UNIVERSIDADE DE LISBOA

Faculdade de Medicina Veterinária

Discovering novel Carbohydrate-Active Enzymes in the cellulosome of anaerobic bacteria

Vânia Ondina Pedro Fernandes

CONSTITUIÇÃO DO JÚRI

PRESIDENTE:

Reitor da Universidade de Lisboa

VOGAIS:

Doutor José António Mestre Prates

Doutor Manuel D'Orey Cancela D'Abreu

Doutor Carlos Mendes Godinho de Andrade Fontes

Doutora Ana Luísa Moreira de Carvalho

Doutora Teresa Paula da Costa Ribeiro

ORIENTADORES

Doutor Carlos Mendes Godinho de Andrade Fontes

Doutor Luís Manuel dos Anjos Ferreira

ORIENTADOR EMPRESARIAL

Doutora Catarina Isabel Proença Duarte Guerreiro

2015

LISBOA

UNIVERSIDADE DE LISBOA

Faculdade de Medicina Veterinária

Discovering novel Carbohydrate-Active Enzymes in the cellulosome of anaerobic bacteria

Vânia Ondina Pedro Fernandes

TESE DE DOUTORAMENTO EM AMBIENTE EMPRESARIAL

(NZYTech, Lda – Genes & Enzymes)

DOUTORAMENTO EM CIÊNCIAS VETERINÁRIAS

ESPECIALIDADE EM CIÊNCIAS BIOLÓGICAS E BIOMÉDICAS

CONSTITUIÇÃO DO JÚRI

PRESIDENTE:

Reitor da Universidade de Lisboa

VOGAIS:

Doutor José António Mestre Prates

Doutor Manuel D'Orey Cancela D'Abreu

Doutor Carlos Mendes Godinho de Andrade Fontes

Doutora Ana Luísa Moreira de Carvalho

Doutora Teresa Paula da Costa Ribeiro

ORIENTADORES

Doutor Carlos Mendes Godinho de Andrade Fontes

Doutor Luís Manuel dos Anjos Ferreira

ORIENTADOR EMPRESARIAL

Doutora Catarina Isabel Proença Duarte Guerreiro

2015

LISBOA

Partes desta tese **foram omitidas** devido à confidencialidade das informações.

Portions of this thesis have been omitted to preserve confidentiality

À minha Família

Agradecimentos

Olhando para trás, a partir do momento em que agora me encontro e reflectindo sobre o que foi esta caminhada de quatro anos de muito trabalho, entrega, dedicação e também de sofrimento, posso dizer que foi um tempo bom. Aprendi e assimilei muitos conhecimentos, valores, experiências e maturidade. Adquiri uma força de vontade tão grande para fazer as coisas o melhor possível e conseguir ultrapassar todas as metas que me foram impostas, incluindo as que eu coloco a mim mesma. Porém, como em tudo na vida, houve também as partes más; a angústia em conseguir resultados, a incerteza e a dúvida sobre os melhores caminhos a seguir, a dura tarefa e disciplina em conseguir conciliar o trabalho de doutoramento com as tarefas e responsabilidades que tenho na empresa que me acolheu. O sofrimento em saber se era capaz… Conhecendo-me como só eu me conheço e tendo à partida uma certa "resistência" em fazer um doutoramento de âmbito académico, não foi fácil a caminhada para chegar até aqui, em seguir uma via mais prática, mais voltada para a realidade do trabalho, das empresas e do mercado. Mas felizmente, muitas foram as pessoas que me ajudaram em todo este processo e aventura que é fazer, não só um doutoramento, mas um doutoramento em ambiente empresarial, onde os clientes são prioridade, as esperas não podem ser prolongadas, o tempo passa rápido... A todas essas pessoas e entidades quero agradecer agora. Assim, o meu mais sincero obrigado:

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

À NZYTech, por me ter motivado a fazer o doutoramento em empresa, por me ter acolhido e financiado parte da minha bolsa. Pelos meios físicos e materiais que me disponibilizou para a realização de todo o trabalho. E por me encher de orgulho todos os dias por fazer parte da equipa;

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 ser mais que um orientador. Por ser meu amigo, meu psicólogo e confidente durante todo este percurso. Por ser, para além de um investigador excepcional e profissional dedicado a 100%, um excelente ser humano. Obrigada por me fazer crescer e por acreditar em mim. Obrigada pela paciência;

Ao Professor Doutor Luís Ferreira, pela sua orientação e pelas palavras sábias no momento certo. Acima de tudo, pela amizade, pelos conselhos e o incentivo com que sempre pude contar e pela sua sensibilidade tocante;

Ao Professor Doutor José Prates, pelos ensinamentos que me transmitiu, pelo sentido de organização e capacidade de trabalho e pela sua disponibilidade e simpatia constantes;

À Nádia Rodrigues, por todo apoio, carinho e dedicação que teve sempre para comigo. Pelo ombro amigo e conselheiro. Por transmitir boa energia e entusiasmo em todos os momentos, principalmente nas horas más. Obrigada por teres estado comigo nessas alturas;

À Joana Brás e à Teresa Ribeiro, por terem sido os meus suportes importantes no início e em muitos outros momentos. Pelo apoio em todas as tarefas laboratoriais, pelos ensinamentos, pelos valores de trabalho e humanos, pela troca de ideias e boa disposição;

À Catarina Guerreiro, à Patrícia Ponte, à Filipa Sequeira e ao André Fernandes, por terem partilhado grande parte do percurso da NZYTech, por termos aprendido juntos muitas coisas sobre “como funciona uma empresa”, pelo apoio incondicional, pela troca de ideias e experiências, pela boa disposição, pelo incentivo, pelo companheirismo. Em particular, à Filipa Sequeira, por partilhar comigo todas as descobertas, frustrações, obstáculos e aventuras do doutoramento em empresa;

À Andreia Peixoto, ao Diogo Comprido, ao Pedro Rio, ao João Belchior e à Vânia Cardoso pela vossa capacidade de trabalho, por terem trazido um novo “fôlego” à NZYTech, pelo espírito de entreajuda e companheirismo. Pela boa disposição, sentido de humor e sorriso sempre presentes;

À Helena Santos, ao Fernando Vaz Dias, à Virgínia Pires, à Maria José Centeno, à Márcia Correia e à Benedita Pinheiro pela ajuda incondicional, pelo espírito de entreajuda que me transmitiram, pela capacidade que tiveram em fazer-me adaptar tão rápido e bem ao laboratório de Nutrição Animal da Faculdade de Medicina Veterinária. Pelas gargalhadas que trocámos, pelas palavras, pelo incentivo, e pela amizade;

