol Biotechnol. 43(1):52–57.

Fontes CMGA, Ponte PIP, Reis TC, Soares MC, Gama LT, Dias FM V, Ferreira LMA. 2004. A family 6 carbohydrate-binding module potentiates the efficiency of a recombinant xylanase used to supplement cereal-based diets for poultry. Br Poult Sci. 45(5):648–656.

Francis Schaeffer, Matuschek M, Guglielmi G, Miras I, Alzari PM, Pierre Béguin. 2002. Duplicated Dockerin Subdomains of Clostridium thermocellum Endoglucanase CelD Bind to a Cohesin Domain of the Scaffolding Protein CipA with Distinct Thermodynamic Parameters and a Negative Cooperativity. Biochemistry. 41(7):2106–2114.

Freire JPB, Guerreiro AJG, Cunha LF, Aumaitre A. 2000. Effect of dietary fibre source on total tract digestibility, caecum volatile fatty acids and digestive transit time in the weaned piglet. Anim Feed Sci Technol. 87(1–2):71–83.

Fries M, Ihrig J, Brocklehurst K, Shevchik VE, Pickersgill RW. 2007. Molecular basis of the activity of the phytopathogen pectin methylesterase. EMBO J. 26(17):3879–3887.

Frommhagen M, Westphal AH, van Berkel WJH, Kabel MA. 2018. Distinct substrate specificities and electron-donating systems of fungal lytic polysaccharide monooxygenases. Front Microbiol. 9:1–22.

Fry SC. 1989. Cellulases, Hemicellulases and Auxin-stimulated Growth - A possible relationship. Physiol Plant. 75(4):532–536.

Fry SC, Nesselrode BHWA, Miller JG, Mewburn BR. 2008. Mixed-linkage (1→3,1→4)-β-D-glucan is a major hemicellulose of Equisetum (horsetail) cell walls. New Phytol. 179(1):104–115.

Fuente JM, Ayala PP, Flores A, Villamide MJ. 1998. Effect of Storage Time and Dietary Enzyme on the Metabolizable Energy and Digesta Viscosity of Barley-Based Diets for Poultry. Poult Sci. 77(1):90–97.

Gardiner ES, Sarko A. 1985. Packing analysis of carbohydrates and polysaccharides. 16. The crystal structures of celluloses IV I and IV II. Can J Chem. 63(1):173–180.

Garron ML, Cygler M. 2010. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. Glycobiology. 20(12):1547–1573.

Gebruers K, Dornez E, Bedõ Z, Rakszegi M, Courtin CM, Delcour JA. 2010. Variability in xylanase and xylanase inhibition activities in different cereals in the HEALTHGRAIN diversity screen and contribution of environment and genotype to this variability in common wheat. J Agric Food Chem. 58(17):9362–9371.

Gerngross UT, Romaniec MPM, Kobayashi T, Huskisson NS, Demain AL. 1993. Sequencing of a Clostridium thermocellum gene cipA encoding the cellulosomal SL-protein reveals an unusual degree of internal homology. Mol Microbiol. 8(2):325–334.

Gibson GR, Probert HM, Loo J Van, Rastall RA, Roberfroid MB. 2004. Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. Nutr Res Rev. 17(02):259.

Gibson GR, Roberfroid MB. 1995. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. J Nutr. 125(6):1401–1412.

Gilbert HJ. 2010. The biochemistry and structural biology of plant cell wall deconstruction. Plant Physiol. 153(2):444–55.

Goh CS, Lan N, Douglas SM, Wu B, Echols N, Smith A, Milburn D, Montelione GT, Zhao H, Gerstein M. 2004. Mining the Structural Genomics Pipeline: Identification of Protein Properties that Affect High-throughput Experimental Analysis. J Mol Biol. 336(1):115–130.

Gonzalo G, Colpa DI, Habib MHM, Fraaije MW. 2016. Bacterial enzymes involved in lignin degradation. J Biotechnol. 236:110–119.

Gopalan N, Rodríguez-Duran L V., Saucedo-Castaneda G, Nampoothiri KM. 2015. Review on technological and scientific aspects of feruloyl esterases: A versatile enzyme for biorefining of biomass. Bioresour Technol. 193.

Gordon AH, Lomax JA, Dalgarno K, Chesson A. 1985. Preparation and composition of mesophyll, epidermis and fibre cell walls from leaves of perennial ryegrass (lolium perenne) and italian ryegrass (lolium multiflorum). J Sci Food Agric. 36(7):509–519.

Gray KA, Zhao L, Emptage M. 2006. Bioethanol. Curr Opin Chem Biol. 10:141–146.

Grizard D, Barthomeuf C. 1999. Non-digestible oligosaccharides used as prebiotic agents: mode of production and beneficial effects on animal and human health. Reprod Nutr Dev. 39(5–6):563–588.

de Groot NS, Espargarö A, Morell M, Ventura S. 2008. Studies on bacterial inclusion bodies. Future Microbiol. 3(4):423–435.

Guglielmetti S, De Noni I, Caracciolo F, Molinari F, Parini C, Mora D. 2008. Bacterial cinnamoyl esterase activity screening for the production of a novel functional food product. Appl Environ Microbiol. 74(4):1284–1288.

Guillén D, Sánchez S, Rodríguez-Sanoja R. 2010. Carbohydrate-binding domains: multiplicity of biological roles. Appl Microbiol Biotechnol. 85(5):1241–1249.

Gunnoo M, Cazade PA, Galera-Prat A, Nash MA, Czjzek M, Cieplak M, Alvarez B, Aguilar M, Karpol A, Gaub H, *et al*. 2016. Nanoscale Engineering of Designer Cellulosomes. Adv Mater. 28(27):5619–5647.

Ha M a., Apperley DC, Jarvis MC. 1997. Molecular Rigidity in Dry and Hydrated Onion Cell Walls. Plant Physiol. 115(2):593–598.

Hajati H, Hassanabadi A, Teimouri Yansari A. 2014. The Effect of Dietary Supplementation of Prebiotic and Probiotic on Performance, Humoral Immunity Responses and Egg Hatchability in Broiler Breeders. Poult Sci J. 2(1):1–13.

Hajati H, Rezaei M. 2010. The application of prebiotics in poultry production. Int J Poult Sci. 9(3):298–304.

Hallac BB, Ragauskas AJ. 2011. Analyzing cellulose degree of polymerization and its relevancy to cellulosic ethanol. Biofuels, Bioprod Biorefining. 5(2):215–225.

Hamberg Y, Ruimy-Israeli V, Dassa B, Barak Y, Lamed R, Cameron K, Fontes CMGA, Bayer EA, Fried DB. 2014. Elaborate cellulosome architecture of Acetivibrio cellulolyticus revealed by selective screening of cohesin–dockerin interactions. PeerJ. 2:e636.

Harholt J, Suttangkakul A, Scheller HV. 2010. Biosynthesis of pectin. Plant Physiol. 153(2):384–95.

Harris P V., Welner D, McFarland KC, Re E, Navarro Poulsen JC, Brown K, Salbo R, Ding H, Vlasenko E, Merino S, *et al*. 2010. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: Structure and function of a large, enigmatic family. Biochemistry. 49(15):3305–3316.

Hart DA, Kindel PK. 1970. Isolation and partial characterization of apiogalacturonans from the cell wall of Lemna minor. Biochem J. 116(4):569–579.

Hashimoto H. 2006. Recent structural studies of carbohydrate-binding modules. Cell Mol Life Sci. 63(24):2954–2967.

Hayashi J, Sufoka A, Ohkita J, Watanabe S. 1975. The confirmation of existences of cellulose IIII, IIIII, IVI, and IVII by the X-ray method. J Polym Sci Polym Lett Ed. 13(1):23–27.

Hayashi T. 1989. Xyloglucans in the Primary Cell Wall. Annu Rev Plant Physiol Plant Mol Biol. 40(1):139–168.

Heiner AP, Sugiyama J, Teleman O. 1995. Crystalline cellulose Iα and Iβ studied by molecular dynamics simulation. Carbohydr Res. 273(2):207–223.

Hellín P, Ralet MC, Bonnin E, Thibault JF. 2005. Homogalacturonans from lime pectins exhibit homogeneous charge density and molar mass distributions. Carbohydr Polym. 60(3):307–317.

Henrissat B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. 280(Pt 2):309–316.

Henrissat B, Bairoch A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. 293(3):781–788.

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

Henrissat B, Coutinho PM, Lombard V, Drula E, Garron M-L, Terrapon N, Lapebie P. 2019. CAZY - Carbohydrate-Active enZYmes. [accessed 2019 Jul 19]. http://www.cazy.org/.

Henrissat B, Daviest G. 1997. Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol. 7(5):637–644.

Henshaw JL, Bolam DN, Pires VMR, Czjzek M, Henrissat B, Ferreira LMA, Fontes CMGA, Gilbert HJ. 2004. The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. J Biol Chem. 279(20):21552–21559.

Hervé C, Rogowski A, Blake AW, Marcus SE, Gilbert HJ, Knox JP. 2010. Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. PNAS. 107(34):15293–8.

Hetland H, Choct M, Svihus B. 2004. Role of insoluble non-starch polysaccharides in poultry nutrition. Worlds Poult Sci J. 60(04):415–422.

Hetland H, Svihus B. 2001. Effect of oat hulls on performance, gut capacity and feed passage time in broiler chickens. Br Poult Sci. 42(3):354–361.

Hetland H, Svihus B, Olaisen V. 2002. Effect of feeding whole cereals on performance, starch digestibility and duodenal particle size distribution in broiler chickens. Br Poult Sci. 43(3):416–423.

Hetland H, Uhlen AK, Viken KHK, Krekling T, Svihus B. 2007. Hagberg falling number and the nutritional value of wheat in broiler chicken diets. Br Poult Sci. 48(1):12–20.

Himmel ME, Bayer EA. 2009. Lignocellulose conversion to biofuels: current challenges, global perspectives. Curr Opin Biotechnol. 20(3):316–317.

Hoffman M, Jia Z, Peña MJ, Cash M, Harper A, Blackburn AR, Darvill A, York WS. 2005. Structural analysis of xyloglucans in the primary cell walls of plants in the subclass Asteridae. Carbohydr Res. 340(11):1826–1840.

Holm L, Rosenström P. 2010. Dali server: Conservation mapping in 3D. Nucleic Acids Res. 38(SUPPL. 2):545–549.

Hsu C-K, Liao J-W, Chung Y-C, Hsieh C-P, Chan Y-C. 2004. Xylooligosaccharides and fructooligosaccharides affect the intestinal microbiota and precancerous colonic lesion development in rats. J Nutr. 134(6):1523–1528.

Hu J, Chandra R, Arantes V, Gourlay K, Susan van Dyk J, Saddler JN. 2015. The addition of accessory enzymes enhances the hydrolytic performance of cellulase enzymes at high solid loadings. Bioresour Technol. 186:149–53.

Hu Y, Walker S. 2002. Remarkable Structural Similarities between Diverse Glycosyltransferases. Chem Biol. 9(02):1287–1296.

Huang C, He J, Li X, Min D, Yong Q. 2015. Facilitating the enzymatic saccharification of pulped bamboo residues by degrading the remained xylan and lignin–carbohydrates complexes. Bioresour Technol. 192:471–477.

Huang XF, Santhanam N, Badri D V., Hunter WJ, Manter DK, Decker SR, Vivanco

JM, Reardon KF. 2013. Isolation and characterization of lignin-degrading bacteria from rainforest soils. Biotechnol Bioeng. 110(6):1616–1626.

Hüttner S, Klaubauf S, de Vries RP, Olsson L. 2017. Characterisation of three fungal glucuronoyl esterases on glucuronic acid ester model compounds. Appl Microbiol Biotechnol. 101(13):5301–5311.

Huyen DT, Giang LQ, Hai TN. 2014. Expression of flagellin FLjB derived from Salmonella enterica serovar typhimurium in Escherichia coli BL21. Tap Chi Sinh Hoc. 36(4):506–514.

Iiyama K, Lam T, Stone BA. 1994. Covalent Cross-Links in the Cell Wall. Plant Physiol. 104(2):315–320.

Ikegami S, Tsuchihashi F, Harada H, Tsuchihashi N, Nishide E, Innami S. 1990. Effect of viscous indigestible polysaccharides on pancreatic biliary secreti on and digestive organs in rats. J Nutr. 120(4):353–360.

Imaizumi K, Nakatsu Y, Sato M, Sedarnawati Y, Sugano M. 1991. Effects of xylooligosaccharides on blood glucose, serum and liver lipids and cecum short-chain fatty acids in diabetic rats. Agric Biol Chem. 55(1):199–205.

Ishii T. 1997. Structure and functions of feruloylated polysaccharides. Plant Sci. 127(2):111–127.

Jacquet G, Pollet B, Lapierre C, Mhamdi F, Rolando C. 1995. New Ether-Linked Ferulic Acid-Coniferyl Alcohol Dimers Identified in Grass Straws. J Agric Food Chem. 43(10):2746–2751.

Jamar C, Jardin P Du, Fauconnier ML. 2011. Cell wall polysaccharides hydrolysis of malting barley (Hordeum vulgare L.): a review. Biotechnol Agron Soc Environ. 15(2):301–313.

Jamroz D, Jakobsen K, Bach Knudsen KE, Wiliczkiewicz A, Orda J. 2002. Digestibility and energy value of non-starch polysaccharides in young chickens, ducks and geese, fed diets containing high amounts of barley. Comp Biochem Physiol - A Mol Integr Physiol. 131(3):657–668.

Jarrett J. Adams, Pal G, Jia Z, Smith SP. 2006. Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin–dockerin complex. PNAS. 103(2):305–310.

Jaskari J, Kontula P, Siitonen A, Jousimies-Somer H, Mattila-Sandholm T, Poutanen K. 1998. Oat β-glucan and xylan hydrolysates as selective substrates for Bifidobacterium and Lactobacillus strains. Appl Microbiol Biotechnol. 49(2):175–181.

Johansen KS. 2016. Lytic Polysaccharide Monooxygenases: The Microbial Power Tool for Lignocellulose Degradation. Trends Plant Sci. 21(11):926–936.

Johnson IT, Gee JM. 1986. Gastrointestinal adaptation in response to soluble non-available polysaccharides in the rat. Br J Nutr. 55(3):497–505.

Jommuengbout P, Pinitglang S, Kyu KL, Ratanakhanokchai K. 2009. Substrate-

Binding Site of Family 11 Xylanase from Bacillus firmus K-1 by Molecular Docking. Biosci Biotechnol Biochem. 73(4):833–839.

Józefiak D, Rutkowski A, Jensen BB, Engberg RM. 2006. The effect of β-glucanase supplementation of barley- and oat-based diets on growth performance and fermentation in broiler chicken gastrointestinal tract. Br Poult Sci. 47(1):57–64.

Józefiak D, Rutkowski A, Jensen BB, Engberg RM. 2007. Effects of dietary inclusion of triticale, rye and wheat and xylanase supplementation on growth performance of broiler chickens and fermentation in the gastrointestinal tract. Anim Feed Sci Technol. 132(1–2):79–93.

Juturu V, Wu JC. 2012. Microbial xylanases: Engineering, production and industrial applications. Biotechnol Adv. 30(6):1219–1227.

Kabel MA, Kortenoeven L, Schols HA, Voragen AGJ. 2002. In vitro fermentability of differently substituted xylo-oligosaccharides. J Agric Food Chem. 50(21):6205–6210.

Kajihara M, Kato S, Konishi M, Yamagishi Y, Horie Y, Ishii H. 2000. Xylooligosaccharide decreases blood ammonia levels in patients with liver cirrhosis. Am J Gastroenterol. 95(9):2514.

Kakiuchi M, Isui A, Suzuki K, Fujino T, Fujino E, Kimura T, Karita S, Sakka K, Ohmiya K. 1998. Cloning and DNA sequencing of the genes encoding Clostridium josui scaffolding protein CipA and cellulase celD and identification of their gene products as major components of the cellulosome. J Bacteriol. 180(16):4303–4308.

Karpol A, Jobby MK, Slutzki M, Noach I, Chitayat S, Smith SP, Bayer EA. 2013. Structural and functional characterization of a novel type-III dockerin from Ruminococcus flavefaciens. FEBS Lett. 587(1):30–36.

Katsimpouras C, Bénarouche A, Navarro D, Karpusas M, Dimarogona M, Berrin JG, Christakopoulos P, Topakas E. 2014. Enzymatic synthesis of model substrates recognized by glucuronoyl esterases from Podospora anserina and Myceliophthora thermophila. Appl Microbiol Biotechnol. 98(12):5507–5516.

Kaur Jashandeep, Kumar A, Kaur Jagdeep. 2018. Strategies for optimization of heterologous protein expression in E. coli: Roadblocks and reinforcements. Int J Biol Macromol. 106:803–822.

Kawai S, Nakagawa M, Ohashi H. 2002. Degradation mechanisms of a nonphenolic β-O-4 lignin model dimer by Trametes versicolor laccase in the presence of 1-hydroxybenzotriazole. Enzyme Microb Technol. 30(4):482–489.

Keegstra K. 2010. Plant cell walls. Plant Physiol. 154(2):483–486.

Keegstra K, Talmadge KW, Bauer WD, Albersheim P. 1973. The structure of plant cell walls. III. A model of the walls of suspension-cultured sycamore cells based on the interconnections of the macromolecular components. Plant Physiol. 51(1):188–196.

Khoury D El, Cuda C, Luhovyy BL, Anderson GH. 2011. Beta Glucan: Health Benefits in Obesity and Metabolic Syndrome. J ofNutrition Metab. 2012:1–28.

Kim GB, Seo YM, Kim CH, Paik IK. 2011. Effect of dietary prebiotic supplementation on the performance, intestinal microflora, and immune response of broilers. Poult Sci. 90(1):75–82.

Kim S, Lee SB. 2006. Rare codon clusters at 5′-end influence heterologous expression of archaeal gene in Escherichia coli. Protein Expr Purif. 50(1):49–57.

Kim S, Lee SB. 2008. Soluble expression of archaeal proteins in Escherichia coli by using fusion-partners. Protein Expr Purif. 62(1):116–119.

Kim S, Shoda M. 1999. Purification and Characterization of a Novel Peroxidase from Geotrichum candidum Dec 1 Involved in Decolorization of Dyes. Society. 65(3):1029–1035.

Kirby J, Martin JC, Daniel AS, Flint HJ. 1997. Dockerin-like sequences in cellulases and xylanases from the rumen cellulolytic bacterium Ruminococcus flavefaciens. FEMS Microbiol Lett. 149(2):213–219.

Van Der Klis JD, Van Voorst A, Van Cruyningen C. 1993. Effect of a soluble polysaccharide (carboxy methyl cellulose) on the physico-chemical conditions in the gastrointestinal tract of broilers. Br Poult Sci. 34(5):971–983.

Korepanova A, Gao FP, Hua Y, Qin H, Nakamoto RK, Cross TA. 2005. Cloning and expression of multiple integral membrane proteins from Mycobacterium tuberculosis in Escherichia coli. Protein Sci. 14(1):148–158.

Kötzler MP, Hancock SM, Withers SG. 2014. Glycosidases: Functions, Families and Folds. eLs.:1–14.

Kroon P, Garcia-Conesa M, Fillingham I, Hazlewood G, Williamson G. 1999. Release of ferulic acid dehydrodimers from plant cell walls by feruloyl esterases. J Sci Food Agric. 79(3):428–434.

Kroon PA, Williamson G, Fish NM, Archer DB, Belshaw NJ. 2000. A modular esterase from Penicillium funiculosum which releases ferulic acid from plant cell walls and binds crystalline cellulose contains a carbohydrate binding module. Eur J Biochem. 267(23):6740–6752.

Laemmli UK. 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature. 227(5259):680–685.

van Laere KMJ, Hartemink R, Bosveld M, Schols HA, Voragen AGJ. 2000. Fermentation of plant cell wall derived polysaccharides and their corresponding oligosaccharides by intestinal bacteria. J Agric Food Chem. 48(5):1644–1652.

Lai GH, Lin YC, Tsai YL, Lien YY, Lin MK, Chen HJ, Chang W Te, Tzen JTC, Lee MS. 2014. High yield production of pigeon circovirus capsid protein in the E. coli by evaluating the key parameters needed for protein expression. BMC Vet Res. 10(1):1–11.

Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: Structures, Functions, and Mechanisms. Annu Rev Biochem. 77(1):521–555.

Lakey DL, Voladri RKR, Edwards KM, Hager C, Samten B, Wallis RS, Peter F, Kernodle DS. 2000. Enhanced Production of Recombinant Mycobacterium tuberculosis

Antigens in Escherichia coli by Replacement of Low-Usage Codons. Infect Immun. 68(1):233–238.

Lam TBT, Kadoya K, Iiyama K. 2001. Bonding of hydroxycinnamic acids to lignin: Ferulic and p-coumaric acids are predominantly linked at the benzyl position of lignin, not the β-position, in grass cell walls. Phytochemistry. 57(6):987–992.

Lamed R, Naimark J, Morgenstern E, Bayer EA. 1987. Specialized cell surface structures in cellulolytic bacteria. J Bacteriol. 169(8):3792–3800.

Lamed R, Setter E, Bayer EA. 1983. Characterization of a cellulose-binding, cellulase-containing complex in clostridium thermocellum. J Bacteriol. 156(2):828–836.

Lan Y, Verstegen MWA, Tamminga S, Williams BA. 2005. The role of the commensal gut microbial community in broiler chickens. Worlds Poult Sci J. 61(1):95–104.

Lan Y, Williams BA, Verstegen MWA, Patterson R, Tamminga S. 2007. Soy oligosaccharides in vitro fermentation characteristics and its effect on caecal microorganisms of young broiler chickens. Anim Feed Sci Technol. 133(3–4):286–297.

Lau JM, McNeil M, Darvill AG, Albersheim P. 1987. Treatment of rhamnogalacturonan I with lithium in ethylenediamine. Carbohydr Res. 168(2):245–274.

Lázaro R, García M, Araníbar MJ, Mateos GG. 2003. Effect of enzyme addition to wheat- , barley- and rye-based diets on nutrient digestibility and performance of laying hens. Br Poult Sci. 44:256–265.