À Ana Sofia Luís, pela amizade, pelo apoio sempre presente, por partilhar comigo parte do trabalho. Pela motivação e força de vontade que transmite;

À Immacolata Venditto, ao Pedro Bule, à Kate Cameron, ao Shabir Najmudin, à Susana Martins, à Sofia Costa, à Marta Madeira e à D. Paula pela simpatia, boa disposição e boa energia que transmitem e pelo apoio prestado sempre que necessário. In particular, Imma, for the important support in the work, as well as in our travels, especially, with the English;

To Dr. Harry Gilbert and Dr. Arun Goyal for collaborating with our group and for being such inspiring scientists;

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

Ao Bruno, pela companhia durante a fase de escrita e pela motivação. Pela distração às vezes necessária. Pela amizade e carinho;

À avó Alice, pela companhia, por me fazer sorrir e pela motivação;
Aos meus pais, Jorge e Maria Inocência, pelo apoio constante, pela motivação, pelo amor e carinho. Por acreditarem em mim. Pela paciência. Agradeço-vos muito tudo o que já fizeram por mim. São o meu porto de abrigo;

À Flávia, por tudo. Por estar sempre comigo mesmo estando longe. Pela motivação e pelo apoio. Por me fazer acreditar nas minhas capacidades. Por tudo o que partilhamos e tudo o que vivemos juntas.

This work was co-funded by Fundação para a Ciência e a Tecnologia, grant SFRH/BDE/51101/2010 from Ministério da Ciência, Tecnologia e Ensino Superior, NZYTech genes & enzymes and by QREN SI IDT projeto nº 30263 – HTProteins.

RESUMO

Descoberta de novas enzimas celulossomais de bactérias anaeróbias que degradam hidratos de carbono

As enzimas que na natureza degradam os hidratos de carbono (CAZymes) são capazes de construir, quebrar ou modificar ligações glicosídicas. Estas enzimas actuam sobre polissacáridos complexos e recalcitrantes, como a celulose e a hemicelulose, e apresentam geralmente uma estrutura modular, podendo incluir módulos catalíticos fundidos através de sequências de ligação a domínios não catalíticos, sendo os mais comuns os módulos de ligação a hidratos de carbono (CBMs). Em algumas bactérias anaeróbias, estas enzimas podem associar-se em complexos multi-enzimáticos de elevada massa molecular designados de celulossomas. Os organismos que produzem estes complexos apresentam um vasto repertório de enzimas envolvidas na degradação da parede celular vegetal e constituem um bom ponto de partida para a descoberta de novas CAZymes. Actualmente, verifica-se uma crescente acumulação de informação genómica e metagenómica a um ritmo superior à capacidade de identificação da função biológica de uma plêiade de genes e proteínas de funções desconhecidas. Para além disso, para a maioria das CAZymes já conhecidas, não foi ainda efectuada uma caracterização estrutural e/ou bioquímica. Neste estudo foram desenvolvidas metodologias inovadoras para a descoberta de novas CAZymes em bactérias celulossomais, bem como se procedeu a uma caracterização bioquímica detalhada para algumas destas enzimas. Desenvolveu-se uma plataforma de alta capacidade para a clonagem, expressão e produção de proteínas celulossomais recombinantes em *Escherichia coli*, tendo como objectivo descobrir novas CAZymes codificadas nos genomas de *Clostridium thermocellum* e *Ruminococcus flavefaciens*. Como resultado, foi construída uma nova série de vectores de expressão (pHTP) a fim de sustentarem um método de clonagem independente de ligação. Para possibilitar a total automatização do processo foram desenvolvidos novos meios de cultura celulares e métodos de purificação de proteínas adaptados a um esquema de produção de alta capacidade. A pesquisa de novas enzimas nos módulos celulossomais de função desconhecida possibilitou a descoberta de uma nova α-L-arabinofuranosidase em *R. flavefaciens*, que se constitui como a enzima fundadora de uma nova família de CAZymes. A fim de potenciar a solubilidade de proteínas recombinantes em *E. coli*, foram desenhadas novas *tags* de fusão, as quais foram incorporadas em vectores derivados do pHTP. Tanto as *tags* Rf1 como Rf47, derivadas de componentes celulossomais, mostraram possuir uma capacidade elevada para potenciar a solubilidade de proteínas, uma vez que as proteínas de fusão contendo quer uma quer outra destas *tags* foram produzidas na forma solúvel em níveis mais elevados do que com parceiros de fusão anteriormente descritos. Confirmou-se que os CBMs afectam a actividade catalítica das CAZymes associadas, tal como ilustrado pelo CBM32 da *Ct*Man5A. Este trabalho forneceu indicações de que os CBMs membros da família 35 têm a capacidade de se ligarem a polímeros de β-manose. A caracterização bioquímica das PL1A, PL1B e PL9 aqui descrita constituiu o primeiro relato de actividade pectinolítica no celulossoma de *C. thermocellum*. Estas enzimas podem estar associadas a CBMs que revelam pouca especificidade de ligação aos substratos. Testou-se a aplicação de β-glucanases na suplementação alimentar animal, tanto como enzimas isoladas, como associadas em mini-celulossomas. Os dados apresentados aqui revelam que são as β-1,3-1,4-glucanases e não as β-1,4-glucanases as enzimas responsáveis por melhorar o valor nutritivo de dietas à base de cevada para frangos. Por outro lado, os resultados mostram que a eficácia dos mini-celulossomas para melhorar o desempenho das enzimas exógenas usadas na suplementação alimentar requer um mecanismo eficaz para proteger as regiões de ligação entre os componentes celulossomais da degradação por proteases.

Palavras-chave: Enzimas degradativas de hidratos de carbono, celulossomas, técnicas de alta capacidade, expressão de proteínas recombinantes, suplementação alimentar animal

ABSTRACT

Discovering novel Carbohydrate-Active Enzymes in the cellulosome of anaerobic bacteria

Carbohydrate-active enzymes (CAZymes) include a range of enzymes that, in nature, make, break or modify glycosidic bonds. CAZymes act on highly recalcitrant polysaccharides, such as cellulose and hemicellulose, and often exhibit a modular architecture including catalytic domains fused through flexible linker regions to non-catalytic domains such as carbohydrate-binding modules (CBMs). In some anaerobic bacteria these enzymes can associate in high molecular mass multi-enzyme complexes termed cellulosomes. Cellulosomal organisms express a vast repertoire of plant cell wall degrading enzymes and constitute a promising source for the discovery of novel CAZymes. Presently, an exponential accumulation of genomic and metagenomic information is observed while the identification of the biological role of both genes and proteins of unknown function is sorely lacking. In addition, for most of the known CAZymes, structure and/or biochemical characterization is missing. In this study we have developed innovative approaches for the discovery of novel CAZymes in cellulosomal bacteria and provide a detailed biochemical characterization of some of those enzymes. A high-throughput platform was designed for cloning, expression and production of recombinant cellulosomal proteins in *Escherichia coli*, aiming at looking for novel cellulosomal CAZymes encoded in the genomes of *Clostridium thermocellum* and *Ruminococcus flavefaciens*. As a result, a series of novel prokaryotic expression vectors (pHTP) were constructed to allow ligation-independent cloning with high levels of soluble recombinant protein production. In addition, to allow total automation of the procedure, both novel cell culture media and protein purification methods have been established. The platform allowed the production of 184 cellulosomal proteins of unknown function that after the implementation of an enzyme discovery screen lead to the discovery of a novel family of α-L-arabinofuranosidases. In order to achieve recombinant soluble expression in *E. coli*, novel fusion tags were designed and incorporated into pHTP-derivatives. Both Rf1 and Rf47 tags, derived from cellulosomal components, were shown to display a high capacity to enhance protein solubility, as fusion proteins containing both these tags were expressed at high levels and in the soluble form in *E. coli*. CBMs were confirmed to affect the catalytic activity of appended CAZymes, as it was illustrated by the CBM32 of *Ct*Man5A. This work revealed that members of family 35 CBM have the capacity to bind β-*mannose*-containing polymers. The biochemical characterization of PL1A, PL1B and PL9 reported here describes the pectinolytic activity expressed by *C. thermocellum* cellulosome. These enzymes are appended to CBMs that display considerable ligand promiscuity. The application of β-glucanases in animal feed supplementation was tested either in the free state or while associated in mini-cellulosomes. This study revealed that β-1,3-1,4-glucanases and not β-1,4-glucanases are necessary to improve the nutritive value of barley-based diets for broilers. In addition, it was shown that mini-cellulosomes designed to improve the efficacy of exogenous enzymes used for feed supplementation require an effective mechanism to protect linker regions from proteolytic cleavage.

Key-words: Carbohydrate-active enzymes, cellulosomes, high-throughput techniques, recombinant protein expression, animal feed supplementation

This thesis was based on the following manuscripts:

I. **Fernandes, V.O.**, Luís, A.S., Gilbert, H.J. Brás, J.L.A., Guerreiro, C.I.P.D., Ferreira, L.M.A., Fontes, C.M.G.A. A high-throughput platform for gene cloning, expression and purification of recombinant proteins in *Escherichia coli*: discovery of novel enzymes in cellulosomes. (Manuscript in preparation).

II. **Fernandes, V.O.**, Sequeira, A.F., Brás, J.L.A., Guerreiro, C.I.P.D., Ferreira, L.M.A., Fontes, C.M.G.A. Identification of novel solubility tags for recombinant protein expression in *Escherichia coli*. (Manuscript in preparation).

III. **Fernandes, V.O***, Mizutani, K.*, Karita, S., Luís, A.S., Sakka, M., Kimura, T., Jackson A., Zhang, X., Fontes C.M.G.A., Gilbert, H., Sakka, K. 2012. Influence of a mannan binding family 32 carbohydrate binding module on the activity of the appended mannanase. *Appl. Environ. Microbiol*. 78(14):4781-4787.

IV. **Fernandes, V.O.***, Chakraborty, S.*, Dias, F.M.V., Prates, J.A.M., Ferreira, L.M.A., Fontes, C.M.G.A., Goyal, A., Centeno, M.S.J. 2015. Role of pectinolytic enzymes identified in *Clostridium thermocellum* cellulosome. *PLOS one*. 10(2):e0116787.

V. **Fernandes, V.O.***, Costa, M.*, Ribeiro, T., Serrano, L., Cardoso, V., Santos, H., Lordelo, M., Ferreira, L.M., Fontes, C.M.G.A. β-1,3-1,4-glucanases and not β-1,4-glucanases improve the nutritive value of barley-based diets for broilers. *PLOS one*. (Submitted).

VI. **Fernandes, V.O.***, Costa, M.*, Ribeiro, T., Serrano, L., Cardoso, V., Lordelo, M., Ferreira., L. M., Fontes C.M. 2014. Construction of GH16 β-glucanase mini-cellulosomes to improve the nutritive value of barley-based diets for broilers. *J. Agric. Food Chem*. 62(30):7496-7506.

* Authors contributed equally to the work

The Author's contribution:

Publication I: Performed all molecular biology and biochemical experiments. Constructed the pHTP expression vector. Wrote the manuscript.

Publication II: Performed all molecular biology experiments. Constructed some of the expression pHTP-derivative vectors. Wrote the manuscript.

Publication III: Was involved in cloning, recombinant protein expression and purification and analysis of CBM and enzyme properties.

Publication IV: Was involved in cloning, recombinant protein expression and purification of recombinant proteins. Performed the enzyme assays for PL1A.

Publication V: Performed all experiments designed to evaluate the biochemical characterization of the enzymes. Major writing contributions.

Publication VI: Was involved in the expression and purification of *Ct*Glc16A mini-cellulosomes. Performed all *in vitro* experiments. Major writing contributions.