Lemaire M, Ohayon H, Gounon P, Fujino T, Béguin P. 1995. OlpB, a new outer layer protein of Clostridium thermocellum, and binding of its S-layer-like domains to components of the cell envelope. J Bacteriol. 177(9):2451–2459.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels. 6(1):1–14.

Li XL, Špániková S, de Vries RP, Biely P. 2007. Identification of genes encoding microbial glucuronoyl esterases. FEBS Lett. 581(21):4029–4035.

Liers C, Bobeth C, Pecyna M, Ullrich R, Hofrichter M. 2010. DyP-like peroxidases of the jelly fungus Auricularia auricula-judae oxidize nonphenolic lignin model compounds and high-redox potential dyes. Appl Microbiol Biotechnol. 85(6):1869–1879.

Liu J, Arcady M. 2003. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. Protein Sci. 12:1418–1431.

Lodish H, Berk A, Zipursky S. 2000. The Dynamic Plant Cell Wall. In: Freeman WH, editor. Molecular Cell Biology. 4th ed. New York.

Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B. 2010. A hierarchical classification of polysaccharide lyases for glycogenomics. Biochem J. 432(3):437–444.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The

carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42(DI):490–495.

Longland JM, Fry SC, Trewavas AJ. 1989. Developmental Control of Apiogalacturonan Biosynthesis and UDP-Apiose Production in a Duckweed. Plant Physiol. 90(3):972–976.

Luis AS, Briggs J, Zhang X, Farnell B, Ndeh D, Labourel A, Baslé A, Cartmell A, Terrapon N, Stott K, *et al*. 2018. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. Nat Microbiol. 3(2):210–219.

Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J, *et al*. 2008. How biotech can transform biofuels. Nat Biotechnol. 26(2):169–172.

Lynd LR, Weimer PJ, Zyl WH van, Pretorius IS. 2002. Microbial Cellulose Utilization: Fundamentals and Biotechnology. Microbiol Mol Biol Rev. 66(3):506–577.

Macfarlane S, Macfarlane GT, Cummings JH. 2006. Review article: Prebiotics in the gastrointestinal tract. Aliment Pharmacol Ther. 24(5):701–714.

Maeda M, Shimahara H, Sugiyama N. 1980. Detailed examination of the branched structure of konjac glucomannan. Agric Biol Chem. 44(2):245–252.

De Maesschalck C, Eeckhaut V, Maertens L, De Lange L, Marchal L, Nezer C, De Baere S, Croubels S, Daube G, Dewulf J, *et al*. 2015. Effects of Xylo-oligosaccharides on broiler chicken performance and microbiota. Appl Environ Microbiol. 81(17):5880–5888.

Mahammad S, Prud'homme RK, Roberts GW, Khan SA. 2006. Kinetics of enzymatic depolymerization of guar galactomannan. Biomacromolecules. 7(9):2583–2590.

Mannall GJ, Titchener-Hooker NJ, Dalby PA. 2007. Factors Affecting Protein Refolding Yields in a Fed-Batch and Batch-Refolding System. Biotechnol Bioeng. 97(6):1523–1534.

Marquardt RR, Brenes A, Zhang Z, Boros D. 1996. Use of enzymes to improve nutrient availability in poultry feedstuffs. Anim Feed Sci Technol. 60(3–4):321–330.

Marrinan HJ, Mann J. 1956. Infrared spectra of the crystalline modifications of cellulose. J Polym Sci. 21(98):301–311.

McCann MC, Carpita NC. 2008. Designing the deconstruction of plant cell walls. Curr Opin Plant Biol. 11(3):314–320.

McCarter JD, Withers SG. 1994. Mechanisms of enzymatic glycoside hydrolysis. Curr Opin Struct Biol. 4(6):885–892.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software . J Appl Crystallogr. 40(4):658–674.

McCracken KJ, Bedford MR, Stewart RA. 2001. Effects of variety, the 1B/1R translocation and xylanase supplementation on nutritive value of wheat for broilers. Br Poult Sci. 42(5):638–642.

McCracken KJ, Owens B, Park R, McNab J. 2008. Lack of relationship between either specific weight or presence of the 1B1R gene and nutritive value of wheat in broiler diets. Br Poult Sci. 49(4):463–474.

McLauchlan WR, Garcia-Conesa MT, Williamson G, Roza M, Ravestein P, Maat J. 1999. A novel class of protein from wheat which inhibits xylanases. Biochem J. 338(Pt 2):441–446.

McNeil M, Darvill AG, Albersheim P. 1980. The Structure of Plant Cell Walls: X. Rhamnogalacturonan I, a structurally complex pectic polysaccharide in the wall of suspension-cultured sycamore cells. Plant Physiol. 66(6):1128–1134.

Miller GL. 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal Chem. 31(3):426–428.

Miras I, Schaeffer F, Béguin P, Alzari PM. 2002. Mapping by site-directed mutagenesis of the region responsible for cohesin - Dockerin interaction on the surface of the seventh cohesin domain of Clostridium thermocellum CipA. Biochemistry. 41(7):2115–2119.

Mirzaei M, Saffar B, Shareghi B. 2016. Cloning, Codon Optimization, and Expression of Yersinia intermedia Phytase Gene in E. coli. Iran J Biotechnol. 14(2):63–69.

Mogk A, Mayer MP, Deuerling E. 2002. Mechanisms of Protein Folding : Molecular Chaperones and Their Application in Biotechnology. Chembiochem. 3:807–814.

Mohnen D. 2008. Pectin structure and biosynthesis. Curr Opin Plant Biol. 11(3):266–277.

Moniz P, Ho AL, Duarte LC, Kolida S, Rastall RA, Pereira H, Carvalheiro F. 2016. Assessment of the bifidogenic effect of substituted xylo-oligosaccharides obtained from corn straw. Carbohydr Polym. 136:466–473.

Monrad RN, Eklöf J, Krogh KBRM, Biely P. 2018. Glucuronoyl esterases: diversity, properties and biotechnological potential. A review. Crit Rev Biotechnol. 38(7):1121–1136.

Monsan P, Paul F. 1995. Oligosaccharide feed additives. In: Wallace RJ, Chesson A, editors. Biotechnology in Animal Feeds and Animal Feeding. New York: VCH. p. 233–245.

Morag E, Lapidot A, Govorko D, Lamed R, Wilchek M, Bayer EA, Shoham Y. 1995. Expression, purification, and characterization of the cellulose-binding domain of the scaffoldin subunit from the cellulosome of Clostridium thermocellum. Appl Environ Microbiol. 61(5):1980–1986.

Moreira LRS, Filho EXF. 2008. An overview of mannan structure and mannan-degrading enzyme systems. Appl Microbiol Biotechnol. 79(2):165–178.

Morgan NK, Wallace A, Bedford MR, Choct M. 2017. Efficiency of xylanases from families 10 and 11 in production of xylo-oligosaccharides from wheat arabinoxylans. Carbohydr Polym. 167:290–296.

Moura P, Barata R, Carvalheiro F, Gírio F, Loureiro-Dias MC, Esteves MP. 2007. In vitro fermentation of xylo-oligosaccharides from corn cobs autohydrolysis by Bifidobacterium and Lactobacillus strains. LWT - Food Sci Technol. 40(6):963–972.

Mourão JL, Ponte PIP, Prates JAM, Centeno MSJ, Ferreira LMA, Soares MAC, Fontes CMGA. 2006. Use of β-glucanases and β-1,4-xylanases to supplement diets containing alfalfa and rye for laying hens: Effects on bird performance and egg quality. J Appl Poult Res. 15(2):256–265.

Murphy TC, McCracken JK, McCann MEE, George J, Bedford MR. 2009. Broiler performance and in vivo viscosity as influenced by a range of xylanases, varying in ability to effect wheat in vitro viscosity. Br Poult Sci. 50(6):716–724.

Murray AG, Fuller MF, Pirie RA. 1977. The effect of fibre in the form of various polysaccharides on the apparent digestibility of protein in the pig. Anim, Prod. 24:139.

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 67(Pt 4):355–367.

Na MH, Kim WK. 2007. Effects of xylooligosaccharide intake on fecal Bifidobacteria, lactic acid and lipid metabolism in Korean young women. Korean J Nutr. 40:154–161.

Nahas J, Lefrancois MR. 2001. Effects of feeding locally grown whole barley with or without enzyme addition and whole wheat on broiler performance and carcass traits. Poult Sci. 80(2):195–202.

Nakamura AM, Nascimento AS, Polikarpov I. 2017. Structural diversity of carbohydrate esterases. Biotechnol Res Innov. 1(1):35–51.

Newman RK, Newman CW. 1988. Nutritive value of a new hull-less barley cultivar in broiler chick diets. Poult Sci. 67(11):1573–1579.

Nguyen VD, Hatahet F, Salo KEH, Enlund E, Zhang C, Ruddock LW. 2011. Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli. Microb Cell Fact. 10(1):1–13.

Niu X, Li N, Chen D, Wang Z. 2013. Interconnection Between the Protein Solubility and Amino Acid and Dipeptide Compositions. Protein Pept Lett. 20(1):88–95.

Noach I, Frolow F, Jakoby H, Rosenheck S, Shimon LJW, Lamed R, Bayer EA. 2005. Crystal structure of a type-II cohesin module from the Bacteroides cellulosolvens cellulosome reveals novel and distinctive secondary structural elements. J Mol Biol. 348(1):1–12.

NRC. 1994. Nutrient Requirements of Poultry. 9 th. Washington, DC.: Natl. Acad. Press.

O'Sullivan AC. 1997. Cellulose: the structure slowly unravels. Cellulose. 4(3):173–207.

Ohara H, Karita S, Kimura T, Sakka K, Ohmiya K. 2000. Characterization of the cellulolytic complex (cellulosome) from Ruminococcus albus. Biosci Biotechnol Biochem. 64(2):254–260.

Okazaki M, Fujikawa S, Matsumoto N. 1990. Effect of Xylooligosaccharide on the Growth of Bifidobacteria. Bifidobact Microflora. 9(2):77–86.

Okazaki M, Koda H, Izumi R, Fujikawa S, Matsumoto N. 1991. In vitro Digestibility and in vivo Utilization of Xylobiose. J Jap Soc Nutr Food Sci. 44:41–44.

Oliveira DM, Finger-Teixeira A, Rodrigues Mota T, Salvador VH, Moreira-Vilar FC, Correa Molinari HB, Craig Mitchell RA, Marchiosi R, Ferrarese-Filho O, Dantas dos Santos W. 2015. Ferulic acid: A key component in grass lignocellulose recalcitrance to hydrolysis. Plant Biotechnol J. 13(9):1224–1232.

Orban JI, Patterson JA, Sutton AL, Richards GN. 1997. Effect of Sucrose Thermal Oligosaccharide Caramel, Dietary Vitamin-Mineral Level, and Brooding Temperature on Growth and Intestinal Bacterial Populations of Broiler Chickens. Poult Sci. 76(3):482–490.

Ovodov YS, Ovodova RG, Bondarenko OD, Krasikova IN. 1971. The pectic substances of zosteraceae: Part IV. Pectinase digestion of zosterine. Carbohydr Res. 18(2):311–318.

Palacios JL, Zaror I, Martínez P, Uribe F, Opazo P, Socías T, Gidekel M, Venegas A. 2001. Subset of hybrid eukaryotic proteins is exported by the type I secretion system of Erwinia chrysanthemi. J Bacteriol. 183(4):1346–1358.

Palframan RJ, Gibson GR, Rastall RA. 2003. Carbohydrate preferences of Bifidobacterium species isolated from the human gut. Curr Issues Intest Microbiol. 4(2):71–75.

Parthasarathi R, Romero RA, Redondo A, Gnanakaran S. 2011. Theoretical Study of the Remarkably Diverse Linkages in Lignin. J Phys Chem Lett. 2(20):2660–2666.

Patterson JA, Orban JI, Sutton AL, Richards GN. 1997. Selective Enrichment of Bifidobacteria in the Intestinal Tract of Broilers by Thermally Produced Kestoses and Effect on Broiler Performance. Poult Sci. 76(3):497–500.

Petersen TN, Brunak S, Heijne G Von, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Publ Gr. 8(10):785–786.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera-A visualization system for exploratory research and analysis. J Comput Chem. 25(13):1605–12.

Pettersson D, Åman P. 1989. Enzyme supplementation of a poultry diet containing rye and wheat. Br J Nutr. 62(1):139–149.

Pettersson D, Åman P. 1993. Effects of Feeding Diets Based on Wheat Bread or Oat Bran Bread to Broiler Chickens. J Cereal Sci. 17(2):157–168.

Pickersgill R, Jenkins J, Harris G, Nasser W, Robert-Baudouy J. 1994. The structure of Bacillus subtilis pectate lyase in complex with calcium. Nat Struct Biol. 1(10):717–723.

Pinto C, Pinho D, Sousa S, Pinheiro M, Egas C, Gomes AC. 2014. Unravelling the diversity of grapevine microbiome. PLoS One. 9(1).

Pirgozliev VR, Birch CL, Rose SP, Kettlewell PS, Bedford MR. 2003. Chemical composition and the nutritive quality of different wheat cultivars for broiler chickens. Br Poult Sci. 44(3):464–475.

Piubelli L, Campa M, Temporini C, Binda E, Mangione F, Amicosante M, Terreni M, Marinelli F, Pollegioni L. 2013. Optimizing Escherichia coli as a protein expression platform to produce Mycobacterium tuberculosis immunogenic proteins. Microb Cell Fact. 12(1):1–14.

Piva A. 1998. Non-conventional feed additives. J Anim Feed Sci. 7(Suppl. 1):143–154.

Pohlschroder M, Canale-Parola E, Leschine SB. 1995. Ultrastructural diversity of the cellulase complexes of Clostridium papyrosolvens C7. J Bacteriol. 177(22):6625–6629.

Pokkuluri PR, Duke NEC, Wood SJ, Cotta MA, Li XL, Biely P, Schiffer M. 2011. Structure of the catalytic domain of glucuronoyl esterase Cip2 from Hypocrea jecorina. Proteins Struct Funct Bioinforma. 79(8):2588–2592.

Ponte PIP, Ferreira LMA, Soares MAC, Gama LT, Fontes CMGA. 2004. Xylanase inhibitors affect the action of exogenous enzymes used to supplement Triticum durum-based diets for broiler chicks. J Appl Poult Res. 13(4):660–666.

Poole DM, Morag E, Lamed R, Bayer EA, Hazlewood GP, Gilbert HJ. 1992. Identification of the cellulose-binding domain of the cellulosome subunit S1 from Clostridium thermocellum YS. FEMS Microbiol Lett. 99(2–3):181–186.

Prates J, Tarbouriech N, Charnock S, Fontes CMGA, Ferreira LMA, Davies GJ. 2001. The Structure of the Feruloyl Esterase Module of Xylanase 10B from Clostridium thermocellum Provides Insights into Substrate Recognition. Structure. 9(1):1183–1190.

Qamsari EM, Kasra KR, Nejad ZM. 2011. Isolation and identification of a novel, lipase-producing bacterium, Pseudomnas aeruginosa KM110. Iran J Microbiol. 3(2):92–98.

Ralet MC, Buffetto F, Capron I, Guillon F. 2016. Cell Wall Polysaccharides of Potato. In: Advances in Potato Chemistry and Technology: Second Edition. Academic Press. p. 33–56.

Reinikainen T, Ruohonen L, Nevanen T, Laaksonen L, Kraulis P, Jones TA, Knowles JKC, Teeri TT. 1992. Investigation of the function of mutated cellulose-binding domains of Trichoderma reesei cellobiohydrolase I. Proteins Struct Funct Bioinforma. 14(4):475–482.

Reiter WD. 2002. Biosynthesis and properties of the plant cell wall. Curr Opin Plant Biol. 5(6):536–542.

Ribeiro T, Lordelo MMS, Ponte PIP, Maçãs B, Prates JAM, Aguiar Fontes M, Falcão L, Freire JPB, Ferreira LMA, Fontes CMGA. 2011. Levels of endogenous ß-glucanase activity in barley affect the efficacy of exogenous enzymes used to supplement barley-based diets for poultry. Poult Sci. 90(6):1245–1256.

Ribeiro T, Lordelo MMS, Prates JAM, Falcão L, Freire JPB, Ferreira LMA, Fontes CMGA. 2012. The thermostable β-1,3-1,4-glucanase from Clostridium thermocellum improves the nutritive value of highly viscous barley-based diets for broilers. Br Poult Sci. 53(2):224–234.

Ridley BL, O'Neill MA, Mohnen D. 2001. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. Phytochemistry. 57(6):929–967.

Rincon MT, Cepeljnik T, Martin JC, Lamed R, Barak Y, Bayer EA, Harry J, Tadej C, Flint HJ. 2005. Unconventional Mode of Attachment of the Ruminococcus flavefaciens Cellulosome to the Cell Surface. J Bacteriol. 187(22):7569–7578.

Rose JKC, Lee S-J. 2010. Straying off the Highway: Trafficking of Secreted Plant Proteins and Complexity in the Plant Cell Wall Proteome. Plant Physiol. 153(2):433–436.

Rotter BA, Neskar M, Guenter W, Marquardt RR. 1989. Effect of enzyme supplementation on the nutritive value of hulless barley in chicken diets. Anim Feed Sci Technol. 24(3–4):233–245.

Round AN, Rigby NM, MacDougall AJ, Morris VJ. 2010. A new view of pectin structure revealed by acid hydrolysis and atomic force microscopy. Carbohydr Res. 345(4):487–497.

Rouvinen J, Bergfors T, Teeri T, Knowles J, Jones T. 1990. Three-dimensional structure of cellobiohydrolase II from Trichoderma reesei. Science (80-). 249(4967):380–386.

Ruiz-Dueiias FJ, Camarero S, Perez-Boada M, Martinez MJ, Martinez AT. 2001. A new versatile peroxidase. Biochem Soc Trans. 29(2):116–122.

Sabathé F, Bélaïch A, Soucaille P. 2002. Characterization of the cellulolytic complex (cellulosome) of Clostridium acetobutylicum. FEMS Microbiol Lett. 217(1):15–22.

Sahdev S, Khattar SK, Saini KS. 2008. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem. 307(1–2):249–64.

Salama-Alber O, Jobby MK, Chitayat S, Smith SP, White BA, Shimon LJW, Lamed R, Frolow F, Bayer EA. 2013. Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: Binding fidelity, promiscuity, and structural buttresses. J Biol Chem. 288(23):16827–16838.

San-Miguel T, Pérez-Bermúdez P, Gavidia I. 2013. Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature. Springerplus. 2(89):1–4.

De Santi C, Willassen NP, Williamson A. 2016. Biochemical characterization of a family 15 carbohydrate esterase from a bacterial marine Arctic metagenome. PLoS One. 11(7):1–22.

Santos A, San Mauro M, Díaz DM. 2006. Prebiotics and their long-term influence on the microbial populations of the mouse bowel. Food Microbiol. 23(5):498–503.

Sára M, Sleytr UB. 2000. S-Layer Proteins. J Bacteriol. 182(4):859–868.

SAS. 2004. SAS User´s Guide: Statistics. SAS Institute In Cary NC.

Saulnier L, Thibault J-F. 1999. Ferulic acid and diferulic acids as components of sugar-beet pectins and maize bran heteroxylans. J Sci Food Agric. 79(3):396–402.

Sauvonnet ١, Poquet I, Pugsley AP. 1995. Extracellular secretion of pullulanase is

unaffected by minor sequence changes but is usually prevented by adding reporter proteins to its N- or C-terminal end. J Bacteriol. 177(18):5238–5246.

Scantlebury-Manning T, Gibson GR. 2004. Prebiotics. Best Pract Res Clin Gastroenterol. 18(2):287–298.

Schein CH, Noteborn MHM. 1988. Formation of Soluble Recombinant Proteins in Escherichia Coli is Favored by Lower Growth Temperature. Nat Biotechnol. 6:291–294.

Scheller HV, Ulvskov P. 2010. Hemicelluloses. Annu Rev Plant Biol. 61:263–289.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, *et al*. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 75(23):7537–7541.

Schoenherr S, Ebrahimi M, Czermak P. 2018. Lignin Degradation Processes and the Purification of Valuable Products. In: Lignin - Trends and Applications. Vol. 2. Matheus Po. p. 64.

Schubot FD, Kataeva IA, Blum DL, Shah AK, Ljungdahl LG, Rose JP, Wang BC. 2001. Structural basis for the substrate specificity of the feruloyl esterase domain of the cellulosomal xylanase Z from Clostridium thermocellum. Biochemistry. 40(42):12524–12532.

Sequeira AF, Brás JLA, Fernandes VO, Guerreiro CIPD, Vincentelli R, Fontes CMGA. 2017. A Novel Platform for High-Throughput Gene Synthesis to Maximize Recombinant Expression in Escherichia coli. In: Domingues L, editor. Methods in Molecular Biology. Vol. 1620. Springer Science+Business Media. p. 113–12.

Sharma S, Bhat TK, Dawra RK. 2000. A spectrophotometric method for assay of tannase using rhodanine. Anal Biochem. 279(1):85–89.

Shrotri A, Kobayashi H, Fukuoka A. 2017. Catalytic Conversion of Structural Carbohydrates and Lignin to Chemicals. Adv Catal. 60:1–57.

Sigoillot JC, Berrin JG, Bey M, Lesage-Meessen L, Levasseur A, Lomascolo A, Record E, Uzan-Boukhris E. 2012. Fungal Strategies for Lignin Degradation. Adv Bot Res. 61:263–308.

Silva SSP, Smithard RR. 2002. Effect of enzyme supplementation of a rye-based diet on xylanase activity in the small intestine of broilers, on intestinal crypt cell proliferation and on nutrient digestibility and growth performance of the birds. Br Poult Sci. 43(2):274–282.

Simões J, Nunes FM, Domingues MR, Coimbra MA. 2011. Demonstration of the presence of acetylation and arabinose branching as structural features of locust bean gum galactomannans. Carbohydr Polym. 86(4):1476–1483.

Singh Arora D, Kumar Sharma R. 2010. Ligninolytic fungal laccases and their biotechnological applications. Appl Biochem Biotechnol. 160(6):1760–1788.

Sinnott ML. 1990. Catalytic mechanism of enzymic glycosyl transfer. Chem Rev. 90(7):1171–1202.