INDEX

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

%	Percentage
ΔG	Gibbs Energy
ΔH	Entalphy change of a system
ΔS	Entropy change of a system
A	Absorbance
AA	Auxiliary Activities
aa	Amino acid
AME	Apparent Metabolizable Energy
ANOVA	Analysis of variance
Ara	L-Arabinose
ATG	Start codon
ATP	Adenosine Triphosphate
B. circulans	*Bacillus circulans*
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
bp	base-pair
BPC	Consolidated bioprocessing
BSA	Bovine serum albumin
BW	Body Weight
C. cellulolyticum	*Clostridium cellulolyticum*
C. cellulovorans	*Clostridium cellulovorans*
C. japonicus	*Clostridium japonicus*
C. josui	*Clostridium josui*
C. lentocellum	*Clostridium lentocellum*
C. papyrosolvens	*Clostridium papyrosolvens*
C. termitidis	*Clostridium termitidis*
C. thermocellum	*Clostridium thermocellum*
$CaCl_2$	Calcium chloride
cAMP	Cyclic adenosine monophosphate
CAP	Catabolite activator protein
CAZyme	Carbohydrate-Active enZyme
CBD	Cellulose-binding domain
CbhA	*C. thermocellum* cellobiohydrolase
CBM	Carbohydrate-Binding Module
CBM32	Family 32 Carbohydrate-Binding Module
CbpA	*C. cellulovorans* scaffolding protein
CcpA	Catabolite control protein
CE	Carbohydrate Esterase
Cel5	Celullase from family 5
CelA	*C. thermocellum* endoglucanase A
celC	*C. thermocellum* gene enconding endoglucanase C
CelE	*C. thermocellum* cellulase E
CelJ	*C. thermocellum* endoglucanase J
CelK	*C. thermocellum* cellobiohydrolase K
CelR	*C. thermocellum* cellulase R

CelS	*C. thermocellum* cellulosomal cellobiohydrolase S
cfu	Colony forming unit
CipA	*C. thermocellum* Cellulosome integrating protein
cipA	*C. thermocellum* cellulosomal gene
cm	Centimetre
CO_2	Carbon dioxide
coh	Cohesin
cP	Centipoise
***Ct*Cel8A**	*C. thermocellum* Cellulase 8A
***Ct*Cel8A**	*C. thermocellum* β-1,4-glucanase 8A
***Ct*Glc16A**	*C. thermocellum* β-1,3-1,4-glucanase 16A
***Ct*Lic26A-Cel5E**	*Clostridium thermocellum* lichanase family 26 A joined to a cellulose from family 5 E
***Ct*Man5**	*C. thermocellum* Mannanase from family 5
CttA	*R. flavefaciens* cellulosomal protein
cttA	*R. flavefaciens* cellulosomal gene
DNA	Desoxyribonucleic Acid
DnaJ	Chaperone protein
DnaK	Chaperone protein
DNSA	3,5-dinitrosalicylic acid
Doc	Dockerin
DsbA	Disulfide Isomerase A
DsbC	Disulfide Isomerase C
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediamine tetraacetic acid
EngB	Endoglucanase B
FAE	Ferulic Acid Esterase
FCR	Feed Conversion Ratio
Fh8	Small protease produced by *Fasciola hepatica*
Fuc	L-Fucose
Gal	D-Galactose
GalA	D-Galacturonic Acid
GalNAc	*N*-Acetylgalactosamine
Gb1	Gb1-domain from protein G
GH	Glycoside Hydrolase
GH16	Family 16 Glycoside Hydrolase
GH8	Family 8 Glycoside Hydrolase
GI	Gastrointestinal tract
Glc	D-Glucose
GlcA	D-Glucuronic Acid
GlyR3	*C. thermocellum* carbon catabolite repressor
GRAVY	Grand Average of Hydropathy
GroEL	Chaperone protein EL
GroES	Chaperone protein ES
GrpE	Chaperone protein E
GST	Glutathione S-transferase
GT	Glycosyl Transferase
h	Hour
H	Fusion tag

HCl	Hydrogen chloride
HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl] ethanesulfonic acid
HG	Homogalacturonan
HGP	Human Genome Project
His	Histidine (H)
His_{10}	Deca-histidine tag
His_6	Hexa-histidine tag
HTP	High-throughput
IB	Inclusion Bodies
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
K_2HPO_4	Dipotassium phosphate
K_a	Association constant
Kan	Kanamycin
kb	kilobase
K_{cat}	Catalytic constant
kDa	kiloDalton
Kg	Kilogram
kJ	kilojoule
K_m	Michaelis constant
L	Litre
lac	Lactose gene
LacI	Lactose repressor
LacNAc	*N*-Acetyl-D-Lactosamine
lacO	Lactose operator gene
LB	Luria Bertani
LIC	Ligation-independent cloning
LPMO	Lytic polysaccharide mono-oxygenases
M	Molar
Man	D-Mannose
MBP	Maltose Binding Portein
MCS	Multiple Cloning Site
MES	2-(*N*-morpholino)ethanesulfonic acid
mg	Milligram
min	Minutes
mL	Milliliter
mM	Millimolar
mol	Molecular replacement
mRNA	Messenger Ribonucleic Acid
MsyB	Acidic protein MsyB
MW	Molecular Weight
N	Nitrogen
Na_2HPO_4	Disodium hydrogen phosphate
NaCl	Sodium chloride
$NaHCO_3$	Sodium bicarbonate
NaOH	Sodium hydroxide
NC	Negative control
NGS	Next generation sequencing
Ni^{2+}	Nickel ion

nm	Nanometer
NRC	National Research Council
NSP	Non-starch polysaccharide
nt	Nucleotide
NusA	N-utilization substance protein A
ºC	Celcius degree
OD	Optical Density
OlpA	*C. thermocellum* cell surface protein
olpA	*C. thermocellum* gene enconding cell surface protein
OlpB	*C. thermocellum* anchoring scaffoldin
olpB	*C. thermocellum* gene enconding anchoring scaffoldin
OlpC	*C. thermocellum* cell surface protein
ORF	Open Reading Frame
Orf2	*C. thermocellum* anchoring scaffoldin
orf2	*C. thermocellum* gene enconding anchoring scaffoldin
P	Phosphorus
PA	Pectin from apple
PAGE	Polyacrylamide Gel Electrophoresis
PC	Pectin from citrus fruits
PC buffer	Phosphate/citrate buffer
PC group	Positive control group
PCR	Polymerase Chain Reaction
PGA	Polygalacturonic Acid
PGL	Pectic galactan from lupin
PGP	Pectic galactan from potato
pH	negative decimal logarithm of the hydrogen ion activity in a solution
p*I*	Isoelectric point
PL	Polysaccharide Lyase
pLysE	Plasmid for expressing T7 lysozyme
pLysS	Plasmid for expressing T7 lysozyme
p*NPA*f	4-nitrophenyl-*α*-arabinofuranoside
R	Universal gas constant
R. flavefaciens	*Ruminococcus flavefaciens*
RBS	Ribosomal Binding Site
RGAP	Rhamnogalacturonan from potato
RGAS	Rhamnogalacturonan from soybean
RG-I	Rahmnogalacturonan-I
RG-II	Rahmnogalacturonan-II
SAS	Statistical Analysis Software
ScaA	Anchoring scaffoldin A
scaA	Anchoring scaffoldin A gene
ScaB	Anchoring scaffoldin B
scab	Anchoring scaffoldin B gene
ScaC	Anchoring scaffoldin C
scaC	Anchoring scaffoldinC gene
ScaE	Anchoring scaffoldin E
scaE	Anchoring scaffoldin E gene
SCFA	Short Chain Fatty Acids
SdbA	*C. thermocellum* anchoring scaffoldin A