Smeets N, Nuyens F, Van Campenhout L, Niewold T. 2014. Variability in the in vitro degradation of non-starch polysaccharides from wheat by feed enzymes. Anim Feed Sci Technol. 187:110–114.

Smith SP, Bayer EA. 2013. Insights into cellulosome assembly and dynamics: From dissection to reconstruction of the supramolecular enzyme complex. Curr Opin Struct Biol. 23(5):686–694.

Smits CHM, Annison G. 1996. Non-starch plant polysaccharides in broiler nutrition - towards a physiologically valid approach to their determination. World Poult Sci. 52(2):203–221.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial Diversity in the Deep Sea and the Underexplored "Rare Biosphere." PNAS. 103(32):12115–12120.

Somerville C. 2006. Cellulose Synthesis in Higher Plants. Annu Rev Cell Dev Biol. 22(1):53–78.

Špániková S, Biely P. 2006. Glucuronoyl esterase - Novel carbohydrate esterase produced by Schizophyllum commune. FEBS Lett. 580(19):4597–4601.

Srivastav A, Kar R. 2009. Characterization And Application Of Tannase Produced By Aspergillus Niger ITCC 6514.07 On Pomegranate Rind. Brazilian J Microbiol. 540:782–789.

Ššković J, Kos B, Goreta J, Matošić S. 2001. Role of lactic acid bacteria and bifidobacteria in symbiotic effect. Food Technol Biotechnol. 39:227–235.

Stahl SW, Nash MA, Fried DB, Slutzki M, Barak Y, Bayer EA, Gaub HE. 2012. Single-molecule dissection of the high-affinity cohesin-dockerin complex. Proc Natl Acad Sci. 109(50):20431–20436.

Stam MR, Danchin EGJ, Rancurel C, Coutinho PM, Henrissat B. 2006. Dividing the large glycoside hydrolase family 13 into subfamilies: Towards improved functional annotations of α-amylase-related proteins. Protein Eng Des Sel. 19(12):555–562.

Sticklen MB. 2008. Plant genetic engineering for biofuel production: Towards affordable cellulosic ethanol. Nat Rev Genet. 9(6):433–443.

Stoll D, Stålbrand H, Warren RAJ. 1999. Mannan-Degrading Enzymes from Cellulomonas fimi. Appl Environ Microbiol. 65(6):2598–2605.

Sugano Y. 2009. DyP-type peroxidases comprise a novel heme peroxidase family. Cell Mol Life Sci. 66(8):1387–1403.

Sugiyama J, Vuong R, Chanzy H. 1991. Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. Macromolecules. 24(14):4168–4175.

Sunner H, Charavgi MD, Olsson L, Topakas E, Christakopoulos P. 2015. Glucuronoyl esterase screening and characterization assays utilizing commercially available benzyl glucuronic acid ester. Molecules. 20(10):17807–17817.

Sutherland IW. 1995. Polysaccharide lyases. FEMS Microbiol Rev. 16(4):323–347.

Świątkiewicz S, Koreleski J, Arczewska-Włosek A. 2011. Effect of inulin and oligofructose on performance and bone characteristics of broiler chickens fed on diets with different concentrations of calcium and phosphorus. Br Poult Sci. 52(4):483–491.

Takahashi N, Koshijima T. 1988. Ester linkages between lignin and glucuronoxylan in a lignin-carbohydrate complex from beech (Fagus crenata) wood. Wood Sci Technol. 22:231–241.

Talbott LD, Ray PM. 1992. Molecular size and separability features of pea cell wall polysaccharides: implications for models of primary wall structure. Plant Physiol. 98(1):357–368.

Tamaki Y, Teruya T, Tako M. 2010. The Chemical Structure of Galactomannan Isolated from Seeds of Delonix regia. Biosci Biotechnol Biochem. 74(5):1110–1112.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 30(12):2725–2729.

Taylor EJ, Gloster TM, Turkenburg JP, Vincent F, Brzozowski AM, Dupont C, Shareck F, Centeno MSJ, Prates JAM, Puchart V, *et al*. 2006. Structure and activity of two metal ion-dependent acetylxylan esterases involved in plant cell wall degradation reveals a close similarity to peptidoglycan deacetylases. J Biol Chem. 281(16):10968–10975.

Taylor EJ, Goyal A, Guerreiro CIPD, Prates AM, Money VA, Ferry N, Morland C, Planas A, Macdonald JA, Stick R V, *et al*. 2005. How Family 26 Glycoside Hydrolases Orchestrate Catalysis on Different Polysaccharides. J Biol Chem. 280(38):32761–32767.

Teitelbaum JE, Walker A. 2002. Nutritional impact of pre- and probiotics as protective gastrointestinal organisms. Annu Rev Nutr. 22:107–138.

Teleman A, Nordström M, Tenkanen M, Jacobs A, Dahlman O. 2003. Isolation and characterization of O-acetylated glucomannans from aspen and birch wood. Carbohydr Res. 338(6):525–534.

Teng P-Y, Kim WK. 2018. Review: Roles of Prebiotics in Intestinal Ecosystem of Broilers. Front Vet Sci. 5:1–18.

Thibault JF, Renard CMGC, Axelos MAV, Roger P, Crépeau MJ. 1993. Studies of the length of homogalacturonic regions in pectins by acid hydrolysis. Carbohydr Res. 238:271–286.

Thurston CF. 1994. The structure and function of fungal laccases. Microbiology. 140(1 994):19–26.

Topakas E, Moukouli M, Dimarogona M, Vafiadi C, Christakopoulos P. 2010. Functional expression of a thermophilic glucuronoyl esterase from Sporotrichum thermophile: Identification of the nucleophilic serine. Appl Microbiol Biotechnol. 87(5):1765–1772.

Topakas E, Vafiadi C, Christakopoulos P. 2007. Microbial production, characterization and applications of feruloyl esterases. Process Biochem. 42(4):497–509.

Tsai AYL, Canam T, Gorzsás A, Mellerowicz EJ, Campbell MM, Master ER. 2012. Constitutive expression of a fungal glucuronoyl esterase in Arabidopsis reveals altered cell wall composition and structure. Plant Biotechnol J. 10(9):1077–1087.

Turchetto J, Sequeira AF, Ramond L, Peysson F, Brás JLA, Saez NJ, Duhoo Y, Blémont M, Guerreiro CIPD, Quinton L, *et al*. 2017. High-throughput expression of animal venom toxins in Escherichia coli to generate a large library of oxidized disulphide-reticulated peptides for drug discovery. Microb Cell Fact. 16(1):1–15.

Udatha DBRKG, Kouskoumvekaki I, Lsson L, Panagiotou G. 2011. The interplay of descriptor-based computational analysis with pharmacophore modeling builds the basis for a novel classification scheme for feruloyl esterases. Biotechnol Adv. 29(1):94–110.

Vahouny G V., Tombes R, Cassidy MM, Kritchevsky D, Gallo LL. 1980. Dietary fibers: V. Binding of bile salts, phospholipids and cholesterol from mixed micelles by bile acid sequestrants and dietary fibers. Lipids. 15(12):1012–1018.

Valbuena A, Oroza J, Hervás R, Vera AM, Rodríguez D, Menéndez M, Sulkowska JI, Marek Cieplak MC-V. 2009. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. PNAS. 106(33):13791–13796.

Vaz-Moreira I, Egas C, Nunes OC, Manaia CM. 2013. Bacterial diversity from the source to the tap: A comparative study based on 16S rRNA gene-DGGE and culture-dependent methods. FEMS Microbiol Ecol. 83(2):361–374.

Vera A, González-Montalbán N, Arís A, Villaverde A. 2007. The Conformational Quality of Insoluble Recombinant Proteins Is Enhanced at Low Growth Temperatures. Biotechnol Bioeng. 96(6):1101–1106.

Villamide MJ, Fuente JM, Ayala PP, Flores A. 1997. Energy Evaluation of Eight Barley Cultivars for Poultry : Effect of Dietary Enzyme Addition. Poult Sci. 76(6):834–840.

Vincken J-P, Schols HA, Oomen RJFJ, McCann MC, Ulvskov P, Voragen AGJ, Visser RGF. 2003. If Homogalacturonan Were a Side Chain of Rhamnogalacturonan I. Implications for Cell Wall Architecture. Plant Physiol. 132(4):1781–1789.

Visser J, Voragen AGJ. 1996. Pectins and Pectinases, Volume 14. In: Progress in Biotechnology. Elsevier Science.

Viveros A, Brenes A, Pizarro M, Castaño M. 1994. Effect of enzyme supplementation of a diet based on barley, and autoclave treatment, on apparent digestibility, growth performance and gut morphology of broilers. Anim Feed Sci Technol. 48(3–4):237–251.

Voulgaridou GP, Mantso T, Chlichlia K, Panayiotidis MI, Pappa A. 2013. Efficient E. coli Expression Strategies for Production of Soluble Human Crystallin ALDH3A1. PLoS One. 8(2):15–17.

Vries RP de, Visser J. 2001. Aspergillus Enzymes Involved in Degradation of Plant Cell Wall Polysaccharides. Microbiol Mol Biol Rev. 65(4):497–522.

Vrsanska M, Buresova A, Damborsky P, Adam V. 2015. Influence of Different Inducers on Ligninolytic Enzyme Activities. J Met Nanotechnologies. 3:64–70.

Wagner DD, Thomas OP. 1978. Influence of diets containing rye or pectin on the intestinal flora of chicks. Poult Sci. 57(4):971–975.

Wang Q, Withers SG, Graham RW, Trimbur D, Warren RAJ. 1994. Changing Enzymic Reaction Mechanisms by Mutagenesis: Conversion of a Retaining Glucosidase to an Inverting Enzyme. J Am Chem Soc. 116(25):11594–11595.

Watanabe T. 1995. Important Properties of Lignin-Carbohydrates Complexes(LCCs) in Environmentally Safe Paper Making. Trends Glycosci Glycotechnol. 7(33):57–68.

Watanabe T, Koshijima T. 1988. Evidence for an ester linkage between lignin and glucuronic acid in lignin-carbohydrate complexes by DDQ-oxidation. Agric Biol Chem. 52(11):2953–2955.

Wei H, Xu Q, Taylor LE, Baker JO, Tucker MP, Ding S-Y. 2009. Natural paradigms of plant cell wall degradation. Curr Opin Biotechnol. 20(3):330–338.

Wende G, Fry SC. 1997. O-feruloylated, O-acetylated oligosaccharides as side-chains of grass xylans. Phytochemistry. 44(6):1011–1018.

Wilkinson DL, Harrison RG. 1991. Predicting the Solubility of Heterologous Proteins Expressed in E. coli. Biotechnol (N Y). 9:443–448.

Williams P, Geraert P, Uzu G, Annison G. 1997. Factors affecting non-starch polysaccharide digestibility in poultry. Cah Options Mediterranéennes. 26:125–134.

Williams S. 2017. Glycosyltransferases. [accessed 2019 Feb 4]. https://www.cazypedia.org/index.php/Glycosyltransferases.

Williamson G, Kroon PA, Faulds CB. 1998. Hairy plant polysaccharides: a close shave with microbial esterases Gary. Microbiology. 144(Pt 8):2011–2023.

Wilson DB. 2008. Three Microbial Strategies for Plant Cell Wall Degradation. Ann NY Acad Sci. 1125:289–297.

Withers SG. 1995. Probing of glycosidase active sites through labeling, mutagenesis and kinetic studies. Prog Biotechnol. 10(C):97–111.

Withers SG, Williams S. 2017. Glycoside hydrolases. CAZypedia. [accessed 2019 Mar 5]. https://www.cazypedia.org/index.php/Glycoside_hydrolases.

Wong DWS. 2006. Feruloyl esterase. Appl Biochem Biotechnol. 133:87–112.

Wong DWS. 2009. Structure and Action Mechanism of Ligninolytic Enzymes. Appl Biochem Biotechnol. 157(2):174–209.

Wong DWS, Chan VJ, McCormack AA, Hirsch J, Biely P. 2012. Functional Cloning and Expression of the Schizophyllum commune Glucuronoyl Esterase Gene and Characterization of the Recombinant Enzyme. Biotechnol Res Int. 2012:1–7.

Wong JMW, Souza R, Kendall CWC, Emam A, Jenkins DJA. 2006. Colonic Health: Fermentation and Short Chain Fatty Acids. J Clin Gastroenterol. 40(3):235–243.

Xia J, Li P. 2018. Pectic enzymes. Ref Modul Food Sci.:1–7.

Xiao R, Power RF, Mallonee D, Routt K, Spangler L, Pescatore AJ, Cantor AH, Ao T, Pierce JL, Dawson KA. 2012. Effects of yeast cell wall-derived mannan-oligosaccharides on jejunal gene expression in young broiler chickens. Poult Sci. 91(7):1660–1669.

Xu Q, Barak Y, Kenig R, Shoham Y, Bayer EA, Lamed R. 2004. A Novel Acetivibrio cellulolyticus Anchoring Scaffoldin That Bears Divergent Cohesins. J Bacteriol. 186(17):5782–5789.

Xu Q, Gao W, Ding S-Y, Kenig R, Shoham Y, Bayer EA, Lamed R. 2003. The cellulosome system of Acetivibrio cellulolyticus includes a novel type of adaptor protein and a cell surface anchoring protein. J Bacteriol. 185(15):4548–4557.

Xu Q, Resch MG, Podkaminer K, Yang S, Baker JO, Donohoe BS, Wilson C, Klingeman DM, Olson DG, Decker SR, *et al*. 2016. Cell Biology: Dramatic performance of Clostridium thermocellum explained by its wide range of cellulase modalities. Sci Adv. 2(2):1–12.

Xue F, Guo H, Hu Y, Liu R, Huang L, Lv H, Liu C, Yang M, Ma L. 2016. Expression of Codon-Optimized Plant Glycosyltransferase UGT72B14 in Escherichia coli Enhances Salidroside Production . Biomed Res Int. 2016:1–8.

Xue GP, Patel M, Johnson JS, Smyth DJ, Vickers CE. 2003. Selectable marker-free transgenic barley producing a high level of cellulase (1,4-beta-glucanase) in developing grains. Plant Cell Rep. 21(11):1088–1094.

Yamada H, Itoh K, Morishita Y, Taniguchi H. 1993. Structure and properties of oligosaccharides from wheat bran. Cereal Foods World. 38:490–492.

Yang Y, Iji PA, Choct M. 2009. Dietary modulation of gut microflora in broiler chickens: A review of the role of six kinds of alternatives to in-feed antibiotics. Worlds Poult Sci J. 65(1):97–114.

Yang Y, Iji PA, Kocher A, Mikkelsen LL, Choct M. 2008. Effects of mannanoligosaccharide and fructooligosaccharide on the response of broilers to pathogenic Escherichia coli challenge. Br Poult Sci. 49(5):550–559.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN : a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40(WS):445–451.

Yip V, Varrot A, Davies G, Rajan S, Yang X, Thompson J, Anderson W, Withers S. 2004. An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from Thermotoga maritima. J Am Chem Soc. 126(27):8354–8355.

Yip V, Withers SG. 2006. Breakdown of oligosaccharides by the process of elimination. Curr Opin Chem Biol. 10(2):147–155.

Zandleven J, Sørensen S, Harholt J, Beldman G, Schols H, Scheller H, Voragen A. 2007. Xylogalacturonan exists in cell walls from various tissues of Arabidopsis thaliana. Phytochemistry. 68(8):1219–1226.

Zechel DL, Withers SG. 2000. Glycosidase mechanisms: Anatomy of a finely tuned catalyst. Acc Chem Res. 33(1):11–18.

Zhang R, Yip VLY, Withers SG, Columbia B. 2010. Mechanisms of enzymatic glycosyl transfer. In: Comprehensive Natural Products II Chemistry and Biology. Canada: Elsevier Ltd. p. 385–422.

Zhang Y, Olsen DR, Nguyen KB, Olson PS, Rhodes ET, Mascarenhas D. 1998. Expression of Eukaryotic Proteins in Soluble Form in Escherichia coli. Protein Expr Purif. 12(2):159–165.

Zhao M, Wu F, Xu P. 2015. Development of a rapid high-efficiency scalable process for acetylated Sus scrofa cationic trypsin production from Escherichia coli inclusion bodies. Protein Expr Purif. 116:120–126.

Zhenping S, Wenting L, Ruikui Y, Jia L, Honghong L, Wei S, Zhongmie W, Jingpan L, Zhe S, Yuling Q. 2013. Effect of a straw-derived xylooligosaccharide on broiler growth performance, endocrine metabolism, and immune response. Can J Vet Res. 77(2):105–109.

Zhou X, Broadbelt LJ, Vinu R. 2016. Mechanistic Understanding of Thermochemical Conversion of Polymers and Lignocellulosic Biomass. Adv Chem Eng. 49:95–198.

Ziegelhoffer T, Raasch JA, Austin-Phillips S. 2001. Dramatic effects of truncation and sub-cellular targeting on the accumulation of recombinant microbial cellulase in tobacco. Mol Breed. 8(2):147–158.

Ziegler MT, Thomas SR, Danna KJ. 2000. Accumulation of a thermostable endo-1,4-β-D-glucanase in the apoplast of Arabidopsis thaliana leaves. Mol Breed. 6(1):37–46.

Zverlov V V., Klupp M, Krauss J, Schwarz WH. 2008. Mutations in the scaffoldin gene, cipA, of Clostridium thermocellum with impaired cellulosome formation and cellulose hydrolysis: Insertions of a new transposable element, IS1447, and implications for cellulase synergism on crystalline cellulose. J Bacteriol. 190(12):4321–4327.

ANNEXES

Supplementary Data – Chapter 2

Table S2.1. Relative weight and length of the gastrointestinal tract and viscosity of digesta. Samples of broilers fed on a barley-based diet NC (non-supplemented with exogenous polypeptides) or PC (supplemented with a commercial enzyme mixture), *Ct*Glc16A (with the recombinant 1,3-1,4-β-glucanase) or *Ct*Cel5E (with the recombinant 1,4-β-glucanase).

	NC	PC	*Ct*Glc16A	*Ct*Cel8A	SEM	P-value
Relative weight (g/kg BW)						
Crop	3,8	2,9	3,5	3,0	0,30	0,118
Gizzard	13,5	14,2	15,7	14,0	1,25	0,578
Liver	27,4	29,5	28,8	29,4	1,17	0,543
Duodenum	9,5	8,3	8,3	9,1	0,40	0,092
Jejunum	$17,2^{a}$	$14,3^{b}$	$15,7^{ab}$	$15,0^{ab}$	0,72	0,042
Ileum	$14,4^{a}$	$10,5^{b}$	$12,8^{ab}$	$12,7^{ab}$	0,80	0,012
Caecum	$3,0^{ab}$	$2,9^{ab}$	$3,5^{a}$	$2,6^{b}$	0,22	0,030
Relative length (cm/kg BW)						
Duodenum	25,4	23,1	22,5	24,4	1,11	0,245
Jejunum	64,0	57,30	57,60	58,9	2,15	0,105
Ileum	$66,8^{a}$	$56,6^{b}$	$59,6^{ab}$	$61,4^{ab}$	2,42	0,031
Caecum	15,0	12,9	13,5	14,1	0,63	0,118
Content Viscosity (cP)						
Duodenum+ jejunum	$11,6^{a}$	$8,0^{bc}$	$6,8^{c}$	$9,9^{c}$	0,73	0,001
Ileum	$20,8^{a}$	$10,5^{b}$	$11,0^{b}$	$17,5^{a}$	1,46	0,001

Means in the same line with different letter superscripts (a, b, c) are significantly different (P<0.05) or tend to be significantly different (P<0.1).

Supplementary Data – Chapter 4

Table S4.1. Relative weight and length of the GI tract of broilers (Experiment 1).
Diet supplemented with a 1,4-β-xylanase mixture (XYL), a xylo-oligosaccharide preparation (XOS) and xylose (OSE). A fourth group of birds was fed on a basal non-supplemented diet (C-).

	C-	XYL	XOS	OSE	SEM	p(*F*)
Relative Weight (g/100 g BW)						
Crop	2,21	2,46	2,36	2,56	0,259	0,803
Gizzard	7,14	7,13	6,66	7,08	0,398	0,804
Liver	20,6	20,70	20,90	20,30	1,088	0,985
Duodenum	5,37	6,06	5,85	5,36	0,292	0,245
Jejunum	11,80	11,30	11,50	11,30	0,642	0,953
Ileum	9,55	9,21	8,53	8,92	0,563	0,627
Cecum	1,80	1,97	1,97	1,88	0,171	0,872
Relative Length (cm/kg BW)						
Duodenum	17,9	17,5	17,9	17,8	0,552	0,954
Jejunum	46,1	48,4	43,7	47,9	1,623	0,191
Ileum	47,5	48,2	45,7	49,1	1,734	0,568
Cecum	10,5	11,2	10,6	10,8	0,382	0,652
Content Viscosity (cP)						
Duodenum+ Jejunum	3,90	4,02	3,87	3,77	0,199	0,832
Ileum	6,70	7,47	8,91	6,93	1,329	0,598

Table S4.2. Relative weight and length of the GI tract of broilers (Experiment 3).
Diet supplemented with a XOS preparation provided at three different incorporation rates, 0.1 g/kg (XOS1), 1 g/kg (XOS2) or 10 g/kg (XOS3). A fourth group of birds was fed on a basal non-supplemented diet (C-).

	C-	XOS1	XOS2	XOS3	SEM	p(*F*)
Relative Weight (g/100 g BW)						
Crop	2,84	2,19	2,46	2,43	0,192	0,141
Gizzard	16,15	13,74	14,93	15,29	0,822	0,233
Liver	22,20	21,45	25,17	23,42	1,183	0,147
Duodenum	5,67	5,32	5,34	5,59	0,367	0,879
Jejunum	12,04	10,76	10,87	11,64	0,832	0,651
Ileum	11,13	10,44	10,05	10,31	0,844	0,824
Cecum	1,88	1,95	2,01	1,96	0,119	0,885
Relative Length (cm/kg BW)						
Duodenum	11,90	11,14	11,17	12,02	0,632	0,658
Jejunum	31,93	28,88	30,78	31,31	1,434	0,479
Ileum	36,22	33,17	33,07	32,64	2,363	0,695
Cecum	8,60	7,80	8,10	8,57	0,440	0,518
Content Viscosity (cP)						
Duodenum+ Jejunum	3,20	3,52	3,99	4,00	0,345	0,221
Ileum	4,60	5,20	5,00	5,40	0,513	0,743

Supplementary Data – Chapter 5

Table S5.1. DNA and amino acid sequences, predicted molecular mass, molecular architecture, CAZy family and origin of enzymes produced in this work.

(Data available in a supplementary excel file)

Supplementary Data – Chapter 6

Table S6.1. Primers used for the PCR amplification of CE15 constructs.

Enzyme	Primer	5'-3' sequence
***Bc*GE15A**	*Bc*GE15A_F	TCAGCAAGGGCTGAGGGTGTTGCCGGGTATCAACGAAC
	*Bc*GE15A_R	TCAGCGGAAGCTGAGGTTAGTGGAATAACCAGCGCTC
***Cc*GE15A**	*Cc*GE15A_F	TCAGCAAGGGCTGAGGTCAGATCTGGTTTATGGAGACCTG
	*Cc*GE15A_R	TCAGCGGAAGCTGAGGTCAAAATGTCGGAACAGTCCAATC
***Cj*GE15A**	*Cj*GE15A_F	TCAGCAAGGGCTGAGGTCTGTTCCTCTTTCCGGTGTGGAG
	*Cj*GE15A_R	TCAGCGGAAGCTGAGGTCAATGTAAACCGGGAGTTTCCCAATC
***Cj*GE15B**	*Cj*GE15B_F	TCAGCAAGGGCTGAGGCAACTTTCTGATCCCCTGTTG
	*Cj*GE15B_R	TCAGCGGAAGCTGAGGTTATTGGAGAGTGGGGGTAGTCCAG
***Ck*GE15A**	*Ck*GE15A_F	TCAGCAAGGGCTGAGGATAGATAATGCTTTGGCTGCTC
	*Ck*GE15A_R	TCAGCGGAAGCTGAGGTTATATCTGTTTCAAATCAGC
***Ma*GE15A**	*Ma*GE15A_F	TCAGCAAGGGCTGAGGGATGCCAGCAAACTACCCACTATTAC
	*Ma*GE15A_R	TCAGCGGAAGCTGAGGTCAATCGAGCGTTGGAGTACTCCAATC
***Mr*GE15A**	*Mr*GE15A_F	TCAGCAAGGGCTGAGGCCGTTGCCGGAATTCAGTCAACTG
	*Mr*GE15A_R	TCAGCGGAAGCTGAGGTCATTTGAGAAGTAAAAGTTTAGTAG
***Ot*GE15B**	*Ot*GE15B_F	TCAGCAAGGGCTGAGGTCGGGTCGTCAGGACGGCTC
	*Ot*GE15B_R	TCAGCGGAAGCTGAGGTTAGGGGAAAAACGGCAGCC
***Rf*GE15A**	*Rf*GE15A_F	TCAGCAAGGGCTGAGGGATAGCGGTCCAGAGCTTATTTAC
	*Rf*GE15A_R	TCAGCGGAAGCTGAGGTCAGTAAAGCCACTTTGAAGCAAATG
***Tt*GE15A**	*Tt*GE15A_F	TCAGCAAGGGCTGAGGGCTCCCTTAACCGCAGATCAGGATC
	*Tt*GE15A_R	TCAGCGGAAGCTGAGGTCACTTGAGCCTGTTCGGGCTTG
***Tt*GE15B**	*Tt*GE15B_F	TCAGCAAGGGCTGAGGTCCAACGAAAACCCCAGCGCCAG
	*Tt*GE15B_R	TCAGCGGAAGCTGAGGTTAGAAGTCCAGGTTGCCGGAC
***Tt*GE15C**	*Tt*GE15C_F	TCAGCAAGGGCTGAGGGAAAATATGGGGTCTGACTGTGATG
	*Tt*GE15C_R	TCAGCGGAAGCTGAGGTCAATCCGCCAGCGTCGGGGTCG
***Zg*GE15A**	*Zg*GE15A_F	TCAGCAAGGGCTGAGGGCACAACCCGAAGCAAATTATGAC
	*Zg*GE15A_R	TCAGCGGAAGCTGAGGTCATTTTACAAAATTTTGCGCCCAAC
***Zp*GE15A**	*Zp*GE15A_F	TCAGCAAGGGCTGAGGCAGGATGGAAAATCTTCTCAGG
	*Zp*GE15A_R	TCAGCGGAAGCTGAGGTTAAAATTGCCAATATTTGCTTG

Table S6.2. Predicted molecular mass, molecular architecture and origin of GEs studied in this work.

Enzyme	GS/PCR	Origin	Organism	Accession number	Modularity	Molecular Architecture	Predicted Molecular Mass	Protein Expression	Yield (mg/L)	Specific activity (U/mg)
***Pc*GE15A**	GS	Eukarya	*Phanerochaete carnosa*	AFM93784.1	single domain	[(1-17)SIGN][(18-406)CE15]	42,771	0	-	-
***Tt*GE15D**	GS	Eukarya	*Thermothelomyces thermophilus*	AEO60464.1	single domain	[(1-18)SIGN][(19-397)CE15]	42,207	0	-	-
***Tr*GE15A**	GS	Eukarya	*Trichoderma reesei*	AAP57749.1	modular	[(1-17)SIGN][(18-53)CBM1][(98-460)CE15]	40,733	0	-	-
***Sc*GE15A**	GS	Eukarya	*Schizophyllum commune*	EFI91386.1, XP_003026289.1	single domain	[(1-19)SIGN][(20-393)CE15]	41,593	0	-	-
***Pa*GE15A**	GS	Eukarya	*Podospora anserina*	CAP60908.1, CDP24923.1, XP_001903136.1	modular	[(1-21)SIGN][(23-59)CBM1][(110-481)CE15]	41,910	0	-	-
***Bc*GE15A**	PCR	Bacteria	*Bacteroides cellulosilyticus*	CDB71655.1	single domain	[(1-436)CE15]	42,246	0	-	-
***Ck*GE15A**	PCR	Bacteria	*Caldicellulosiruptor kristjanssonii*	ADQ41707.1	modular	[(1-33)SIGN][(38-174)CBM4][(192-330)CBM4][(361-695)GH10][(714-872)CBM9][(885-1065)CBM9][(1082-1255)CBM9][(1310-1721)CE15]-UNK-[(1985-2048)SLH][(2102-2159)SLH]	47,969	+	8,48	1,61
***Cj*GE15A**	PCR	Bacteria	*Cellvibrio japonicus*	ACE84474.1	single domain	[(1-28)SIGN][(29-477)CE15]	43,900	+	15,21	3,50
***Cc*GE15A**	PCR	Bacteria	*Clostridium cellobioparum*	WP_027630575.1	single domain	[(1-26)SIGN][(27-484)CE15]	50,646	+++	163,71	7,08

***Cj*GE15B**	PCR	Bacteria	*Cellvibrio japonicus*	ACE84239.1	modular	[(1-34)SIGN][(35-115)CBM60][(169-198)CBM10][(246-638)CE15]	43,922	0	-	-
***Ma*GE15A**	PCR	Bacteria	*Marinimicrobium agarilyticum*	WP_027329784.1	modular	[(1-17)SIGN][(52-190)CBM57][(234-601)CE15]	41,279	0	-	-
***Mr*GE15A**	PCR	Bacteria	*Melioribacter roseus*	AFN75319.1	single domain	[(1-41)SIGN][(42-506)CE15]	54,228	0	-	-
***Ot*GE15B**	PCR	Bacteria	*Opitutus terrae*	ACB75551.1	modular	[(1-23)SIGN][(38-447)CE15]-UNK	49,576	+++	69,81	3,97
***Pr*GE15A**	GS	Bacteria	*Prevotella ruminicola*	ADE83410.1	single domain	[(1-18)SIGN][(19-394)CE15]	42,358	+++	195,50	14,51
***Rf*GE15A**	PCR	Bacteria	*Ruminococcus flavefaciens*	CAB55348.1	modular	[(1-40)SIGN][(41-264)CE3][(282-768)CE15]	56,312	++	64,90	94,30
***Tt*GE15C**	PCR	Bacteria	*Teredinibacter turnerae*	ACR11863.1	modular	[(1-17)SIGN][(60-199)CBM57][(229-608)CE15]	42,654	+	29,25	4,94
***Tt*GE15A**	PCR	Bacteria	*Teredinibacter turnerae*	ACR12575.1	single domain	[(1-22)SIGN][(23-451)CE15]	48,396	+++	126,91	105,82
***Tt*GE15B**	PCR	Bacteria	*Teredinibacter turnerae*	ACR14261.1	modular	[(13-198)GH11]-CBM60-[(386-447)CBM5][(513-644)CBM57][(668-1051)CE15]	42,358	++	28,10	11,36
***Zg*GE15A**	PCR	Bacteria	*Zobellia galactanivorans*	CAZ95028.1	single domain	[(1-402)CE15]	47,250	+	110,29	8,34
***Zp*GE15A**	PCR	Bacteria	*Zunongwangia profunda*	ADF53344.1	single domain	[(1-24)SIGN][(25-457)CE15]	50,235	++	32,28	3,04

Table S6.3. DNA and amino acid sequences of GEs studied in this work.

Enzyme	Recombinant DNA Sequence	Recombinant Amino Acid Sequence
***Pc*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGACCCCG	MGSSHHHHHHSSGPQQGLR
	GCTAATATCCCGTTCAATGATGACAAACTGCCGGACCCGTTCCTGTTCAATGATGGCACCCCGGT	TPANIPFNDDKLPDPFLFNDG
	TCGTTCTCTGACCGATTGGTCATGCCGTCGCCAGCAACTGGCCTCGCTGATTCAGGGCTATGAA	TPVRSLTDWSCRRQQLASLIQ
	GCCGGTACGCTGCCGCCGAAACCGCCGATCGTGACGTCAACCTTTTCGCAAAACGGCCTGACCG	GYEAGTLPPKPPIVTSTFSQN
	GTAATCTGACGGTTACCGCCGGCTTCCCGGGTAACACCACGACCTTTAGCTCTCCGGTCACCTTC	GLTGNLTVTAGFPGNTTTFSS
	CCGAATGGTACGGTGCCGACCGAAGGTTGGCCGCTGCTGATTGCGTATAGCGGCCTGTCTATTC	PVTFPNGTVPTEGWPLLIAYS
	CGATCCCGGATGGTATTGCTGTGCTGACGTACGACAACAGTGCGATCGGCGAACAGAATGATCA	GLSIPIPDGIAVLTYDNSAIGEQ
	AACCAGCCGTGGCGTGGGCCAGTTTTTCGACGTTTACGGCCATAACGCGACCGCCAGTGCAATG	NDQTSRGVGQFFDVYGHNAT
	TCCGCGTGGGTTTGGGGTGTCAGTCGCATTATCGATGTGCTGGAAGTTACCCCGGCGGCCCACG	ASAMSAWVWGVSRIIDVLEVT
	TTAATACGGCCAAAATTGCAGTCACCGGTTGCAGCCGTGACGGTAAAGGTGCACTGATGGCTGG	PAAHVNTAKIAVTGCSRDGKG
	TGCGTTTGAAGAACGCATTGCACTGACGATCCCGCAAGAAAGTGGCTCCGGCGGTGATACCTGT	ALMAGAFEERIALTIPQESGS
	TGGCGTCTGAGCAAATTTGAACAGGATTCTGGTGACGTGGTTCAGCAAGCCACGGAAATTGTCCA	GGDTCWRLSKFEQDSGDVV
	AGAAAACGTGTGGTTTAGCACCAACTTCGATAACTTCGTTTTCAACATCAGCGTCCTGCCGTATGA	QQATEIVQENVWFSTNFDNFV
	CCATCACTCTCTGGCTGGTCTGATTGCGCCGCGCCCGATGATCTCTTACGAAAACACGGATTTTG	FNISVLPYDHHSLAGLIAPRPM
	AATGGCTGTCACCGCTGTCGGGCTTCGGTTGCATGACCGCAGCTCATCCGATTTGGGAAGCGAT	ISYENTDFEWLSPLSGFGCMT
	GGGCGTGCCGGATAACCACGGCTTTGTCCAAGTGGGTAATCATAGTCACTGTGAATTTCCGTCCG	AAHPIWEAMGVPDNHGFVQV
	ATCTGAATCCGACCCTGTTTGCGTTTTTCGACAAATTCCTGCTGGGTAAAGAAGCAAACACGACC	GNHSHCEFPSDLNPTLFAFFD
	ATCTTTGAAACCAATGAAGTTTTCAATGGCACCGTTTGGAACCCGAGTCAATGGATCAACTGGAC	KFLLGKEANTTIFETNEVFNGT
	CACCCCGACCCTGTCACAC	VWNPSQWINWTTPTLSH

***Tt*GE15D**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGTCGGAC	MGSSHHHHHHSSGPQQGLRS
	AACTACCCGACCGTCAACTCTGCGAAACTGCCGGACCCGTTCACGACCGCTTCAGGCGAAAAA	DNYPTVNSAKLPDPFTTASGEK
	GTCACCACGAAAGACCAATTTGAATGCCGTCGCGCGGAAATTAACAAAATCCTGCAGCAATATG	VTTKDQFECRRAEINKILQQYEL
	AACTGGGCGAATACCCGGGTCCGCCGGATTCTGTGGAAGCAAGTCTGTCCGGTAATAGTATTA	GEYPGPPDSVEASLSGNSITVR
	CCGTGCGTGTTACGGTCGGCTCAAAATCGATTAGCTTTTCTGCTAGTATCCGCAAACCGTCTGG	VTVGSKSISFSASIRKPSGAGPF
	CGCAGGTCCGTTCCCGGCTATTATCGGTATTGGCGGTGCATCGATTCCGATCCCGAGCAACGT	PAIIGIGGASIPIPSNVATITFNND
	TGCTACCATCACGTTTAACAATGATGAATTTGGTGCACAGATGGGCTCCGGTTCACGTGGTCAG	EFGAQMGSGSRGQGKFYDLFG
	GGTAAATTTTATGACCTGTTTGGTCGCGACCATTCTGCAGGCAGTCTGACCGCATGGGCTTGGG	RDHSAGSLTAWAWGVDRLIDG
	GTGTGGATCGTCTGATTGACGGCCTGGAACAGGTTGGTGCGCAAGCCTCGGGCATCGACACCA	LEQVGAQASGIDTKRLGVTGCS
	AACGTCTGGGCGTCACGGGTTGCAGCCGTAACGGTAAGGGTGCGTTTATTACCGGTGCCCTGG	RNGKGAFITGALVDRIALTIPQE
	TTGATCGTATTGCGCTGACGATCCCGCAGGAATCCGGTGCCGGCGGTGCAGCCTGTTGGCGCA	SGAGGAACWRISDQQKAAGAN
	TTTCAGACCAGCAAAAAGCAGCTGGCGCGAATATCCAGACCGCGGCCCAAATTATCACGGAAA	IQTAAQIITENPWFSRNFDPHVN
	ATCCGTGGTTTAGCCGCAACTTCGATCCGCATGTGAACAGCATCACCTCAGTTCCGCAGGACCA	SITSVPQDHHLLAALIVPRGLAV
	TCACCTGCTGGCAGCTCTGATTGTCCCGCGTGGTCTGGCGGTGTTTGAAAACAATATCGATTGG	FENNIDWLGPVSTTGCMAAGR
	CTGGGTCCTGTGAGCACCACGGGTTGCATGGCAGCAGGCCGCCTGATTTATAAAGCCTACGGT	LIYKAYGVPNNMGFSLVGGHNH
	GTTCCGAACAATATGGGCTTTTCCCTGGTCGGCGGTCATAATCACTGTCAGTTCCCGAGCTCTC	CQFPSSQNQDLNSYINYFLLGQ
	AGAACCAAGATCTGAACAGCTATATCAACTACTTCCTGCTGGGTCAGGGTTCGCCGAGCGGTGT	GSPSGVEHSDVNVNVAEWAP
	GGAACACAGTGACGTGAACGTTAATGTCGCGGAATGGGCCCCGTGGGGCGCAGGTGCTCCGA	WGAGAPTLALEQKLISEEDLNS
	CCCTGGCTCTGGAACAAAAACTGATCTCCGAAGAAGACCTGAATAGTGCTGTGGAC	AVD
***Tr*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGGCTCTG	MGSSHHHHHHSSGPQQGLRAL
	CCGGGCTCGATTACCCTGCGTTCCAATGCGAAACTGAATGACCTGTTTACGATGTTCAATGGCG	PGSITLRSNAKLNDLFTMFNGD
	ACAAAGTGACCACCAAAGATAAATTCAGCTGCCGTCAAGCGGAAATGTCTGAACTGATTCAGCG	KVTTKDKFSCRQAEMSELIQRY
	TTATGAACTGGGTACCCTGCCGGGTCGTCCGTCCACCCTGACGGCCTCATTCTCGGGTAACAC	ELGTLPGRPSTLTASFSGNTLTI

	CCTGACGATCAATTGTGGTGAAGCAGGCAAAAGCATTTCTTTTACCGTGACGATCACCTACCCG	NCGEAGKSISFTVTITYPSSGTA
	AGCTCTGGTACCGCACCGTATCCGGCTATTATCGGCTACGGCGGTGGCAGCCTGCCGGCACC	PYPAIIGYGGGSLPAPAGVAMIN
	GGCTGGTGTCGCCATGATTAACTTCAACAATGATAATATCGCAGCACAAGTGAACACGGGCTCT	FNNDNIAAQVNTGSRGQGKFY
	CGTGGTCAGGGCAAATTTTATGACCTGTACGGTAGTTCCCATAGTGCAGGTGCAATGACCGCAT	DLYGSSHSAGAMTAWAWGVS
	GGGCTTGGGGTGTCTCCCGTGTGATTGATGCACTGGAACTGGTGCCGGGTGCACGTATTGACA	RVIDALELVPGARIDTTKIGVTG
	CCACGAAAATCGGTGTTACGGGCTGCAGTCGTAACGGTAAAGGCGCAATGGTTGCGGGTGCCT	CSRNGKGAMVAGAFEKRIVLTL
	TTGAAAAACGCATCGTCCTGACCCTGCCGCAGGAAAGCGGTGCGGGTGGCTCGGCATGTTGG	PQESGAGGSACWRISDYLKSQ
	CGTATTAGCGATTATCTGAAATCTCAGGGTGCAAATATCCAAACCGCTAGCGAAATTATCGGCG	GANIQTASEIIGEDPWFSTTFNS
	AAGACCCGTGGTTTAGTACCACGTTCAACTCCTACGTTAATCAGGTGCCGGTTCTGCCGTTTGA	YVNQVPVLPFDHHSLAALIAPR
	TCATCACTCACTGGCAGCTCTGATTGCCCCGCGCGGTCTGTTCGTCATTGATAACAATATCGAC	GLFVIDNNIDWLGPQSCFGCMT
	TGGCTGGGTCCGCAGTCGTGCTTTGGCTGTATGACGGCAGCACATATGGCATGGCAAGCACTG	AAHMAWQALGVSDHMGYSQIG
	GGCGTGTCAGACCACATGGGTTATTCGCAGATTGGCGCACATGCTCACTGCGCATTCCCGAGT	AHAHCAFPSNQQSQLTAFVQK
	AACCAGCAATCCCAGCTGACCGCTTTTGTTCAAAAATTCCTGCTGGGCCAGTCTACGAATACCG	FLLGQSTNTAIFQSDFSANQSQ
	CGATCTTTCAGAGCGATTTTAGTGCCAACCAAAGCCAATGGATTGACTGGACGACCCCGACCCT	WIDWTTPTLS
	GTCT	
***Sc*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGACCCCG	MGSSHHHHHHSSGPQQGLRT
	GCGACCGTCTCAGGCTACTCGAACAGCGCACTGCCGGACCCGTTTACCTTCAATGATGGCTCC	PATVSGYSNSALPDPFTFNDGS
	CCGGTTACGACCGCAGAAGATTGGGAATGCCGTCGCTCTCAGATTCTGGCCCTGATCCAAGGC	PVTTAEDWECRRSQILALIQGY
	TATGAAAGTGGTGCAGCACCGCCGGAACCGGAATCCGTGACCGGTACGGCAAGCGGTAACTCT	ESGAAPPEPESVTGTASGNSLS
	CTGAGTGTGCAGGTTTCTTATGGCGGTAAAAGTATCACCTTCAACAACTCCATCACCTACCCGA	VQVSYGGKSITFNNSITYPSGTA
	GCGGTACGGCACCGGCCGAAGGTTGGCCGGTTATTATCGCGTACGAATTTCCGTCGCTGCCGA	PAEGWPVIIAYEFPSLPIPSNVA
	TTCCGAGCAACGTCGCGACCCTGTCGTTCCAGAATAGCGCCATGGGCAAACAAGATAGCACGA	TLSFQNSAMGKQDSTSSRGQG
	GCTCTCGTGGCCAGGGTCTGTTCTATGACCTGTACGGCAGTTCCTCAAATGCAAGCGCTATGAC	LFYDLYGSSSNASAMTAWAWG