sdbA	*C. thermocellum* gene enconding anchoring scaffoldin A
SDS	Sodium Dodecyl Sulfate
SLH	S-layer-homologous
SLIC	Sequence- and ligation-independent cloning
SSCF	Simultaneous saccharification and co-fermentation
SUMO	Small ubiquitin-like modifier
T	Absolute temperature
TB	Terrific Broth
Tev	Tobacco etch virus
TF	Trigger factor
TLC	Thin Layer Chromatography
Tm	Melting temperature
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
tRNA	Transfer RNA
Trx	Thioredoxin
U	Enzymatic unities
Ub	Ubiquitin
w/v	Weight per volume
w/w	Weight per weight
WGS	Whole-genome shotgun
Xyl	D-Xylose
XynC	Glucuronoxylanase XynC
XynZ	Endo-1,4-β-xylanase Z
ZZ	Double Z-domain from Staphylococcal protein A
µg	Microgram
µL	Microliter
µM	Micromolar

1. INTRODUCTION AND THESIS OUTLINE

In the post-genomic era there is an urgent need to assign biological functions to proteins of unknown role encoded by sequenced genomes and metagenomes. Even for annotated proteins, structural and biochemical characterization would allow a more insightful understanding of their biological roles in nature. This knowledge is obviously of fundamental importance to science but could also uncover a large range of biotechnological applications of great importance to our societies, since they may present remarkable impacts in the industry, agriculture, and medicine sectors. Plant cell walls represent the most abundant and renewable source of carbon available in nature and their biotechnological utilization remains elusive due to the recalcitrance of their major constituents, cellulose and hemicellulose. Recently, as a consequence of pressures for an increasingly green society, aiming to develop sustainable approaches to degrade plant biomass for the generation of biofuels, there is an exponential increase in the identification of carbohydrate active-enzymes (CAZymes). These enzymes are involved in plant cell wall deconstruction by breaking or modifying glycosidic bonds in a multitude of structural polysaccharides. Besides their use in the production of renewable fuels through the production of fermentable sugars from plant biomass, CAZymes can also be applied in other industrial and agricultural processes, such as animal feed supplementation. Addition of exogenous enzymes in dietary preparations for simple-stomach animals enhances the nutrient availability through the degradation of anti-nutritive soluble polysaccharides in the gastrointestinal tract. The major impacts of adding exogenous enzymes to monogastric animal diets relate with the significant reduction in digesta viscosity associated with enzyme activity. Today, genetic engineering allows the development of appropriate tailor-made strategies to design more efficient enzymes which could improve the cost-effectiveness of converting biomass to fuels and also improve feed nutritive value. The remarkable gene diversity and complexity found in genomes of the microbial flora colonizing ecosystems capable of deconstructing polysaccharides provides an ideal resource for mining novel CAZymes. In particular, genomes of organisms containing cellulosomes, which are highly efficient nanomachines involved in the deconstruction of cellulose and hemicelluloses, usually comprehend a large array of potentially effective plant cell wall-degrading enzymes. The optimal utilization of genomic sequence data requires, however, the development of high-throughput methods for gene cloning, recombinant protein expression and protein purification that could support the rapid characterization of proteins, including their structure determination. In this respect, several high-throughput strategies for parallel cloning and expression of large numbers of genes are being applied in laboratories worldwide. Ligation-independent cloning methods allow efficient cloning of multiple targets simultaneously, while miniaturization of cell growth conditions and automation of the expression and purification steps accelerate the process of recombinant protein production.

In addition, several approaches are being designed for the efficient production of soluble recombinant proteins in economic and simple bacterial expression systems, which usually involve the application of protein fusion technologies.

The focus of this thesis was the development of novel methods for the discovery of novel CAZymes in cellulosomal bacteria, which will permit the isolation and biochemical characterisation of several unkown enzymes. The aim of this project was not only to provide highly efficient biocatalysts that could benefit many biotechnological applications, but also to elucidate several questions concerning the function and properties of cellulosomal enzymes. The work described here was developed with the aim of establishing fundamental research that could support the development of innovative applied science. Thus, what was envisaged in this project was a link over the fundamental work that has been developed at the University and an applied perspective given by a Company, which has in its genetic make-up the drive for exploring the applications driven from scientific knowledge. Besides the integration of the novel CAZymes described here in the NZYTech's market portfolio, many other contributions of this project to improve the competitiveness of the Company can be reported, being the development of a completely novel high-throughput strategy that could be used for cloning, expression and recombinant protein production in *Escherichia coli* the most prominent one.

Following this introductory section, this thesis is divided into 7 additional chapters. The second chapter reviews our current knowledge on the microbial degradation of plant cell walls by CAZymes, with a special focus on the cellulosomes, in particular the cellulosomes of *Clostridium thermocellum* and *Ruminococcus flavefaciens*. Subsequently, the biotechnological use of CAZymes and cellulosomes are revised to allow extrapolating the relevance of genomic information data that is currently available. A description of the high-throughput methods for recombinant protein expression in the post-genomic era is provided. At the end of the bibliographic review, the objectives of this work are clearly defined. Chapters 5, 6 and 7 are organized in papers based on scientific manuscripts, already published or submitted to international peer reviewed journals. Chapters 3 and 4 are also based on scientific manuscripts which are currently in preparation. Finally, the last chapter discusses and integrates the results presented. Future perspectives for the scientific knowledge attained with this work will also discussed.