	CGCGTGGGCCTGGGGTGTCTCTCGCATTATCGATGCAATCGAAAGTACCCCGGACGCTAAACT	VSRIIDAIESTPDAKLNPAAVGV
	GAACCCGGCAGCTGTCGGCGTGACGGGTTGTTCTCGTAATGGCAAAGGTGCACTGATGGCAG	TGCSRNGKGALMAGALEPRVA
	GTGCTCTGGAACCGCGTGTTGCACTGACGCTGCCGCAGGAATCCGGTTCAGGCGGTGATGCAT	LTLPQESGSGGDACWRLSRYE
	GCTGGCGTCTGTCCCGTTATGAAGAACAGCAAGGCTCACAGGTGCAAACCGCAACGGAAATTG	EQQGSQVQTATEIVGENCWFS
	TTGGCGAAAACTGTTGGTTTTCCGCTGGTTTCGATCAGTATGTTAACAATCTGGATTCACTGCCG	AGFDQYVNNLDSLPYDHHLLAA
	TACGACCATCACCTGCTGGCAGCCCTGGTCGCACCGCGTGGTCTGATCTCGTATGCTAACACC	LVAPRGLISYANTDYVWLSGMS
	GACTACGTGTGGCTGAGCGGCATGTCGAGCTTTGGTTGCATGACGGCAGCTCATGCGGTTTAT	SFGCMTAAHAVYEALGVPENH
	GAAGCCCTGGGCGTCCCGGAAAATCACGGTTTCGAACAAGTGGGCGGTCATAGTCACTGTCAG	GFEQVGGHSHCQWPSQLDGS
	TGGCCGTCTCAACTGGATGGCAGTCTGAACGCGTTTATTAATAAATTCCTGCTGGGTCAGGATG	LNAFINKFLLGQDVSTDYFESN
	TTTCGACCGACTACTTTGAAAGCAACAATCAATTTAATGGCGTCACCTGGTCCGAATCGCAATG	NQFNGVTWSESQWINWETPTL
	GATCAACTGGGAAACCCCGACCCTGAAC	N
***Pa*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGAACACG	MGSSHHHHHHSSGPQQGLRN
	CCGAGCGGTCTGGGCACGCCGGTCGCAAACCAACTGAATGATCCGTTCACGTTCCACAATGGC	TPSGLGTPVANQLNDPFTFHNG
	AATAAAGTCACGAGCAAAGCAGATTGGGCTTGCCGTCAACGCGAAATTAGCGAACTGCTGCAG	NKVTSKADWACRQREISELLQR
	CGTTATGAACTGGGCACCCTGCCGCCGAAACCGAGCTCTGTGACCGCCAGTTTTTCCGGTTCA	YELGTLPPKPSSVTASFSGSTL
	ACGCTGTCGATCAGCGTTTCTGAAGGCGGTAAAAGTATTTCCTTCACGGTCTCAATCAACAATC	SISVSEGGKSISFTVSINNRPSG
	GTCCGTCGGGTGCAGGTCCGCATCCGGCTATTATCAACTTTGGCACCTTCGGTGCAAGCCTGC	AGPHPAIINFGTFGASLPVPAGV
	CGGTCCCGGCGGGCGTGGCAACGATTAATTTTAACAATGATGACATCGCACAGCAACAGGGCG	ATINFNNDDIAQQQGGSSRGR
	GTAGTTCCCGTGGTCGTGGTAAATTCTATGATCTGTACGGCTCATCGCACAGTGCAGGTGCACT	GKFYDLYGSSHSAGALTAWAW
	GACCGCATGGGCTTGGGGCGTTTCCCGTATTGTCGATGCGCTGGAACTGACCCAAGCCCAGAC	GVSRIVDALELTQAQTGIDPTRL
	GGGTATCGACCCGACCCGTCTGGGCGTGACGGGTTGCAGCCGTAACGGTAAAGGTGCAATTGT	GVTGCSRNGKGAIVAGALEPRI
	TGCAGGTGCACTGGAACCGCGTATCGCACTGACCCTGCCGCAGGAAAGCGGTGCTGGCGGTT	ALTLPQESGAGGSGCWRIATW
	CTGGTTGTTGGCGCATTGCGACCTGGCAGAAAAACAATGGTCAAAACGTCCAGGATTCTACGCA	QKNNGQNVQDSTQIVQENVWF

	AATCGTGCAGGAAAATGTTTGGTTTAGTCCGAACTTCAATTCCTATGTGAACAATGTTAACCAGC	SPNFNSYVNNVNQLPFDHHLLA
	TGCCGTTTGATCATCACCTGCTGGCTGGCCTGATTGCACCGCGTGCTCTGTATGTCATGGAAAA	GLIAPRALYVMENVDMEWLGKI
	TGTGGACATGGAATGGCTGGGTAAAATTAGTACCTACGGCTGCATGGGTATCGCGCGCAAACA	STYGCMGIARKQWEALGALDN
	ATGGGAAGCGCTGGGCGCCCTGGACAACTTTGGTTACTCACAGGTTGGCGGTAATTCACATTG	FGYSQVGGNSHCSFPSSQQGS
	TTCGTTCCCGAGCTCTCAACAGGGTTCGGAACTGAATGCCTTTATTGAAAAATTCCTGCTGAAAC	ELNAFIEKFLLKRSGGNTNIFRS
	GTAGCGGCGGTAACACCAATATCTTTCGCTCTACGCAAACGCACTCTTCTTTTAATCTGAACAAC	TQTHSSFNLNNWSPWAVPSLN
	TGGTCCCCGTGGGCCGTCCCGTCTCTGAAT	
***Bc*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGGTGTTG	MGSSHHHHHHSSGPQQGLRVL
	CCGGGTATCAACGAACTGCCGGTAGTAAAAACACTGACCGATCCCTTTCAATGGTCCGATGGCA	PGINELPVVKTLTDPFQWSDGS
	GTGGCCGTTCAACCAACTTCAAGGACTGGAGCCGACGGCGTGCAGAGATAGCCCGCGAAATAG	GRSTNFKDWSRRRAEIAREIEH
	AACACTACGAAATCGGCGAGAAACCCATAGTGTCGAAAAAAGACATTACCGCGGATATCGTGGA	YEIGEKPIVSKKDITADIVDDTLR
	TGATACCCTTAGAGTAAATGTAACAGTGAACGGACAAACGCTCACCCTGAAAGCAAAAATTACTT	VNVTVNGQTLTLKAKITYPAGK
	ATCCGGCAGGTAAAGGCCCTTTCCCAGCCATCATTGGAATAGGCAGAGGCACGGGTAGTCTTC	GPFPAIIGIGRGTGSLPPTIFTSR
	CTCCCACCATCTTCACGAGCCGCAATATAGCACAAATCGCATTCAACTTCACACAAGTCATGTC	NIAQIAFNFTQVMSHTQKRGNE
	GCACACACAAAAGCGCGGCAACGAGCCTATCAACAAACTTTATCCTGACCGTACAGACATGGGA	PINKLYPDRTDMGAYCAWSWG
	GCCTATTGCGCCTGGTCCTGGGGAGTGAGCCGTATCATTGACGGATTGGAAATCGTAGGTAAA	VSRIIDGLEIVGKKSKIDTKRLAIS
	AAGAGCAAGATAGATACCAAACGACTGGCCATATCCGGTTGTTCCTTCGCCGGGAAAATGGCTC	GCSFAGKMALFAGAFDERIALTI
	TTTTTGCCGGAGCATTCGATGAGCGTATTGCACTGACCATTGCTCAGGAACCGGGTGGTGGCG	AQEPGGGGAAAWRVSETLGEV
	GTGCGGCAGCATGGCGTGTTTCCGAAACGCTGGGTGAAGTGGAAACTTTGGGAAGAACCAGTC	ETLGRTSHAWFKESMFQYKED
	ATGCCTGGTTTAAAGAAAGTATGTTCCAATATAAAGAAGATAATGTATCGAAATTGCCGTATGAC	NVSKLPYDHHELCAMVAPRALL
	CACCACGAACTGTGTGCCATGGTAGCACCGCGTGCACTGCTGGTACTCGGAAACCCGGACTAT	VLGNPDYVWLADESGYVSCRA
	GTGTGGCTGGCTGATGAATCGGGTTATGTCTCCTGCCGTGCTGCCCGTAAAGTATGGGAGACC	ARKVWETFGIADRMGFSIVGGH
	TTCGGCATTGCCGACAGAATGGGCTTCTCCATCGTAGGCGGACATGGGCACTGCCAGCTGCCC	GHCQLPESQYPEVEAFVDKFLL

	GAAAGCCAGTATCCTGAAGTAGAAGCGTTTGTCGACAAGTTCCTGCTTGGAAAGAAAGACGCGA ATACAGAAGTAACCATTGCACCACTTTACGAAAAGGTGGACTATGAGCGCTGGTTATTCCAC	GKKDANTEVTIAPLYEKVDYER WLFH
***Ck*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGATAGAT AATGCTTTGGCTGCTCTCAAGCAGGCATTGAACAGTTTAAGAAGAAGCGATAAATATCCAGAAC CAGAAACACTGCCTATTGTAGAGACATTGCCTGATTCGTTTACCTTCTACGATGGAACAAAAGTA CAGAGATTGAGTGATTGGCCAAAGAGAGCACAGGAGCTCAAGGATTTATATCAATTTTACATGTA TGGTTACAAGCCTGATACTTCTGTGGAAGATGTAACTTACTCAGTAAATGGCAATACATTAACCA TAACTGTCAAGGTCGGTGATAAACAGGCATCATTCAATGCTACAGTAAGGTTACCTCAAGCAAAT TCAGGATATCAGCCACCATATCCAGTTATTATTTCACTTGGTTATCTTGCAGGATTCAACTGGCA AACATGGCAGTTTATTGATTATTCTACTAACGCAGTTAACAGAGGTTATGCAGTAATTTCGTTCAT GCCAAATGATGTTGCGCGCGATGATTCCTCGTATACAGGAGCTTTTTACACCTTATATCCGCACT CTAATAAGGTAGAAAATGATACGGGAGTATTGATGGCATGGGCATGGGGAGCATCAAAGATTTT AGATGCTCTTGAAAAAGGTGCAATTCCAGAAATTGATGCTAAAAAAGCTATTGTAACTGGTTTTT CAAGATATGGTAAAGCTGCTTTGGTCGCTGGTGCATTTGATGAAAGGTTTGCTGTAGTTAATCCA CATGCTTCAGGACAAGGCGGTGCTGCATCATTCAGATACAGTTTTGCTGGCAAGCAATATAGCT GGGGTGTAGCAGGCAATGCAGAGGCATTTTCTAACTTACAGGGAAACACTGAAGGCCACTGGT TCAATGCAGTGTTCAGAGAGTTCAAAGACCCACGCCAGCTTCCATTTGACCAGCACGAACTTAT TGCACTTTGTGCACCAAGAACAGTGCTGATTACAGGTGGTTATTCAGATTGGGGAACAAATCCA GAGGGTACATGGGTATCATTTGTTGGTGCAAGAAAAGTATATGAGTTCTTAGGTGTTGCCGATA GGATAGGATTTGCTTTAAGAGATGGAAGTCATGCTATAACTGAAGAAGATGTAAATAACTTGCTA GATTTCTGTGACTGGCAACTGAGGGGCATACAGCCAACAAAAGATTTTTCAACAAGCCGTTTTG CTATTGATCCAGCATGGGACACTATCTCTGTGCCTACACTTTACCGAAATGCTGATTTGAAACAG ATA	MGSSHHHHHHSSGPQQGLRID NALAALKQALNSLRRSDKYPEP ETLPIVETLPDSFTFYDGTKVQR LSDWPKRAQELKDLYQFYMYG YKPDTSVEDVTYSVNGNTLTITV KVGDKQASFNATVRLPQANSG YQPPYPVIISLGYLAGFNWQTW QFIDYSTNAVNRGYAVISFMPN DVARDDSSYTGAFYTLYPHSNK VENDTGVLMAWAWGASKILDA LEKGAIPEIDAKKAIVTGFSRYG KAALVAGAFDERFAVVNPHASG QGGAASFRYSFAGKQYSWGVA GNAEAFSNLQGNTEGHWFNAV FREFKDPRQLPFDQHELIALCA PRTVLITGGYSDWGTNPEGTW VSFVGARKVYEFLGVADRIGFA LRDGSHAITEEDVNNLLDFCDW QLRGIQPTKDFSTSRFAIDPAW DTISVPTLYRNADLKQI

***Cj*GE15B**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGCAACTT	MGSSHHHHHHSSGPQQGLRQ
	TCTGATCCCCTGTTGCCGGCCGGTGTGGAAAACTCCGGTGCGGGATGTTATGTCCCGGCATTG	LSDPLLPAGVENSGAGCYVPAL
	CCGGCTTATGCGAACCTGACCGCGGTGGCAGCCTTGCCTGACCCTTTCCGCGCGTTAAATGGC	PAYANLTAVAALPDPFRALNGN
	AATCGCATTACCAGCAAATACGATTGGACCTGCCGCCGCGCCGAAGTGAATAAGCAGGCGCAA	RITSKYDWTCRRAEVNKQAQF
	TTTTACGAATTGGGCGAGAAGCCAGAGCCGGGCAGCAATGCGGTAACGGCTTCTGCATCGGGC	YELGEKPEPGSNAVTASASGN
	AATAGTATTACGGTGACCGTGCAGGATGCGGGTCGCTCTATTTCCTTCAGTGCCAGCATTCAGT	SITVTVQDAGRSISFSASIQLPS
	TGCCCTCGTCGGGGACGGCACCTTACCCGGCGGTGATTGGCATGGGTGGCTCTTCATTGAATA	SGTAPYPAVIGMGGSSLNNNGL
	ACAATGGTTTGCTCAGTCGCGGTGTGGCAGTGATTAATTTTAATAACAATGATATTGCTGAGCAG	LSRGVAVINFNNNDIAEQTNGS
	ACGAACGGCAGTTCCCGCGGTCTTGGCAAGTTTTATACGCTCTATGGCAATAACCACAGTGCCG	SRGLGKFYTLYGNNHSAGAMT
	GTGCTATGACTGCCTGGGCCTGGGGGGGTGAGCCGTTTGATCGATGCCCTGGAAAAAACACCGG	AWAWGVSRLIDALEKTPAARID
	CAGCGCGCATTGATGCAAAACGCCTGGGTGTGACCGGCTGTTCGCGCAATGGCAAGGGTGCC	AKRLGVTGCSRNGKGALIAGAL
	TTAATTGCCGGCGCGTTGGATGAGCGCATTCTATTGACGATTCCACAGGAGAGCGGCTCCGGT	DERILLTIPQESGSGGSAAWRV
	GGTTCGGCGGCCTGGCGCGTATCCGATGCGCAGCGCAGTGCGGGGCAGAATGTGCAGACCCT	SDAQRSAGQNVQTLSQIVTENV
	GAGCCAGATCGTGACCGAGAACGTGTGGTTCCGCAGTAATTTTTCACAGTTTGCCAACACGGCA	WFRSNFSQFANTATRLPFDHH
	ACACGCTTGCCCTTTGACCATCACCAGGTGATGGGGCTGGTGGCGCCGCGCGCACTGCTGGT	QVMGLVAPRALLVLENTSMEW
	ATTGGAAAATACCTCCATGGAGTGGCTGGGTAATGTGAGCACATATACCGCGTCGGTCGCGGC	LGNVSTYTASVAAREIWTALGIT
	GCGTGAAATCTGGACGGCGCTGGGGATTACCGATCGTATGGGCGTATCGCAACTCGGTGGCTA	DRMGVSQLGGYQHCVFPSSQ
	CCAGCATTGTGTGTTCCCCTCATCCCAACAGGCGGTACTGGATGCGTTTGTCGATAAATTCCTG	QAVLDAFVDKFLRGVSSANTNV
	CGCGGTGTGAGTTCTGCCAATACCAATGTGATGCGCACGGATGGCAATTACAGCGTGGATCGC	MRTDGNYSVDRARWINWTTPT
	GCGCGTTGGATTAACTGGACTACCCCCACTCTCCAA	LQ
***Cj*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGTCTGTT	MGSSHHHHHHSSGPQQGLRS
	CCTCTTTCCGGTGTGGAGAACTCCGGGCGCGATTGTTTTGTGCCGGCATTGCCCGCCTATGCC	VPLSGVENSGRDCFVPALPAYA
	AATTTGGCGTCGAGCATAGCTGCCTTGCCTGATCCATTTCGGGGGTTGGATGGTGAATCGATTA	NLASSIAALPDPFRGLDGESISS

	GCAGCAAACATGATTGGACTTGTCGCCGCGCTGAGGTGAGCAAACAGGCGCAGTATTATGAGC	KHDWTCRRAEVSKQAQYYELG
	TGGGTGAAAAACCGGAACCTGATACCGATGCAGTGACCGCCGCTGTGGTCGATAACACGATCA	EKPEPDTDAVTAAVVDNTITVTV
	CGGTAACGGTCGAGGATAATGGCAGTGCTATTGCGTTTAGCGCGACTATCCAATTACCGACAAC	EDNGSAIAFSATIQLPTTGVAPY
	AGGCGTTGCACCTTATCCGGCGATGATCGGTATGGGTGGATCATCGCTCAATAATACTGAACTG	PAMIGMGGSSLNNTELLNRGIA
	TTAAATCGCGGTATCGCGGTGATTAACTTTAATAATAATGATATTGCCGAGCAGACTAACGGTAG	VINFNNNDIAEQTNGSSRGRGK
	TTCCCGCGGTCGTGGCAAGTTTTACACACTCTATGGCAGCAACCACGGCGCGGGCGCCATGAC	FYTLYGSNHGAGAMTAWAWG
	GGCCTGGGCCTGGGGTGTCAGCCGTTTGATTGACGCGCTGGAAAAAACACCGGATAGCCAGAT	VSRLIDALEKTPDSQIDTRYLGV
	AGATACCCGCTACCTGGGTGTCACCGGATGTTCGCGCAATGGTAAGGGGGCGCTGATTGCCG	TGCSRNGKGALIAGALDERIALT
	GGGCCCTGGATGAGCGTATTGCGCTCACCATTCCACAGGAAAGCGGCTCTGGCGGCTCGGCT	IPQESGSGGSAAWRVSDAQKA
	GCCTGGCGTGTGTCCGATGCGCAAAAAGCGGCGGGGCAAAACGTACAGACCCTGAGCCAGAT	AGQNVQTLSQIVTENVWFRSSF
	AGTGACGGAGAATGTGTGGTTTCGCAGTAGTTTCGCGCAGTTTTCTAGCACTGCAACGCGTTTG	AQFSSTATRLPFDHHQVMGLV
	CCTTTCGATCACCATCAGGTGATGGGATTGGTTGCGCCACGCGCGCTGTTGGTGTTGGAAAATA	APRALLVLENTSMEWLGNVSTY
	CCTCTATGGAGTGGTTGGGTAATGTGAGCACCTATACCGCCGCCGTGGTCGCGCGCGAAATCT	TAAVVAREIWIALGVTDHMAVS
	GGATCGCTTTGGGCGTTACCGACCATATGGCGGTTTCCCAGCTTGGCGGCTATTGGCATTGCA	QLGGYWHCTLPESQQPLVDAF
	CATTACCGGAATCCCAACAGCCCTTGGTGGATGCCTTTGTGGATAAATTTTTAAAAGGCATCAC	VDKFLKGITTADTDVVRTDGEY
	GACTGCGGATACCGATGTGGTGCGCACTGACGGTGAGTATACGGTTGACCGGGCACGCTGGA	TVDRARWIDWETPGLH
	TTGATTGGGAAACTCCCGGTTTACAT	
***Cc*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGTCAGAT	MGSSHHHHHHSSGPQQGLRS
	CTGGTTTATGGAGACCTGAACGGCGATAGGGAAGTTAATGCTATTGACCTTTCATTAATGAAAGA	DLVYGDLNGDREVNAIDLSLMK
	ATACTTGCTCGGAAAGATCAGCACTTTTCCGGGCGGACAGGGGAAAGAAGCTGCTGATGTGGA	EYLLGKISTFPGGQGKEAADVD
	CTTGAGCGGCAGCATAGATGCAATCGATATGTCATATTTAAAACAATTCCTGCTGGGAACCATTG	LSGSIDAIDMSYLKQFLLGTIDG
	ACGGGTTGCCTGTCGGAGAACCTCTGGAACCGATACAAATCCCTTCCTTTGATTCATTGCAGCC	LPVGEPLEPIQIPSFDSLQPNAK
	AAATGCCAAACTTCCCGACCCCTTTATGTTTAAGACCGGCAGCAAGAAAGGTACTCGTATTACAA	LPDPFMFKTGSKKGTRITSKSQ

	GCAAGAGCCAGTGGACAGCCCGCCGTGCCGAAATCTCAGCCTTGGCCCAGGCCTTCGAATTTG	WTARRAEISALAQAFEFGVKPP
	GTGTAAAACCGCCGAAACCGCAGACTGTCACAGGCTCCTTTAATAACAACTCAATAACGGTAAC	KPQTVTGSFNNNSITVTCSQNG
	CTGCAGCCAAAACGGAAAATCAATTTCCTTTAGCTGTGCAATCCAGTATCCAACTACCGGTACA	KSISFSCAIQYPTTGTAPYPAMI
	GCCCCTTATCCCGCTATGATAGGTGTAAATATGAACACACTTAATACCTCAGAGATACTCAAATT	GVNMNTLNTSEILKLGVALITFP
	GGGGGTTGCCTTGATAACTTTTCCTGCCGACCAGATCGGAAAGGAAGATAACGCCGGTTCCAG	ADQIGKEDNAGSRGQGKFFDL
	GGGACAGGGTAAATTCTTTGACCTGTATGGAAGCAGTTACGATGCCGGTGCATTGATCACATGG	YGSSYDAGALITWAWGVDCLID
	GCCTGGGGTGTGGACTGTCTGATAGACGCGTTGGAAATGACTCCGGCAGCGAGAATCGACCC	ALEMTPAARIDPKKLGVTGGSR
	GAAAAAGCTGGGTGTAACAGGCGGTTCCCGCAATGGAAAGGGAGCACTTGCCATCGGCGCCTT	NGKGALAIGAFDERIALTVPQES
	TGACGAAAGAATTGCTCTTACCGTACCTCAGGAATCCGGGAACGGCGGCGCATCGGGCTGGAG	GNGGASGWRTADAQKAAGED
	AACCGCCGATGCACAGAAAGCGGCCGGTGAAGATGTTCAGACACTTTCCCAGATAATTACTGAA	VQTLSQIITENCWFAKSLNQFS
	AATTGCTGGTTCGCCAAATCTCTCAACCAGTTCAGCGGCCAGACAAACAAACTGCCGTATGACC	GQTNKLPYDHHELMALCAPRG
	ACCATGAATTAATGGCTCTGTGTGCCCCCCGTGGATTGCTTGTAATAGAAAATCCTGATTTTACA	LLVIENPDFTWLGNLSCFNTSTA
	TGGTTGGGCAACCTCAGCTGCTTTAATACTTCAACTGCCGCACATATGGCATATGAAGGTCTTG	AHMAYEGLGVPDNMGYSSVG
	GCGTACCCGACAACATGGGCTATTCCTCTGTGGGAGGCCACGGCCATTGCCAGTTCCCCGCAT	GHGHCQFPASQQPELTAYIQKF
	CACAGCAGCCGGAACTGACAGCCTATATACAGAAATTCCTGCTGGGACAGAATAGCAACACCAA	LLGQNSNTKVFRSDRNYTFDKA
	GGTCTTCCGGTCGGACAGGAACTATACTTTCGATAAAGCAAAATGGGTTGATTGGACTGTTCCG	KWVDWTVPTF
	ACATTT	
***Ma*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGGATGCC	MGSSHHHHHHSSGPQQGLRD
	AGCAAACTACCCACTATTACGACACTGCCGGATCCATTTATGGGGCTGGATGGTGAGCGCATTG	ASKLPTITTLPDPFMGLDGERIA
	CCAGCAAAGACGATTGGCGCTGCCATCGTCAGAATCTGAACTGGCAACTGCAAGCGTACGAAG	SKDDWRCHRQNLNWQLQAYE
	CTGGCATTAAACCCTATAAGCCCGAAACGGTATCGGGCAGCGTGGGCAATGACGCCATCGAAG	AGIKPYKPETVSGSVGNDAIEVT
	TGACTGTCGGTCATGGCGGAGAGAAAATTACTTTCTCGGCCTCGGTGACGTTGCCCAGCGCGG	VGHGGEKITFSASVTLPSAGEA
	GCGAGGCGCCATATCCCGCCATGATCGGAATCGGTGCGTCCAACCTGGACAACGGCTATTTAG	PYPAMIGIGASNLDNGYLARQGI