2. BIBLIOGRAPHIC REVIEW AND OBJECTIVES

2.1. Microbial degradation of plant cell wall

2.1.1. Plant cell wall components

Plant cell walls are composed of complex polymers tightly organized in a meshwork with a well-defined structural organization within each cell type, allowing cell walls to perform their mechanical and biochemical functions. Although the overall architectures of grasses and dicotyledonous cell walls are similar in that they both consist of a network of cellulose embedded in a matrix of non-cellulosic polysaccharides, they differ significantly in the type and relative amounts of these constituent polymers (Carpita & Gibeaut, 1993). During cell wall maturation the appearance of a secondary cell wall leads to a chemical composition change, which reflects the dynamic nature of plant cell walls (King *et al*., 2011).

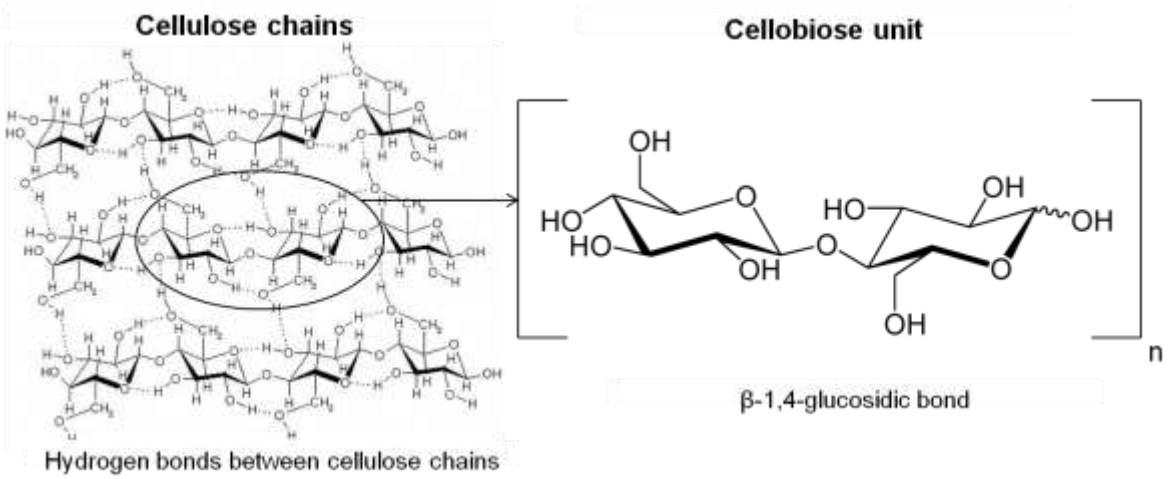
Plant cell wall structural polysaccharides are generated directly from the products of photosynthesis, which are used by plants, not only as building blocks, but also as their only source of energy (Sarkar, Bosneaga, & Auer, 2009). Around one-third of the total mass of several plants is cellulose, mainly in the form of crystalline microfibrils (Somerville, 2006). Cellulose plays a vital role as a load-bearing network and, due to its physical properties, it is important in determining the structural bias of the cell wall (Cosgrove, 1997). In addition to cellulose, the other polysaccharides of plant cell walls can be divided into two main groups: (1) hemicelluloses, such as xyloglucans, xylans, mannans or glucomannans (Scheller & Ulvskov, 2010) that bind to cellulose thus preventing the direct contact of each microfibril by acting as a lubricating coating, and (2) pectins, which act as a gel matrix in which the cellulose-hemicellulose network is embedded (Cosgrove, 1997) and modulate the cell wall porosity by constituting the major adhesive material between cells (Willats *et al*, 2001). In addition to polysaccharides, plant cell walls also contain structural proteins, enzymes with diverse roles and phenolic compounds, such as lignin (Cosgrove, 1997).

Based on their polysaccharide composition, primary cell walls are usually classified as type I or type II. Type I walls consist of a cellulose-xyloglucan framework (about 50% of the wall mass) surrounded by a hydrated matrix of pectic polysaccharides (about 30% of the total mass). Other hemicelluloses such as gluco- and galactoglucomannans, can also interlock the cellulose microfibrils (Carpita & Gibeaut, 1993). In Type II walls, arabinoxylan and/or glucomannans constitute the major hemicellulose. Furthermore, type II walls contain a higher percentage of cellulose and residual amounts of pectins and proteins. When the secondary walls form (a process that does not occur in all cell types), water is largely replaced by lignin. Lignin confers great mechanical strength and structural reinforcement and allows cell walls to become more resistant to the action of enzymes and solutes (Pauly & Keegstra, 2008).

2.1.1.1. Cellulose

Cellulose, the major constituent of plant cell walls, represents the most abundant renewable natural resource produced on the planet. Its biodegradation by microbial cellulases represents a major step in the carbon flux from fixed carbon present in cellulosic biomass to atmospheric CO_2. Cellulose consists of a collection of parallel β-1,4-linked glucan chains that interact with each other through an extensive hydrophobic and hydrogen bond network (Somerville, 2006). Sequential glucose molecules are twisted 180°, forming a snap wherein the repeating unit is cellobiose. The degree of polymerization, *n*, varies between 10.000 and 15.000, depending on the cellulose source material (Samir, Alloin & Dufresne, 2005) (Figure 2.1). The assembly of large number of glucan chains forms a crystalline microfibril, cable-like structure, contributing to the robust nature of plant cell walls (Taylor, 2008). Microfibrils cover the plant cell walls in spatially oriented overlapping layers by providing resistance to enzymatic hydrolysis (Cosgrove, 2005) as well as to osmotic pressures (Somerville, 2006).

Figure 2.1| Molecular structure of cellulose.