	CCAGACAGGGCATCGCGGTCATTAACGTCAATAACAACGAACTGGGTGCGCAAAGCGGCGGTG	AVINVNNNELGAQSGGGSRGT
	GTTCACGTGGCACCGGCCTGTTCTATGACCTCTACGGCAGAGAGCACTCGGCCAGCTCAATGA	GLFYDLYGREHSASSMTAWAW
	CTGCCTGGGCCTGGGGGGTAAGCCGGTTGATCGACGTGCTGGAGACCCCGGAGGCGGAGCT	GVSRLIDVLETPEAELIDATRLG
	GATTGACGCCACTCGGCTCGGAGTCACCGGATGCTCGCGCAACGGCAAAGGGGCGCTTCTTG	VTGCSRNGKGALLAGALDQRIA
	CCGGTGCTTTGGATCAGCGTATCGCCCTGACCATCCCTCAAGAATCTGGGGCCGGCGGCGCG	LTIPQESGAGGAMAWRVAQEM
	ATGGCCTGGCGGGTCGCTCAGGAAATGGCGGACAGCGGTACCAACATCCAGACACTCAGCCA	ADSGTNIQTLSHAAGEQPWFR
	CGCCGCCGGTGAGCAACCCTGGTTCCGGGAATCCTTCGGCGCCGACTTCGGAGGCCCGAACA	ESFGADFGGPNTVTLPFDHHQL
	CCGTCACACTGCCATTCGATCACCACCAACTGATGGGTATGGTGGCTCCCCGCGGCCTGTTGG	MGMVAPRGLLVLDNNIDWLGP
	TCCTCGACAACAATATCGACTGGTTGGGTCCGAGAGCGGCCTATATTGGAACGGCTGCAGCGA	RAAYIGTAAAKEIYRALGAPENI
	AGGAAATTTACCGTGCCCTGGGTGCACCCGAGAACATCGCCTACTCGGAAAACGGCGGCCACG	AYSENGGHGHCQFPAHQQDIL
	GTCACTGTCAGTTTCCGGCCCATCAGCAGGACATACTGGCGGCATTCGTAAAGCGCTTTCTGTT	AAFVKRFLLGELGSTEVMRSTQ
	GGGGGAGCTCGGTAGCACCGAGGTGATGCGATCCACTCAGGCTGATGAAGAGGATGTTGGGG	ADEEDVGDWVDWSTPTLD
	ATTGGGTTGATTGGAGTACTCCAACGCTCGAT	
***Mr*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGCCGTTG	MGSSHHHHHHSSGPQQGLRPL
	CCGGAATTCAGTCAACTGCCGGTAATCAACGAACTGCCGGACCCATTCGAATGGTCTGACGGC	PEFSQLPVINELPDPFEWSDGR
	AGAGGCAGGATTACGAGCAAACTCGAATGGAGATGCCGGCGCGCTGAAATAGGCGAAGAAATT	GRITSKLEWRCRRAEIGEEIQYY
	CAATATTACGAACTCGGGTCTAAGCCTCCCAAACCGGAGAATATGACGGTTACATATACCGACA	ELGSKPPKPENMTVTYTDTLMT
	CTCTCATGACAATTGTTATTGTCGAGAATGGCGATACGCTGGTTATAAACGCGCCTATTAGTCTA	IVIVENGDTLVINAPISLPEGEGP
	CCTGAAGGGGAAGGACCGTTTCCCGCGGTTATAGGAGTGGGATTTACGCCGACAGGCAGTTTG	FPAVIGVGFTPTGSLPEDLFTSR
	CCGGAGGATCTTTTTACAAGCAGGGGAATTGCCACAATTCATTTTATGGAATGGCAGCTTACCA	GIATIHFMEWQLTNGWSGSRG
	ACGGTTGGTCGGGCAGCCGTGGAGACGGACAGTTTTATAAACTTTATCCTGATAAAAAGAGGG	DGQFYKLYPDKKRGKFIAWAW
	GTAAATTTATTGCCTGGGCATGGGGCGTTAGCAGAATTATCGACGCGCTTGAAATGTCGCCCGA	GVSRIIDALEMSPESKIDLKHLAI
	ATCAAAAATCGATTTAAAACATCTTGCTATAACCGGTTGTTCTTATGCGGGCAAAATAGCTTTGTT	TGCSYAGKIALFSGALDERIALTI

	TTCCGGGGCGCTCGATGAACGGATTGCATTGACTATAGCGCAGGAGCCGGGCGGCGGAGGCG	AQEPGGGGDAAWRVTEKLPGS
	ACGCTGCATGGCGCGTGACCGAAAAATTGCCCGGCAGCAGGGAACGACTGCGAAACGCGCAA	RERLRNAQGYAWYHQDLGRFN
	GGTTATGCATGGTATCATCAGGATTTAGGACGCTTTAATTCAGCCGTTACTAAATTGCCGTACGA	SAVTKLPYDHHELMAMIAPRAL
	TCATCATGAATTGATGGCAATGATTGCACCACGCGCCTTGTTGGTGCTGGGCAACCCGGATTAC	LVLGNPDYEWLAEESGHVACK
	GAATGGTTAGCGGAAGAATCCGGTCACGTGGCATGTAAAGCCGCTCATGAAGTTTGGAAAGCT	AAHEVWKALGVPDRFGFSKVA
	CTCGGCGTGCCCGACCGTTTCGGATTTTCGAAAGTGGCTGGTCATCAGCACTGCCAATTGCCC	GHQHCQLPGVQRPEVTAFIEKF
	GGTGTACAACGTCCCGAAGTTACTGCATTCATCGAAAAATTCTTACTCGATATCGATACCGTAAA	LLDIDTVNTNIEISPYDPDLSSWI
	TACTAATATCGAAATAAGTCCTTACGACCCGGATTTGTCGAGCTGGATTACGTGGACAACCCCG	TWTTPELSDTLASVGQNIIEKEY
	GAACTTTCCGACACGCTTGCTTCAGTAGGCCAAAATATTATTGAAAAAGAATATGAATTGATGCA	ELMHNYPNPFNPVTKIVYRIPEN
	TAATTATCCGAATCCGTTCAATCCGGTTACTAAAATAGTTTATAGAATTCCCGAGAACGCTTTTAT	AFISLKVYNSLGQEIRTLFEGYRI
	AAGTCTGAAAGTATATAATTCGCTTGGGCAGGAAATAAGAACTTTATTCGAAGGGTATCGAATCA	KGKYEVLFDAADLTTGVYLCRL
	AAGGTAAGTATGAAGTCCTATTCGATGCCGCCGATTTGACTACCGGCGTTTATTTGTGCAGATTA	SSLNFDKTTKLLLLK
	TCGAGTCTAAATTTTGATAAAACTACTAAACTTTTACTTCTCAAA	
***Ot*GE15B**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGTCGGGT	MGSSHHHHHHSSGPQQGLRS
	CGTCAGGACGGCTCGCGCATTCCGTTGGTCTACGACGTCGAAAACACCGGCGCGCACTACCGT	GRQDGSRIPLVYDVENTGAHY
	GCGCCCAGTTTCCCGGATTTCGACTGGCTGCCCATCGTTCGGCCGTTGCCCGATCCGTTCCGA	RAPSFPDFDWLPIVRPLPDPFR
	TTCGAGAACGGCACTCGCAGCACCTCGTTCAGAAATTGGGAACGTCGGCGCAACGAAATCAAG	FENGTRSTSFRNWERRRNEIKA
	GCGGCGATCGAGAAATATGAGATCGGCAGGAAGCCCGACGCGTCGGACCTCACGATCGCCGC	AIEKYEIGRKPDASDLTIAATYTP
	GACGTATACCCCGCCCGCCGCCGGTACGAGCATCGGCCAGCTGGTCGTGGTGGTAACGCGAA	PAAGTSIGQLVVVVTRNSNAKT
	ACTCGAACGCGAAAACCGTCACGCTGACTTCGAAGGTGTTCATCCCGCAGGGTTGGGGCGAAG	VTLTSKVFIPQGWGEGPFPALIP
	GGCCGTTCCCCGCGCTGATCCCGATGACGTTCTTTGCCAGTCCCACCGGGCCCAACTACGGCA	MTFFASPTGPNYGSLAAVLSTR
	GCCTGGCCGCCGTGTTGTCGACCCGCCCGATCGCCACGGTGGATTTCGTGCACAACCAGGTC	PIATVDFVHNQVTRYGGGDKTP
	ACGCGCTATGGCGGCGGCGACAAAACGCCCGACCCATTCTATCAGATGTATCCGGAATTCAAG	DPFYQMYPEFKAPGGPVDSDS

	GCGCCGGGCGGACCCGTTGACAGCGACTCGGGCCAATACGCGGCCTGGTCGTGGGGCGTGA	GQYAAWSWGVSRLIDGIVLASQ
	GCCGGCTGATCGACGGCATCGTGCTCGCCAGTCAGCAGGAGGAAAACCCGCTGCCGATCGAC	QEENPLPIDPDYLAVTGCSYAG
	CCTGACTACCTCGCGGTGACCGGCTGCTCCTACGCCGGCAAGATGGCGCTTTTCGCCGGCGC	KMALFAGALDERVALTIPLESG
	GCTCGACGAGCGGGTCGCGTTGACGATCCCGCTCGAGAGCGGCGGCGGCGGCGCGCCCTCG	GGGAPSWRVSQEIEGDRVVEA
	TGGCGGGTCTCGCAGGAGATTGAGGGCGACCGCGTGGTGGAGGCGCTGACGCATACGGATCG	LTHTDRRWFASQLFQFSGNNV
	GCGCTGGTTCGCCTCGCAGCTGTTCCAGTTCAGCGGGAACAACGTCTACAAGCTGCCGTATGA	YKLPYDHHELMAMVAPRALLVT
	TCACCACGAGCTGATGGCCATGGTCGCACCCCGCGCGCTGCTCGTCACCGGCAACACGGATTT	GNTDFLWLSNRSNYVTSRATQ
	TCTGTGGCTTTCGAACCGCTCCAACTACGTCACGTCGCGCGCCACCCAGCGGATTTACGAGAC	RIYETLGIGDRFGFYIDGGHGH
	GCTCGGGATCGGCGACCGGTTCGGTTTCTACATCGATGGCGGACACGGCCACTGTCAGGTGC	CQVPASQYPAIANFVDKFLFGV
	CGGCGAGCCAGTACCCGGCGATCGCGAACTTCGTCGACAAGTTCCTGTTCGGCGTCGAGGAC	EDADTQVRVHPYGEDFDYRRW
	GCCGACACCCAGGTGAGAGTCCATCCCTACGGCGAAGACTTCGACTATCGTCGTTGGACGGCG	TAWWGHGLPFFP
	TGGTGGGGTCACGGGCTGCCGTTTTTCCCC	
	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGTCCAAC	MGSSHHHHHHSSGPQQGLRQ
	GAAAACCCCAGCGCCAGCTGTAATGTAGGTCCGCTGCCCGGAGCAGTATCTGGCACCAACAAC	QGGQLPLAYQVENTGAAFAKP
	CTGCCTGATCCGTTCCGCAAACTCGACGGCACGCGCATGACGAGCAAGACCGAATGGCGTTGT	AMPAANELPVIKDLPDPLKGVN
	CGCCGTGCGGAAATCCTCGAGCAGGCCTACACCTATATTTACGGTGAAAAACCCGCCAAACCC	GFSDWARKRSEIAALIQHYGIG
	AGCAGTGTGACGGGTACTGTCAGCAGCTCCTCGGTGCAGGTCAGCGTGTCGCACAACGGCCG	EKPAVKKENIKARMSGDTLIVDV
***Pr*GE15A**	TTCAACCTCGTTCAGCGCGTCTGTCGAACTGCCTTCCACGGGTTCCGCTCCCTACCCCGCAGT	TVNGQTLTLTSEIRYPKTGKAPY
	CATTGGCTTCGGTGGTGGATTCTTCGGTATCGCAGGTGGCATGAAAGATGTCCTTAAATCTCAG	PLMIGSSMIALPRQLFEDRPIAT
	GGCGTTGCGATTATCAACTTCGATCCATACCAAGTGGGTGCAGAAGGCAGCGGACAAGGCAAC	MNFHEKQVNDYGQWGKHHER
	GGACGCTTCTACGACATCTATGGCTCAAGCCAGGCGGGTCTGCTTACCGCGTGGGCATGGGG	GEHNFDRLYPQLKDNGAYSEW
	CGCCAGCCGTATTATCGACGTGTTGGAAAGCTCTAGCCTGATCGACGCTGACAGCGTGGGTGT	AWGFSRLIDGLELLGPEVTKIDT
	TACCGGGTGTTCGCGCTTCGGTAAAGCCGCCTTTATCGCCGGCGCATTCGACGGCCGTGTCGC	KRIGVTGCSYAGKMALYCGAFD

	TTTGACCATCCCCAACGAGTCTGGTATCGGCGGCGTACCCGCCTTGCGCCTGGTACCCGTGGT	ERVALTIAQEPGGGGAAAWRY
	CGACAGCGGTGGCGAACAACCTTCGCACGCGATCAACTACGAACCCTGGTTCTCGCCAAATGC	SHLQDSVENLDKTDYHWFLES
	TTTCCGCACTTTCGCCAGCAGCCCGAACCGCTTGCCGATTGATACCCATGAGGTTATTGGTTTG	QLTNFHGDSVYQLPYDQHELC
	GTGGCGCCACGCGGCCTGCTTATTCTGGATAACCCACACATCGGCAATCTCGACTATCGCTCCT	ALICPRALLLLGNPDYKWLADD
	CCTACGCTGCAGCCGCAGCCGGTAAGCGCATTTATGAAGCACTGGGTGTGAGCAGCAACATGA	AMLVSAKAAKKVWERFGIADR
	GCTTCCACAATAACGTAAGTGACGGCAGCCACTGCGCGTGGAAAAGCGAGTTTGCCCAACCTC	MGWSIVGGHGHCQLPECQWP
	TGCGCGATAACGTGTCGAAGTTCCTGAAAGGCGGCAGCGGCAATACAGGCTCTATCAATGCCC	EVLAFIDKFLLGKDTKTSDIRVY
	GCTCTGGTAGCACAGTGAACGTGGACTCCTACATCAACTGGACCACACCAAACCTGTCCGGCA	SKTLIK
	ACCTGGACTTC	
***Rf*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGGATAGC	MGSSHHHHHHSSGPQQGLRD
	GGTCCAGAGCTTATTTACGGTGACCTTGACGGAGATAAGACAATAACCTCCTTTGACGCTGTTA	SGPELIYGDLDGDKTITSFDAVI
	TCATGCGAAAGGGACTCATAAACGACTTCAAGGATAACAACGTAAAGAAAGCCGCTGATATAGA	MRKGLINDFKDNNVKKAADIDQ
	TCAGAACGGCAAGGCAGAGGTAGCGGACCTTGTCCAGCTCCAGAGCTTTATAATCGGTAAGAT	NGKAEVADLVQLQSFIIGKIKEF
	AAAAGAGTTCACCGTGGCCGAAAAAACTGTCACAGAAAAGCCTGTTTTTGAAAAGAGCTACAAC	TVAEKTVTEKPVFEKSYNFPAV
	TTCCCTGCTGTAAATCAGCTGAAAAGCTCAAAGGATATTCCAGATCCTTTCATATTTATGGACGG	NQLKSSKDIPDPFIFMDGSKVE
	TTCAAAGGTAGAATCCACCGACGACTGGTGGAAGCGTCAGAGCGAGATAAGCTGTATGTATGA	STDDWWKRQSEISCMYEYYMY
	GTACTATATGTACGGCAAGTGGATAGACGGCTCTGATGACGAAACCACCTACAGCATAAGCGG	GKWIDGSDDETTYSISGNSMTI
	CAACAGCATGACCATAAACGTTAAGCGCAAGAGCACAGGTAAGACAGCTTCATTCAAGGCGGTT	NVKRKSTGKTASFKAVINLPKN
	ATAAATCTTCCTAAGAATGTTCGCCATGAGGGCGGAGCTCCTGTTATACTTGGTATGCACAAGG	VRHEGGAPVILGMHKGISESTA
	GCATTTCCGAGAGCACAGCTACCTCAAACGGCTATGCTGTTATTACCTACGACAGTGACGGAAT	TSNGYAVITYDSDGMFSAPGTA
	GTTCAGCGCTCCTGGAACTGCACAGGATAACAACCAGCACAAGGGTGCTTTCTATGATCTCTAT	QDNNQHKGAFYDLYPYGRNW
	CCTTACGGAAGAAACTGGGACGAGCAGACAGGCGACCTTATGGCGTGGTCATGGGGAATCAGC	DEQTGDLMAWSWGISRILDALY
	AGGATACTCGACGCTCTTTACAACGGAGCCGCAAAGGAGCTGAACATCAATCCTGACAGCTCTA	NGAAKELNINPDSSIVTGVSRY

	TCGTTACAGGTGTTTCAAGATACGGTAAAGCAGCTTCTGTATGCGGAGCATTCGATACACGTAT	GKAASVCGAFDTRIKMCAPSCS
	CAAGATGTGTGCTCCCTCATGCTCGGGCGCAGGCGGTCTGGCACTGTATCGCTACAGCTCAGT	GAGGLALYRYSSVGKTYDFSSK
	GGGAAAGACCTATGACTTCTCCAGCAAGGGCGGATCTTCATCGTATACATATAAGGAAAATGAG	GGSSSYTYKENEPLGSLQASG
	CCTCTCGGAAGCTTACAGGCTTCGGGCGAGCAGGGCTGGTTCAACGGAAGATTCATGGAGTTC	EQGWFNGRFMEFRNAEQFPM
	AGAAACGCCGAACAGTTCCCAATGGATCAGCATATGCTGGGCGCTCTCTGCTGTGATCCTGACA	DQHMLGALCCDPDRYLFIIGSC
	GGTATCTGTTTATAATCGGCTCATGCGAGAGCGAGGACTGGGTAAATGCTCCGTCAGTATGGAT	ESEDWVNAPSVWMAYLGMKH
	GGCTTACCTCGGTATGAAGCACGTATGGGATTATGTCGGTATCAGCGATCATCTTGCTATAAATA	VWDYVGISDHLAINIHKSGHAVI
	TCCATAAATCAGGCCACGCAGTCATAGCAGAGGACATCGAGAAAATGGTACAGTATTTCGACTA	AEDIEKMVQYFDYHVYGIQPKM
	TCACGTATACGGAATACAGCCGAAAATGAACCTTGAAGAACTTCAGACATCTGTATTTGCTCTTC	NLEELQTSVFALPKNKDSFADT
	CAAAGAATAAAGACTCATTCGCAGATACATTTGCTTCAAAGTGGCTTTAC	FASKWLY
	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGGAAAAT	MGSSHHHHHHSSGPQQGLRE
	ATGGGGTCTGACTGTGATGTTAGTGGTTTAGTCGCTGTTGAAGATTTGGCCGTCAACCCGAACT	NMGSDCDVSGLVAVEDLAVNP
	TGCCAGATCCTTTCACCAAGCTGGATGGTAATCGCATGACGGCTAAATCCGAGTGGCGCTGCC	NLPDPFTKLDGNRMTAKSEWR
	GCCGTCAGGAAATACAGCGTGAAGCGGAAGCCTATATGTATGGCTTTAAGCCGCCTAAGCCGG	CRRQEIQREAEAYMYGFKPPK
	AAAAAGTCACCGGTTCGGTCACCTCTGATACGATTACCGTCAACGTTGAGAACGGCGGAAAGAT	PEKVTGSVTSDTITVNVENGGK
	GACGTCATTCACTGCTGAAGTTGATCTCCCGGATGGCGACGGTCCTTTTCCTGTGTTGTTTATG	MTSFTAEVDLPDGDGPFPVLFM
***Tt*GE15C**	TACTCCAGCTTTGCGCACGCAGATCTCGCAAAAGCTGAAGGCGTGGCGGTAGTTAAATACAACC	YSSFAHADLAKAEGVAVVKYNP
	CCTACGATGTTGGTGCAGAGAGCTTTGGTGGTGGTAGCAGAGCGAGCAAGAAAGGCGCGTTTT	YDVGAESFGGGSRASKKGAFY
	ACGATATTTACGGTTCTGACAGCGAAACCGGCCTGCTGGTTGCTTGGGGTTGGGGGGTCAGCC	DIYGSDSETGLLVAWGWGVSR
	GTTTGATCGATGTGATCGAAGCGTCGGATGGTTCGATTTTGCGTGCAGCCGATACTGCGGTTAC	LIDVIEASDGSILRAADTAVTGC
	GGGCTGTTCACGATTTGGTAAAGGCGCCTTCATTGCGGGTGCTTTCGATCAGCGTATCGACTTA	SRFGKGAFIAGAFDQRIDLTIPF
	ACGATTCCATTTGAATCCGGTTCTGGCGGCGTGCCAATTATGCGTGGTCTTCCTGGTGAAGGTG	ESGSGGVPIMRGLPGEGAQSP
	CTCAATCTCCAGGTAGCATCTATGGCGAGCAATACTGGATGGGCGATGCTTTCAAAGATTTCGC	GSIYGEQYWMGDAFKDFATNT

	CACCAACACCAAGGTGAAGAAATTGCCGATCGATACTCATGAAGTCGTTGGTATGGTCGCTCCT	KVKKLPIDTHEVVGMVAPRGLLI
	CGCGGCTTGTTGATTCTGGATAACCCGCATATCGACAACCTTGGGCCACTGTCTGCGCACGTG	LDNPHIDNLGPLSAHVAALGGA
	GCAGCGTTGGGCGGTGCTGAAATTTATAAAGCGCTGGGTGCAGAAGCCAATATTTCGTATCAGT	EIYKALGAEANISYQSNVADGG
	CCAATGTTGCTGATGGCGGTCACTGCTCCGCTCGCCCTGAGTTCGAGCAGCCCTTACGCGACA	HCSARPEFEQPLRDNFAKFLLH
	ATTTTGCGAAGTTCCTGCTGCACACTGGTGATGTACCTGGAGAAATGAATCCACACACCAATAA	TGDVPGEMNPHTNKTGDLSSW
	AACCGGCGACCTGTCTAGCTGGATCGATTGGACGACCCCGACGCTGGCGGAT	IDWTTPTLAD
	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGGCTCCC	MGSSHHHHHHSSGPQQGLRA
	TTAACCGCAGATCAGGATCACGCACAGCTACTTCACGTGTTGGGCATTGAAAACCTGCGCCGC	PLTADQDHAQLLHVLGIENLRR
	GGCGCTGACGGGAATACCGATTCTCCCTTCGCGGCAAATACAGATGAGGCCAAGGCGAATACT	GADGNTDSPFAANTDEAKANT
	GCACTGGACTCTCTGCCACCACTACTTACTAGCGTCAGCGGTCAAGCCATCGCAAGCGCCACC	ALDSLPPLLTSVSGQAIASATD
	GACTGGGAGGCTAACCGGCCAGCATTGCTCAACACTTTTTCTCAGGAGATATACGGCTATGTTC	WEANRPALLNTFSQEIYGYVPG
	CCGGCGGTGCGCCGGAACTGCACTGGAAGGCAGGTTCCACCACACCAATTGACGACAGCGGC	GAPELHWKAGSTTPIDDSGTSA
	ACCAGTGCGATTCGTCAGCATTTCACCAGCACCCTCGTTCACCCAGAGAATGCAGCACTCAATC	IRQHFTSTLVHPENAALNLSLNF
	TCTCGCTAAACTTTACCCTGGTCCTGCCGAAATCTAACAAGCCCGTACCGGTGGTAGTAGTGAT	TLVLPKSNKPVPVVVVMSFDPG
***Tt*GE15A**	GAGCTTCGACCCTGGCATTTGGGAGCGTTTCCGCGATCGTATGCCCGCAGAGCGCTACGCCCA	IWERFRDRMPAERYAQIQADNA
	AATACAGGCGGACAATGCCCGCTGGCGGGAGCAGGTTGTTAATGCTGGCTGGGGCTACGCGG	RWREQVVNAGWGYAEIIPTEF
	AAATTATTCCTACCGAGTTCCAGGCCGACTCGGGCGACGGCTTATCACAAGGTATTATCGGATT	QADSGDGLSQGIIGFVNNGKPR
	TGTTAACAATGGCAAGCCACGCAATCCCACCGACTGGGGTGCATTGAGGGCGTGGGCTTGGAG	NPTDWGALRAWAWSASQVLTY
	TGCATCGCAAGTACTCACTTATTTGCAAACCGATAGCCGGGTTGCAGCAGACAGAATCAGTGTC	LQTDSRVAADRISVHGHSRFGK
	CACGGCCACTCGCGCTTTGGCAAGGCGGCACTGGTCGCCATGGCTTTCGATAATCGATTCGCC	AALVAMAFDNRFAAGFISSSGE
	GCGGGATTTATCAGCTCTTCCGGCGAAGGCGGTGCAAAGCTCTGGCGTCGCAACTTTGGGGAA	GGAKLWRRNFGEQVGNLAGA
	CAAGTGGGTAACCTTGCCGGCGCCGGAGAGTACCACTGGATGGCAGGTAATTTCGTCAAGTAT	GEYHWMAGNFVKYAGPKKVN
	GCAGGCCCGAAGAAAGTTAACGACATTCCCGTCGATGCGCATCAGTTACTCGCGCTCTGTGCT	DIPVDAHQLLALCAPRPVLVSV

	CCACGCCCCGTGCTGGTGAGCGTCGGCAGCCAGGGCGAGAGTTGGGTCGATCCGAAAGGTAT GCTGCTCGCGGCCTACCATGCCACGCCGGCTTACGCGCTGTTCGGCGAACAAGGCGTGACCC AAAACGAGTTACCGGCGGTGGGCAACGGGTTACTTGCAGGTAAACTTGCCTTCAGGCAGCATG AAGGTGGACATACACCTGCGCCCAATTGGGAAACCTTTATCACCTTTGCAACGCGCCAGTGGG CAAGCCCGAACAGGCTCAAG	GSQGESWVDPKGMLLAAYHAT PAYALFGEQGVTQNELPAVGN GLLAGKLAFRQHEGGHTPAPN WETFITFATRQWASPNRLK
***Tt*GE15B**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGTCCAAC GAAAACCCCAGCGCCAGCTGTAATGTAGGTCCGCTGCCCGGAGCAGTATCTGGCACCAACAAC CTGCCTGATCCGTTCCGCAAACTCGACGGCACGCGCATGACGAGCAAGACCGAATGGCGTTGT CGCCGTGCGGAAATCCTCGAGCAGGCCTACACCTATATTTACGGTGAAAAACCCGCCAAACCC AGCAGTGTGACGGGTACTGTCAGCAGCTCCTCGGTGCAGGTCAGCGTGTCGCACAACGGCCG TTCAACCTCGTTCAGCGCGTCTGTCGAACTGCCTTCCACGGGTTCCGCTCCCTACCCCGCAGT CATTGGCTTCGGTGGTGGATTCTTCGGTATCGCAGGTGGCATGAAAGATGTCCTTAAATCTCAG GGCGTTGCGATTATCAACTTCGATCCATACCAAGTGGGTGCAGAAGGCAGCGGACAAGGCAAC GGACGCTTCTACGACATCTATGGCTCAAGCCAGGCGGGTCTGCTTACCGCGTGGGCATGGGG CGCCAGCCGTATTATCGACGTGTTGGAAAGCTCTAGCCTGATCGACGCTGACAGCGTGGGTGT TACCGGGTGTTCGCGCTTCGGTAAAGCCGCCTTTATCGCCGGCGCATTCGACGGCCGTGTCGC TTTGACCATCCCCAACGAGTCTGGTATCGGCGGCGTACCCGCCTTGCGCCTGGTACCCGTGGT CGACAGCGGTGGCGAACAACCTTCGCACGCGATCAACTACGAACCCTGGTTCTCGCCAAATGC TTTCCGCACTTTCGCCAGCAGCCCGAACCGCTTGCCGATTGATACCCATGAGGTTATTGGTTTG GTGGCGCCACGCGGCCTGCTTATTCTGGATAACCCACACATCGGCAATCTCGACTATCGCTCCT CCTACGCTGCAGCCGCAGCCGGTAAGCGCATTTATGAAGCACTGGGTGTGAGCAGCAACATGA GCTTCCACAATAACGTAAGTGACGGCAGCCACTGCGCGTGGAAAAGCGAGTTTGCCCAACCTC TGCGCGATAACGTGTCGAAGTTCCTGAAAGGCGGCAGCGGCAATACAGGCTCTATCAATGCCC	MGSSHHHHHHSSGPQQGLRS NENPSASCNVGPLPGAVSGTN NLPDPFRKLDGTRMTSKTEWR CRRAEILEQAYTYIYGEKPAKPS SVTGTVSSSSVQVSVSHNGRS TSFSASVELPSTGSAPYPAVIGF GGGFFGIAGGMKDVLKSQGVAI INFDPYQVGAEGSGQGNGRFY DIYGSSQAGLLTAWAWGASRII DVLESSSLIDADSVGVTGCSRF GKAAFIAGAFDGRVALTIPNESG IGGVPALRLVPVVDSGGEQPSH AINYEPWFSPNAFRTFASSPNR LPIDTHEVIGLVAPRGLLILDNPH IGNLDYRSSYAAAAAGKRIYEAL GVSSNMSFHNNVSDGSHCAW KSEFAQPLRDNVSKFLKGGSG NTGSINARSGSTVNVDSYINWT

	GCTCTGGTAGCACAGTGAACGTGGACTCCTACATCAACTGGACCACACCAAACCTGTCCGGCA ACCTGGACTTC	TPNLSGNLDF
***Zg*GE15A**	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGGCACAA	MGSSHHHHHHSSGPQQGLRA
	CCCGAAGCAAATTATGACGAAACCAAGGTTCCCAAGTTTACCGTACCGGACCCATTATTGACCT	QPEANYDETKVPKFTVPDPLLT
	TTAGTGGCGAAGCCATTACCGATAGCGACCAATGGGAAGAAAAACGAAAATTGGAAATCTATCA	FSGEAITDSDQWEEKRKLEIYQ
	ATTTTTCGAAAAACAGATTTACGGCAAGGTGCCCGCCATGCTCGACGAGTATTCGTTCGAAGTC	FFEKQIYGKVPAMLDEYSFEVIE
	ATAGAGGAAGACAACAACGCTCTGGGTGGCAAGGCTCAGCGCAAACAAATCGCCGTTGGCCTA	EDNNALGGKAQRKQIAVGLKKN
	AAAAAGAACAACCGTTCGCTCAATTTCAACATCCTCCTCTATCTCCCCAAGGGAAACGAAAACG	NRSLNFNILLYLPKGNENAPVFL
	CTCCCGTATTCCTAGGCTATAATTTTCACGGTAACCACACGGTGACCGACGACCCCAAGGTCAT	GYNFHGNHTVTDDPKVIITKAW
	CATAACCAAAGCCTGGAACGCGAACGACATCGCCTTGAACATCCAGAACAACAAGGCCACCGA	NANDIALNIQNNKATEASRGSR
	AGCCTCTAGGGGCTCTAGAAAAAACCGCTGGGCCATTGACAAAATACTCGACAACGGATTTGGA	KNRWAIDKILDNGFGLATVYYG
	CTAGCGACGGTCTACTACGGTGAAATCGACCCTGATAAAAACGATTTTTCCGATGGATTGCACA	EIDPDKNDFSDGLHSLFYREGQ
	GCCTCTTCTATAGGGAAGGACAAACAAGACCCAAAACAAACGAATGGGGAAGCATTGCCGCAT	TRPKTNEWGSIAAWAFGLSRA
	GGGCCTTTGGCCTTAGCCGTGCCATGGATTACCTCGAAAACGACACCAACATTTCCAAGGTTAT	MDYLENDTNISKVIVFGHSRLG
	TGTTTTTGGCCATTCAAGACTGGGCAAGGCCGCACTTTGGGCCGGGGCCTCCGACCCTCGCTT	KAALWAGASDPRFDGVISNNS
	TGACGGAGTGATCAGCAACAATTCCGGATGCGGTGGCGCGGCCTTATCTAAACGCAAATATGG	GCGGAALSKRKYGETIGHINNS
	GGAAACCATAGGCCACATCAACAATTCCTTTCCGCATTGGTTTTCGGAAAGCTTCAAGAAATACA	FPHWFSESFKKYSNKEEILPVD
	GCAACAAAGAAGAAATATTGCCCGTAGACCAGCACCAACTATTGGCATTGATCGCCCCAAGGCC	QHQLLALIAPRPLYVASAAEDE
	TTTATATGTGGCAAGTGCCGCCGAAGACGAATGGGCCGACCCCAAGGGCGAATTTCTTTCGGC	WADPKGEFLSAQYATPVYALY
	ACAGTACGCCACACCGGTATACGCCCTATATGGCAAAAAAGGAATCTCTAAAAGCGATAGCCCC	GKKGISKSDSPVVDRPIQETLAY
	GTTGTCGACAGGCCTATTCAAGAAACCCTAGCCTACCATATACGATCGGGAAAACACGATGTTA	HIRSGKHDVTDYDWDQYISWA
	CCGATTACGATTGGGACCAATACATCAGTTGGGCGCAAAATTTTGTAAAA	QNFVK

	ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGCAGGAT	MGSSHHHHHHSSGPQQGLRQ
	GGAAAATCTTCTCAGGATAGTATTCATAAACTAAACTATCAGCATTTTATGCAGATGAAACATCAG	DGKSSQDSIHKLNYQHFMQMK
	TTAGGAATCACTGCTAAAAATAGACAGGGGCCTTCTGGAGATCCAACTGATCCAAATGCTGCCA	HQLGITAKNRQGPSGDPTDPN
	ATTTTGATGAAAGTAAAGTAAGATCTTATGTTTTACCTGAAGTATTAGTTAGCAATGAAGGTAAAC	AANFDESKVRSYVLPEVLVSNE
	AAATAAATACTTCCAGGGAATGGTGGGAAATCCGAAGACCCGAAATTGTTGAAGATTTTGAAACT	GKQINTSREWWEIRRPEIVEDF
	TCAATCTATGGTCATTTACCCGATCATATTCCAGATGTAAAATGGAATATTGTTTCTAAAAAAGAT	ETSIYGHLPDHIPDVKWNIVSKK
	ACTCTTATAAACTCTTATCCGGTAAGCGAGCTTCTTTTAGAAGGGCTTGTGGATAATTCTGCCTA	DTLINSYPVSELLLEGLVDNSAY
	TCCGGCAGTTGAGGTGAAAATCGAGCTGTTAGTTGGGATACCTAAAAGCGCAAAAAAAGCAGTG	PAVEVKIELLVGIPKSAKKAVPL
	CCATTAGCTATGGAATTCGGGTTTATAAAATCTCCTTTCATGAGGGCTAATGATGAACCAGACAG	AMEFGFIKSPFMRANDEPDSYF
	CTACTTCTTTTCTTCCTATGAGCCTAAATGGAAACAGCAATTATTATCGCAAAATATGGGGTATG	FSSYEPKWKQQLLSQNMGYAIL
***Zp*GE15A**	CTATATTAGTGCCTAGTAGTATTCAGGCTGATAATGGTGCAGGGCTGACTTCCGGAATCATAGG	VPSSIQADNGAGLTSGIIGLVNH
	ATTGGTAAATCATGGTAAGCAAAGAAAACCCGATGAATGGGGAGTTTTAAGAGCCTGGGCGTGG	GKQRKPDEWGVLRAWAWGAS
	GGCGCAAGTCGAGCCATCGATTATTTTGAAACGAACCCAAAAATAGATGAAAACCGAATTGCAG	RAIDYFETNPKIDENRIAVEGVS
	TAGAAGGCGTATCCAGATATGGTAAAGCTGCCATAGTTAGTATGGCTTTCGACCAAAGAATTTCT	RYGKAAIVSMAFDQRISLGFFG
	CTTGGTTTCTTTGGTTCGGCCGGGGCGGGAGGAATTTCGTTGCTACGTAGAAATTTTGGAGAAC	SAGAGGISLLRRNFGEQVENLA
	AAGTAGAAAATCTTGCAAGCCCAGGTGAGTATCATTGGTTTTGCGGAAATTTTATAAAATATGCC	SPGEYHWFCGNFIKYASVMET
	TCGGTAATGGAAACCCAAGATTTACCTGTAGATGCCCATCAATTAATTGCTCTTTGTGCACCAAG	QDLPVDAHQLIALCAPRPVFISA
	ACCAGTATTTATAAGTGCTGGTTCACATCTTATCGAAGGACAGTGGGTAGACGCCAAAGGAATG	GSHLIEGQWVDAKGMFLSSVY
	TTTTTGTCTAGTGTTTACGCCACACCGGTGTATGAACTGTTAGGGAAAAAAGGACTTGAAACTTC	ATPVYELLGKKGLETSGFPKMG
	AGGATTTCCAAAAATGGGTACCGCTCTAGTAGATGGAGAAATTGCCTTTAGACAGCATGCAGGC	TALVDGEIAFRQHAGGHSTGPN
	GGCCATAGTACCGGCCCAAATTGGAGTACATGGATTGCATGGGCAAGCAAATATTGGCAATTT	WSTWIAWASKYWQF

Supplementary Data – Chapter 7

Table S7.1. Primers used for the PCR amplification of FAEs constructs.

Enzyme name	Primer Fw 5'-3' sequence	Primer Rv 5'-3' sequence
HA146	TCAGCAAGGGCTGAGGGCCGTAAACGATCATCTGCATTTTAC	TCAGCGGAAGCTGAGGTCATAAAATCTGAGGGTATTGAGCATTG
NA143	TCAGCAAGGGCTGAGGAATATAAAAGTCGGCACAAATATTTAC	TCAGCGGAAGCTGAGGTCATTTTATAAAAATCTGCTTAGAGAAATC
PP148	TCAGCAAGGGCTGAGGGTAAACCCGGAGGTAAAGGTGGGAAC	TCAGCGGAAGCTGAGGTCAAATCTTCCATAGATTGGCATAC
NA149	TCAGCAAGGGCTGAGGGCAGATATTTCAGAAAATCAGTTTAAAG	TCAGCGGAAGCTGAGGTCATTTCTGTTTTTTCTGACTTAGTAAC
NA147	TCAGCAAGGGCTGAGGGATGATATGGTAAAGGAAGATTTTCAG	TCAGCGGAAGCTGAGGTCATTTAAATAATAATTGAGCATACTGATAC
IN144	TCAGCAAGGGCTGAGGTTAGTTTTTGTGGCTATCCTGCTGTG	TCAGCGGAAGCTGAGGTCAATTGAATAATAAAGGTGCTAGTTC
NA137	TCAGCAAGGGCTGAGGAAACACGCACCTAATTTTCGCATAG	TCAGCGGAAGCTGAGGTCATTGAACGTCGACATGGAAAACTC
NA139	TCAGCAAGGGCTGAGGGACACCAGCCACAACAAAACCCTG	TCAGCGGAAGCTGAGGTCACTCAGGGTCGCCGCTACCTAAATG
NA138	TCAGCAAGGGCTGAGGAACGCTCAAGTTGCCGCTGGAGTG	TCAGCGGAAGCTGAGGTCAATCCGCAGGTACGTCCAGCCC
NA142	TCAGCAAGGGCTGAGGTCAGAAGTAATAAACATAAAGATTCC	TCAGCGGAAGCTGAGGTCACTTCTGCAAATTCGTCTCAG
IN141	TCAGCAAGGGCTGAGGGCCACCCTCACCCAGGTGTCCG	TCAGCGGAAGCTGAGGTCAATCGCCGTCGAGTTCCCAGAAC
PP182	TCAGCAAGGGCTGAGGCCTGAGCCGTTGCTGCCACG	TCAGCGGAAGCTGAGGTTATGGGCCGACGGCCTCCG
NA180	TCAGCAAGGGCTGAGGCCGCTCGATCCGCAGCTCGC	TCAGCGGAAGCTGAGGTTAATCCGCCCGAACGAACC
PP181	TCAGCAAGGGCTGAGGACGGACGGTTTGGAAAAGCGCAC	TCAGCGGAAGCTGAGGTTATGCCCTGTCGAGGAAGTC
NA184	TCAGCAAGGGCTGAGGTTCGAGCAGTTTGAAAGCAAG	TCAGCGGAAGCTGAGGTTAATCCAGCTTTCCGAAGAAG
IN183	TCAGCAAGGGCTGAGGGTCCAGCGCGGCCGCCGGCTG	TCAGCGGAAGCTGAGGTTAGTCACGGCTAAGAAGGTTC
HA134	TCAGCAAGGGCTGAGGGCAGTGATGAAAATCGAGTATTACTC	TCAGCGGAAGCTGAGGTCAAGTCAGTCTCTCTTCTAATTTG
HA135	TCAGCAAGGGCTGAGGGCAAGTATTGCAATTGAATACCATTC	TCAGCGGAAGCTGAGGTCAAGACAACCTTTCTTCTTTTTGG
NA133	TCAGCAAGGGCTGAGGAATCAATCCTACTTTTATCTAAAAATGAAAG	TCAGCGGAAGCTGAGGTCACCATTTTTCTGCAAAAAATCTCAG
IN131	TCAGCAAGGGCTGAGGGCTTTTTTTCAGATAGAATACAATTC	TCAGCGGAAGCTGAGGTCAGGACAACCGTTCTTCTTTTTG
IN132	TCAGCAAGGGCTGAGGCACGATGGGCAAAATATTTTTTACAG	TCAGCGGAAGCTGAGGTCATCCTTTCAATGTGAATTTCAG
IN128	TCAGCAAGGGCTGAGGGCGTACCAACAAATTGAAAAGTCG	TCAGCGGAAGCTGAGGTCACTGATTGCTTAGTTCGTTTTTCC
HA213	TCAGCAAGGGCTGAGGAACGACGACGTCCCCACCTGG	TCAGCGGAAGCTGAGGTTAACCCCCGTCTCGGACCTC
NA214	TCAGCAAGGGCTGAGGGCACTCGATCCTGCGGTCAAG	TCAGCGGAAGCTGAGGTTAGCCGGCGAACTGCTTGC
PP212	TCAGCAAGGGCTGAGGACGCTCGACCCCGACATGGC	TCAGCGGAAGCTGAGGTTACTCCGGGGTGAGCAACTC