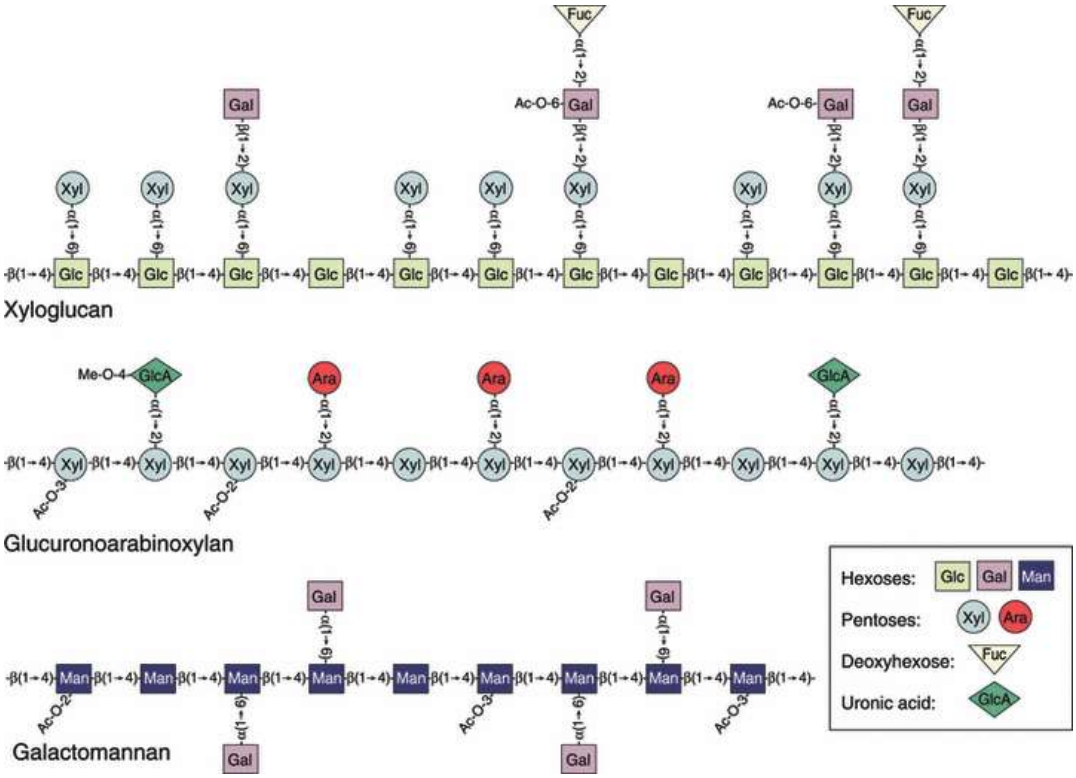
Parallel glucan chains aggregate through hydrogen bonds. The repeating unit cellobiose is indicated showing the β-1-4 glucosidic bond. Adapted from Poletto, Pistor, & Zattera (2013).

There are many polymorphs of crystalline cellulose (I, II, III and IV). Cellulose I, or “native” cellulose, is the structure found in nature and comprises two forms (Iα and Iβ) that coexist in various proportions depending on the cellulose source. Iα contains one single chain in a triclinic unit and is the dominant form of the cellulose produced by bacteria and algae, while Iβ, containing two chains in a monoclinic unit cell, dominates in the cellulose produced by the higher plants (Attala & Vanderhart, 1989; Sullivan, 1997). It is important to note that, within cellulose fibrils there are regions where the cellulose chains are arranged in a highly ordered crystalline structure, making the hydrolysis there more difficult, and regions that are loosely ordered termed “amorphous” (Nishiyama, 2009).

2.1.1.2. Hemicelluloses

Hemicelluloses constitute a heterogeneous group of noncrystalline glycans that bind tightly to the surface of cellulose microfibrils, thereby strengthening the cell wall (Cosgrove, 1997). These polysaccharides are characterized by β-1,4-linked backbones of sugars in an equatorial configuration that include xyloglucans, xylans, mannans, glucomannans and β-1,3-1,4-glucans (Scheller & Ulvskov, 2010). The backbone of hemicelluloses resembles that of cellulose. However, the presence of branches and other modifications that can decorate their backbones prevent them to form microfibrils (Cosgrove, 2005) (Figure 2.2). In xylans, mannans, and xyloglucans, the backbone sugars are β-1,4-D-Xyl, β-1,4-D-Man, and β-1,4-D-Glc, respectively, while in glucomannan the backbone consists of randomly dispersed β-1,4-Glc and β-1,4-Man sugars. Xyloglucan and arabinoxylan are two of the most abundant hemicelluloses. Xyloglucan is present in large quantities in the primary cell wall of dicotyledons and has the glucan backbone decorated with xylose branches on 3 out of 4 glucose residues. Some of the xylosyl side chains are extended by the addition of galactose (Gal) or galactose-fucose (Fuc) residues (McNeil *et al*., 1984; Cosgrove, 2005). Arabinoxylan consists of a β-1,4-D-xylan linked backbone branched with arabinose and is predominantly found in the cell walls of monocotyledons. Other residues, such as glucuronic acid (GlcA) and ferulic acid esters, may also be found decorating arabinoxylans (Cosgrove, 2005).

Figure 2.2| Chemical structure of the most abundant hemicelluloses.



Examples of hemicelluloses structures present in plant cell walls, including monosaccharides, and their linkage and ester constituents. Glc – D-Glucose; Gal – D-Galactose; Man – D-Mannose; Xyl – D-Xylose; Ara – L-Arabinose; Fuc – L-Fucose; GlcA – D-Glucuronic acid. (Pauly & Keegstra, 2008).

2.1.1.3. Pectins

Pectins, the most soluble of the cell wall constituents, form a complex and heterogeneous group of acid-rich polysaccharides (Cosgrove, 1997). Homogalacturonan (HG), rhamnogalacturonan-I (RG-I) and rhamnogalacturonan-II (RG-II) are three of the major pectic polysaccharides that occur in all primary cell walls and all of them are rich in galacturonic acid (GalA). HG (also known as polygalacturonic acid) is a linear homopolymer of 1,4-α-linked-D-galacturonic acid that is deposited in the cell wall in a form that has 70–80% of the GalA residues methyl esterified (O'Neill, Albersheim & Darvill, 1990). RG-I consists of as many as 100 repeats of the disaccharide (1,2)-α-L-rhamnose-(1,4)-α-D-galacturonic acid. The rhamnose residues can be substituted with β-1,4-galactan, branched arabinan, and/or arabinogalactan side chains (Willats *et al*., 2001; Harholt, Suttangkakul, & Vibe Scheller, 2010). RG-II is a highly complex branched carbohydrate containing an HG backbone with 11 different sugar residues and forms dimers through borate esters (Cosgrove, 2005).