IN211	TCAGCAAGGGCTGAGGACTCGACAGCAGATCTCCAGCAC	TCAGCGGAAGCTGAGGTTAACTCTGCGCGAGGAAGTCAG
HA121	TCAGCAAGGGCTGAGGGTAATAAACACACATCTTTCAATGAAG	TCAGCGGAAGCTGAGGTCAGTTCTTGAAAGCTATGCGGAAG
NA120	TCAGCAAGGGCTGAGGTATAACTACGCATCAAATATGCAGTTC	TCAGCGGAAGCTGAGGTCACTGTCCTTCATGGAAGAAAGAAG
HA113	TCAGCAAGGGCTGAGGGCATCCGAAACACAAGATCAAGATC	TCAGCGGAAGCTGAGGTCACTTAAAGATACGTTGCGAAAATTG
NA111	TCAGCAAGGGCTGAGGACGGATTCCCGGTATTTGAAACGAAC	TCAGCGGAAGCTGAGGTCAGGAGCGATCTAAAAAATACATTAG
LA112	TCAGCAAGGGCTGAGGGCTCTTATTCAATGCCAGTTTTATTC	TCAGCGGAAGCTGAGGTCACCGCCCCTTGAGCGGTAGCC
PP109	TCAGCAAGGGCTGAGGGCTCTTATTCAATGTCAGTTTTATTC	TCAGCGGAAGCTGAGGTCATTGCTTTTTGAGCGGTAGCCAC
NA200	TCAGCAAGGGCTGAGGATTTCCGCCTCTCCGGATGTAC	TCAGCGGAAGCTGAGGTTACTTCCCTGCATAGAAGAGC
NA107	TCAGCAAGGGCTGAGGCAGCTGGACCAGCAATATGCCGC	TCAGCGGAAGCTGAGGTCACTCTTCGTTTACAAATAAATCCG
NA108	TCAGCAAGGGCTGAGGGCTTGGCTGCATATTGAGTATTACTC	TCAGCGGAAGCTGAGGTCATTCCTCCTCCTTGGGCGTGC
NA241	TCAGCAAGGGCTGAGGACCCTCGAAATCGTTTCCAG	TCAGCGGAAGCTGAGGTTACGCCAGCGCTTCGGCGTG
IN243	TCAGCAAGGGCTGAGGAAACGCCTCGCCCATGTGATG	TCAGCGGAAGCTGAGGTTAGCGCAGCAGAAAGGCAAC
IN199	TCAGCAAGGGCTGAGGGAACACCATATCACGATACGCAG	TCAGCGGAAGCTGAGGTTAAATCGTCCAGTCGTTCCAC
IN115	TCAGCAAGGGCTGAGGGCTAAGGTAGATACGGTGCAGAC	TCAGCGGAAGCTGAGGTCATTTTCCCTTCTTATTAAAAAAACC
NA114	TCAGCAAGGGCTGAGGCTACCTGCCGATTTTCCTGTTATG	TCAGCGGAAGCTGAGGTCATTTAAAGAGTTTCTGAACAAATTC
LA116	TCAGCAAGGGCTGAGGCAAACGATAAATGTACCAGCAGATG	TCAGCGGAAGCTGAGGTCACTTGAAAAGATGAGGAATAAACTG
HA104	TCAGCAAGGGCTGAGGCAGGAAAAGCCCGAAGACGTTTC	TCAGCGGAAGCTGAGGTTAGCGGAAGATCCGCTGCGCGAAG
PP189	TCAGCAAGGGCTGAGGACGGGACGAATGAACATCAG	TCAGCGGAAGCTGAGGTTAGGTGGCCAACGTCCGCG
NA240	TCAGCAAGGGCTGAGGACCACCTCACCACCGAGCCC	TCAGCGGAAGCTGAGGTTAGCCCGCGAAGGGCAGGG
HA238	TCAGCAAGGGCTGAGGTCTGTTGCTGACGAGAAACC	TCAGCGGAAGCTGAGGTTAGCTGTGCAGCGCCTCGC
IN239	TCAGCAAGGGCTGAGGATCGGCAAGCATTTCGAACC	TCAGCGGAAGCTGAGGTTATATAACCGCTTTTGCTGTG
PP103	TCAGCAAGGGCTGAGGGGCAAAGTATATGATAATCTCTCGATG	TCAGCGGAAGCTGAGGTCAATATTGATGAAAAGCGTCGGATAC
PP207	TCAGCAAGGGCTGAGGAGCGACTCGAGCACCCGCGAAG	TCAGCGGAAGCTGAGGTTAGTTCTCGGCGAGGGCGC
NA208	TCAGCAAGGGCTGAGGGATGTCAGCATCGATCCGGAG	TCAGCGGAAGCTGAGGTTAACGCGCGGGACGCAACG
NA209	TCAGCAAGGGCTGAGGAATGTCCTGCATCCGCAGGTC	TCAGCGGAAGCTGAGGTTATCTGAGCAGCGCGGCGAG
NA210	TCAGCAAGGGCTGAGGACCTTGCGCATTGACCCCGAG	TCAGCGGAAGCTGAGGTTACAGCTGCTGGAGGCGGC
PP229	TCAGCAAGGGCTGAGGCCCAGCTTGGACCAGACCGC	TCAGCGGAAGCTGAGGTTACCCGTGCAGCGCCTCGC
NA228	TCAGCAAGGGCTGAGGACCGCAAGCACGCTTGCCGAG	TCAGCGGAAGCTGAGGTTATGATGCGTACCCCGATC
NA227	TCAGCAAGGGCTGAGGGGCCTGCAGATCGACCAGGAG	TCAGCGGAAGCTGAGGTTAGATGGTCCGCAGCCGCC
IN230	TCAGCAAGGGCTGAGGTGCATGGCCAGTGTGACTTC	TCAGCGGAAGCTGAGGTTAGCTGGCCGGAGTCGCCTG
NA219	TCAGCAAGGGCTGAGGAGCCTCGACCCGCAGATCGC	TCAGCGGAAGCTGAGGTTAGCGCTGCAATAGCTCGG

NA220	TCAGCAAGGGCTGAGGACCCCGCACGCACTTGATCC	TCAGCGGAAGCTGAGGTTACGGCGATACGGTCGACG
IN221	TCAGCAAGGGCTGAGGAAGGTGAGGCTGCTGGCAACAG	TCAGCGGAAGCTGAGGTTATCGCACGGTCACGGTCAG
IN222	TCAGCAAGGGCTGAGGAAGGTGAGGCTGCTGGCAACAG	TCAGCGGAAGCTGAGGTTATGGCCGCGAACGTCCACGTC
HA160	TCAGCAAGGGCTGAGGCTGATCCAGGTGCCCGGCGG	TCAGCGGAAGCTGAGGTTAACTGGCGGAAGTCGCCGAC
NA152	TCAGCAAGGGCTGAGGACTTCTCCCCCACTGGATCC	TCAGCGGAAGCTGAGGTTACCGCGTTGGGGTGGAATG
IN153	TCAGCAAGGGCTGAGGCGGCTGATGCGAACGCTCCG	TCAGCGGAAGCTGAGGTTAGTCGCGCTCGGGCGGATTTG
HA179	TCAGCAAGGGCTGAGGAACAAACACGTCGCTGCTATG	TCAGCGGAAGCTGAGGTTATTGATGCTTATCTTTGAG
NA174	TCAGCAAGGGCTGAGGAAACGACATGCAATTTATTTCGC	TCAGCGGAAGCTGAGGTTATTTAAACAGTTTTTGCAC
NA176	TCAGCAAGGGCTGAGGAGCTATCAACCGCAAACTGAAG	TCAGCGGAAGCTGAGGTTAAGCCCGTGCGAGGAAATTG
NA178	TCAGCAAGGGCTGAGGATGAGGCTCAATATTGCTCCG	TCAGCGGAAGCTGAGGTTATCGGGTCTCCCTGGCGAC
NA175	TCAGCAAGGGCTGAGGAACGGTTTCTTTTCTTCCGTCG	TCAGCGGAAGCTGAGGTTAACCCTGGCTGGCGATGATC
PP173	TCAGCAAGGGCTGAGGAACAAACACGTCGCTGCTATG	TCAGCGGAAGCTGAGGTTATCTTTGAGATTTTGGTC
IN172	TCAGCAAGGGCTGAGGAACGGTTTTTATTCGTCCGTCG	TCAGCGGAAGCTGAGGTTAACTATGGCTGGCGATAAGC
IN177	TCAGCAAGGGCTGAGGACGGGGTTTCGTGAACAAGGAAG	TCAGCGGAAGCTGAGGTTAATTTGTGCATTCATCGCGC
HA215	TCAGCAAGGGCTGAGGATTCCTAAAATAAATAGTGAAG	TCAGCGGAAGCTGAGGTTATGTATTATCAGGACAAAG
NA216	TCAGCAAGGGCTGAGGGAACTGCTCGAAGAGCACCGCTG	TCAGCGGAAGCTGAGGTTAGGAGAAGAGGTGCCGGG
NA217	TCAGCAAGGGCTGAGGAAACGACATGCTATTTACTTCGC	TCAGCGGAAGCTGAGGTTATTTGAACAACTTCTGAACG
NA218	TCAGCAAGGGCTGAGGATGACGGGTTTTCGTGAACAAG	TCAGCGGAAGCTGAGGTTAATTTGTGCATTCATCGCGC
PP098	TCAGCAAGGGCTGAGGGCAAAAGTAGACACTTTACAAATTGC	TCAGCGGAAGCTGAGGTCATTTCTTAGTAACTACTTGATTTTTAAG
NA101	TCAGCAAGGGCTGAGGGCAAAAATTGACACAATTCAGGTTTTTAG	TCAGCGGAAGCTGAGGTCATTTCTTTTTCTCTTTGAAAAAATTATCAAAG
NA099	TCAGCAAGGGCTGAGGCAGACCGAAACAACAGGGAAAATC	TCAGCGGAAGCTGAGGTCATTTGTTTGTATTTGTTTTTTTCTGC
LA100	TCAGCAAGGGCTGAGGTATATGCTATTTAACGGCAAAGAACC	TCAGCGGAAGCTGAGGTCACTTCTCGATAGTATAGGCAG
NA161	TCAGCAAGGGCTGAGGGAACTCATTGAAAAACATGCCAG	TCAGCGGAAGCTGAGGTTAACGCATATTCAGTTTATTGG
NA164	TCAGCAAGGGCTGAGGGAACTCATTGAAAAACATGCCAG	TCAGCGGAAGCTGAGGTTAACGCATATTCAGTTTATTGG
NA166	TCAGCAAGGGCTGAGGGAAATGCTCGAAGAGCACCGCTG	TCAGCGGAAGCTGAGGTTACTTCAGTAAATACTGCGC
NA169	TCAGCAAGGGCTGAGGAATATAAAAATTGCTGCTTTAAC	TCAGCGGAAGCTGAGGTTATTTGAATAATTTCTGTACG
NA170	TCAGCAAGGGCTGAGGAATATAAAAATTGCTGCTTTAAC	TCAGCGGAAGCTGAGGTTATTTAAATAATTTCTGTACG
NA201	TCAGCAAGGGCTGAGGAACAGTTTTTACTCGCAACATGC	TCAGCGGAAGCTGAGGTTACGCCGCCATACACTCCG
NA165	TCAGCAAGGGCTGAGGACAGAGCAGAAACGACCGGTAC	TCAGCGGAAGCTGAGGTTATTGTTCATCAAGCACGGCC
LA203	TCAGCAAGGGCTGAGGACGGGCTTTCGTGAACAGGGAAG	TCAGCGGAAGCTGAGGTTAATTTGTACATTCATCGCGC

PP204	TCAGCAAGGGCTGAGGACGGGTTTTCGTGAACAGGGAAG	TCAGCGGAAGCTGAGGTTAATTTGTGCATTCATCGCGC
IN162	TCAGCAAGGGCTGAGGGAACTCATTGAAAAACATGTCAG	TCAGCGGAAGCTGAGGTTAACGCATATTCAGTTTATTGG
IN205	TCAGCAAGGGCTGAGGACGGGTTTTCGTGAACAGGGAAG	TCAGCGGAAGCTGAGGTTAATTTGTGCATTCATCGCGC
IN167	TCAGCAAGGGCTGAGGAAAATTACGGATCACAAACTTTC	TCAGCGGAAGCTGAGGTTAATAATGGCATGCGTCAGC
IN202	TCAGCAAGGGCTGAGGAACAGTTTTTACTCGCAGCAGG	TCAGCGGAAGCTGAGGTTACTCCGTCATACACCCGG
HA041	TCAGCAAGGGCTGAGGGCACAGGTGACCAGCGCTATGGCC	TCAGCGGAAGCTGAGGTCATGGCTTCAACTGACGGTCCAG
NA091	TCAGCAAGGGCTGAGGATGATTCTGGTGCCGGGCAGCGTG	TCAGCGGAAGCTGAGGTCACTTGAAGATCTTCTGCACAAATTC
PP039	TCAGCAAGGGCTGAGGTTTAAATATGAGTCGGCCGTGCAATATC	TCAGCGGAAGCTGAGGTCATTCATGGAAGAAATATGGAAGTG
HA040	TCAGCAAGGGCTGAGGGCATCCTTGCCAACCATGCCGCCTTC	TCAGCGGAAGCTGAGGTCAGTTTCCATCCCTCGTCAATC
HA086	TCAGCAAGGGCTGAGGGCAACATCAACAAGTCCAACTATG	TCAGCGGAAGCTGAGGTCACTTACCCACAGTATATCCTG
HA090	TCAGCAAGGGCTGAGGAAGGGAATACACATGTCGGACATTTTG	TCAGCGGAAGCTGAGGTCAAACCCGTCCTGACAGAAATTTTTTC
NA087	TCAGCAAGGGCTGAGGAAGCTTGTACTCACAGATGAAGCAG	TCAGCGGAAGCTGAGGTCAGTTTTTAAACAAATCTGGCAATATG
IN088	TCAGCAAGGGCTGAGGGCATTTTTACAAACAAATTTTTATTCAAAAA CATTAAAAAAAC	TCAGCGGAAGCTGAGGTCATTCGTTAATCCAAAGTATTGC
IN089	TCAGCAAGGGCTGAGGAAAGGGACAATTTCTATAAGAGAAG	TCAGCGGAAGCTGAGGTCATTTTTTACTCATCAAAAACTCTAATG
NA081	TCAGCAAGGGCTGAGGGTACATAACGACTTATCCTTTGATG	TCAGCGGAAGCTGAGGTCATTTGTTATCTACGAATAAATCTG
PP079	TCAGCAAGGGCTGAGGGCAATTGGCAAATTGGTTCTCACTG	TCAGCGGAAGCTGAGGTCATTTAAATAAATCAGGTAAAATGTCATAA ATATAC
PP084	TCAGCAAGGGCTGAGGGCTTCCTTACCAACTACTCCACCTTC	TCAGCGGAAGCTGAGGTCAGAAATCTTCTGCTTCTATTTG
IN038	TCAGCAAGGGCTGAGGGCAGGTGCAGCAGTACCCAGTGCC	TCAGCGGAAGCTGAGGTCAGAATTGTGTGAAGAATTGCCAG
IN072	TCAGCAAGGGCTGAGGGCAGGTGCAGCAGTACCCAGTGCC	TCAGCGGAAGCTGAGGTCAAAACTGTGTAAAAAACTTCCAGG
IN071	TCAGCAAGGGCTGAGGGCAGGTGCAGCAGTACCCAGTGCC	TCAGCGGAAGCTGAGGTCAGAATTGTGTGAAGAATTGCCAG
HA078	TCAGCAAGGGCTGAGGGATAAAACTACTGCCACAATTAGTAC	TCAGCGGAAGCTGAGGTCATTTAAATATTCTCTTAGCAAAATTATAA AGTG
LA074	TCAGCAAGGGCTGAGGGCAGATACGCTGCCAACGACACCAC	TCAGCGGAAGCTGAGGTCATGGTTCAGCAGGAACAGTGTC
NA076	TCAGCAAGGGCTGAGGGCTGCTGATGCTAGCGAGAAAGTTTTC	TCAGCGGAAGCTGAGGTCAACCCAAGGTTTGAATAGGTTC
NA077	TCAGCAAGGGCTGAGGGCTGCCATTGATAATTTGGTACATAAC	TCAGCGGAAGCTGAGGTCATTTGTTGTCCAAGAACAAATCTG
NA075	TCAGCAAGGGCTGAGGGCCACCCAGAAACAAATTGATGCAC	TCAGCGGAAGCTGAGGTCATTTAAATAGATCAGGCAATATATC
NA237	TCAGCAAGGGCTGAGGACTGACACTGTGATTCATACCG	TCAGCGGAAGCTGAGGTTATCGGCGCAGGAACTCGAG
HA235	TCAGCAAGGGCTGAGGAGCGCGCACTTCACCGACGAC	TCAGCGGAAGCTGAGGTTAGCGTCGCGCCCGGAATCGTG
IN236	TCAGCAAGGGCTGAGGCGAGGCTCCATGAACCTGTTAC	TCAGCGGAAGCTGAGGTTACGCGTAACCGCCGTGCTG

NA063	TCAGCAAGGGCTGAGGCAGCAAGGGAAAGTGTATGAGACAC	TCAGCGGAAGCTGAGGTCAATATTGAGTAAATATTCTAGAAACAAATC
LA064	TCAGCAAGGGCTGAGGGCCAAAGTAGATACGTTATTAGTG	TCAGCGGAAGCTGAGGTCATTTCTTGAAGAATTTATCAAAGAAC
NA059	TCAGCAAGGGCTGAGGGCACCGAAAAACGGCACAGTGCAAG	TCAGCGGAAGCTGAGGTCAAAGCATTTTTTTCAGCGCCG
HA067	TCAGCAAGGGCTGAGGCAAGAAACTAAGGCTGACGATCAG	TCAGCGGAAGCTGAGGTCATTTTTCGTTAAATGCGAGCTGG
NA057	TCAGCAAGGGCTGAGGGCTATGAAAACAGGAGTCATTCAG	TCAGCGGAAGCTGAGGTCACAGCATTTTTTTTAAAGCGGGG
NA056	TCAGCAAGGGCTGAGGATACTTGTGCAGAATGGAGTAATGAAC	TCAGCGGAAGCTGAGGTCAGAAACTCCTGCCATTCCGTG
IN196	TCAGCAAGGGCTGAGGGATATCACCACCCATCGAAC	TCAGCGGAAGCTGAGGTTAGACGTGCGCGCCGAACCAG
IN194	TCAGCAAGGGCTGAGGCATCCGCGATGCCGGCCTGG	TCAGCGGAAGCTGAGGTTAGTGCGCATCGACGAAGG
NA062	TCAGCAAGGGCTGAGGGCAAGGGTAGATACGCTGATGGTG	TCAGCGGAAGCTGAGGTCACTCTCCGGCAAAGAACTTGCTG
IN049	TCAGCAAGGGCTGAGGGCGTCGTTGCAGCAGGTCAAG	TCAGCGGAAGCTGAGGTTAGGTGCCGGCCGACTGGC
IN069	TCAGCAAGGGCTGAGGCAAAAGCATGTAGAGATTAAAAAC	TCAGCGGAAGCTGAGGTTAATCCTTTAAAAACTCAACTG
HA206	TCAGCAAGGGCTGAGGGGTTGGCTCGATCGGCTCGATC	TCAGCGGAAGCTGAGGTTAATACAGGGCGCGCCGCAAC
LA186	TCAGCAAGGGCTGAGGACGAAGCTCACCCTCGCCGG	TCAGCGGAAGCTGAGGTTAGCTTCCCGTCACGCGCC
NA145	TCAGCAAGGGCTGAGGCTAGTATTTGCCAATTATGAAGGAG	TCAGCGGAAGCTGAGGTTACTTATTATCATCCCTATATATC
PP198	TCAGCAAGGGCTGAGGGCCGTTCAATCGTTTTCGTTTTTC	TCAGCGGAAGCTGAGGTTATCCGATCTCAGCCGGCG
NA054	TCAGCAAGGGCTGAGGGCGTTGACCAGGGTGACCGG	TCAGCGGAAGCTGAGGTTAGTGGTCGAGGCCGAGGAAC
IN225	TCAGCAAGGGCTGAGGCGCACAGCACTGCTGTCGGTG	TCAGCGGAAGCTGAGGTTAGGGCAGCACGTGGTTGC
IN234	TCAGCAAGGGCTGAGGTCAATTCATAGCAATAAC	TCAGCGGAAGCTGAGGTTAAACATTGTGCCAACCTCCC
IN231	TCAGCAAGGGCTGAGGGCACTGATTCGTATTAATTTTATG	TCAGCGGAAGCTGAGGTTAATTAGCTTTAATATGCCCAG
IN232	TCAGCAAGGGCTGAGGACATCGATGGAATTTAAG	TCAGCGGAAGCTGAGGTTATTTAAACGCGGCCAGTGC
HA085	TCAGCAAGGGCTGAGGGAAACAAATTTGAAAGATGCTGG	TCAGCGGAAGCTGAGGTTATTTAAAAATCCTCTTTGC
PP052	TCAGCAAGGGCTGAGGGCCACGTTGACCCGGGTCAC	TCAGCGGAAGCTGAGGTTAACCGCTGTTCAAGCCCAG
NA073	TCAGCAAGGGCTGAGGTCTTCAAATGCAAGTTTAAAC	TCAGCGGAAGCTGAGGTTAAGCTTGATTTGCTGGAAATG
HA046	TCAGCAAGGGCTGAGGGCCATACAACCAACTATACCG	TCAGCGGAAGCTGAGGTTACTGAGATGTATACCCAACG
PP157	TCAGCAAGGGCTGAGGGACAACCCTGCCGACGAGAAAC	TCAGCGGAAGCTGAGGTTAGCCCTGCAGGGCGGTTC
IN119	TCAGCAAGGGCTGAGGCGTACGCTGATCCGCCTGGTTC	TCAGCGGAAGCTGAGGTTAGCGCGGCAACTGGCTGAAG
LA061	TCAGCAAGGGCTGAGGTGCGCATTGGCGTATTGCTTC	TCAGCGGAAGCTGAGGTTATTTGTTGAATAACAATTGTG

Table S7.2. DNA and amino acid sequences, predicted molecular mass, molecular architecture and origin of FAEs studied in this work.

(Data available in a supplementary excel file)