2.1.2. Plant cell wall models

It is well established that all plant cell walls are composed of cellulose microfibrils that form the scaffold of the wall and a cross-linked matrix phase that fills the space among the fibrils framework (Carpita & Gibeaut, 1993; Cosgrove, 1997; Popper, 2008; Keegstra, 2010). At the same time, it is also clear that not all plant cell walls have the same chemical composition and architecture (Carpita & Gibeaut, 1993; Cosgrove, 1997; Scheller & Ulvskov, 2010; King *et al*., 2011). Over the years, many models have been proposed to illustrate how structural polysaccharides are organized in such ordered structures that are plant cell walls. The most popular and generally accepted models were proposed by Hayashi (1989) and Fry (1989) and both defend that long xyloglucan chains tether the cellulose microfibrils together while pectin polysaccharides and structural proteins occupy the space between the cellulose-xyloglucan network without being covalently bound to them. More recently, other two models were suggested: (1) the multicoat model, which proposes cellulose microfibrils to be coated with several sheaths, being hemicelluloses the inner and most tightly bound sheath; the linkage between microfibrils is made via non-covalent bindings between polysaccharide layers (Talbott & Ray, 1992), and (2) the stratified model that proposes that the cellulose-hemicellulose framework is separated by pectic layers (Ha, Apperley, & Jarvis, 1997).

2.1.3. Plant cell wall hydrolysis

Plant cell wall polysaccharides, primarily cellulose and hemicelluloses, are the most abundant source of organic carbon and energy on the planet. The photosynthetically fixed carbon is recycled by the initial action of microbial enzymes that convert cell wall polysaccharides to oligosaccharides and monosaccharides, a fundamental biological process that is of immense industrial importance. However, plant cell walls are considered to be

recalcitrant structures to biological depolymerisation, which makes the carbon cycle a relatively inefficient process. The physical association between polysaccharides and between polysaccharides and lignin restricts the accessibility to the microbial enzymes that participate in plant cell wall polysaccharide deconstruction (Gilbert, 2010) and only a limited number of microorganisms have acquired the capacity to produce these enzymes (Fontes & Gilbert, 2010). So, at one side we have plant cell strategies to protect their rich chemical energy stores, while on the other side we have microbial strategies for the efficient breakdown of cell walls and thus breach of such protective efforts developed by plants.

Throughout evolution, microbes have evolved an extensive arsenal of hydrolytic enzymes, generally termed as Carbohydrate-Active enZymes (CAZymes), for the efficient attack of a heterogeneous insoluble and highly recalcitrant substrate that constitutes the majority of plant cell wall. The plant cell wall-degrading apparatus of aerobic and anaerobic microorganisms differ in their macromolecular organization (Warren, 1996). Aerobes produce extracellular enzymes in large quantities, which although do not physically associate, display extensive biochemical synergy to convert polysaccharides to soluble products that are transported into the cells. In contrast, in most anaerobic microorganisms, plant cell wall-degrading enzymes frequently assemble into a large multienzyme complex, termed the "cellulosome" (Gilbert, 2007; Fontes & Gilbert, 2010). It is believed that the anaerobic environment imposes selective pressures for the evolution of this highly efficient plant cell wall degrading nanomachine (Bayer *et al*., 2004).

2.1.3.1. Carbohydrate-Active Enzymes (CAZymes)

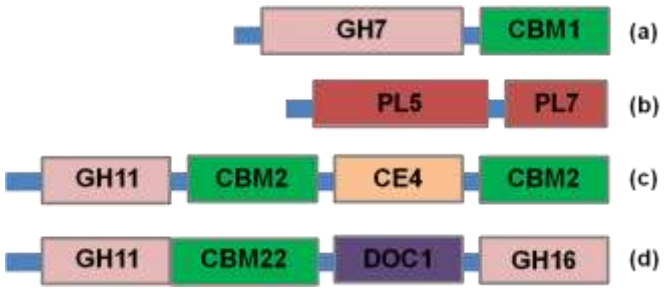
As polysaccharides exhibit remarkable diversity, it is not surprising that their degradation by microbes involves diverse enzymes with different specificities and modes of action (Warren, 1996). Carbohydrate-active enzymes (CAZymes) are frequently modular, where a module can be defined as a structural and functional unit (Figure 2.3). Usually, CAZymes contain one catalytic module connected through a flexible linker sequence to non-catalytic modules involved in protein-carbohydrate interaction, also termed carbohydrate-binding modules (CBMs, see section 2.1.3.3) or modules involved in protein-protein interactions (the most predominant of those being the dockerins, see section 2.1.3.4). In the last years, many CAZymes have been identified, characterized and their individual modules grouped into multiple families according to their sequence and structural similarities (Henrissat, 1991; Henrissat, Teeri & Warren, 1998). These protein families are accessible at the constantly updated CAZy database (http://www.cazy.org) (Cantarel *et a*l., 2009). CAZymes of the same family display a common fold, while the catalytic apparatus and mechanism are similarly conserved (Gilbert, 2010). Significant sequence similarity (usually over 30%) is a strong sign of folding similarities. Thus, if the three-dimensional structure of one member of a family is known, it is possible to do homology modelling and deduce structural insights for other family

members. Consequently, this system of cataloguing CAZymes in families is used to classify protein modules of unknown function, of which the only recognised feature is sequence similarity. When novel families are created (sequence homology less than 30%), previously released information is reanalyzed to take the additional new family into account (Henrissat, 1991; Cantarel *et al*., 2009). However, it should be noted that classification of a protein module within a family does not directly establish a function for an enzyme, since substrate specificity is not conserved among CAZyme families.

CAZymes are a class of enzymes which make, break or modify glycosidic bonds and they fall into four main categories: glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and glycosyl transferases (GTs). It is important to note that not all of these enzymes contribute to degrade carbohydrates from the plant cell wall, as exemplified by lytic polysaccharide mono-oxygenases (LPMO) included in the Auxiliary Activities (AA) families recently incorporated on the CAZy database, which also includes lignin degrading enzymes (Levasseur *et al*., 2013).

CAZyme terminology was proposed by Henrissat *et al*. (1998) based on the family to which the enzyme belongs and its target substrate. Thus, the first three letters of the enzyme abbreviation identify the substrate, followed by the family number and by an uppercase letter corresponding to the order by which the catalytic domain was reported. For example, a family 5 GH will be named Cel5 or Man5, depending on its substrate (cellulose or mannose respectively) and by Cel5A or Cel5B if there were two catalytic domains with the same specificity but reported at different times. The microorganism abbreviation may also be included before the enzyme name, in order to differentiate similar enzymes of different origins. For example, the enzyme from *Clostridium thermocellum* composed of a licheninase family 26 A (first to be discovered) catalytic domain fused to a cellulase of family 5 E (fifth one to be published) will be *Ct*Lic26A-Cel5E, written in the conventional sense from the amino- to the carboxyl-terminus of the protein.

Figure 2.3| Modular architecture of Carbohydrate-Active enZymes (CAZymes).

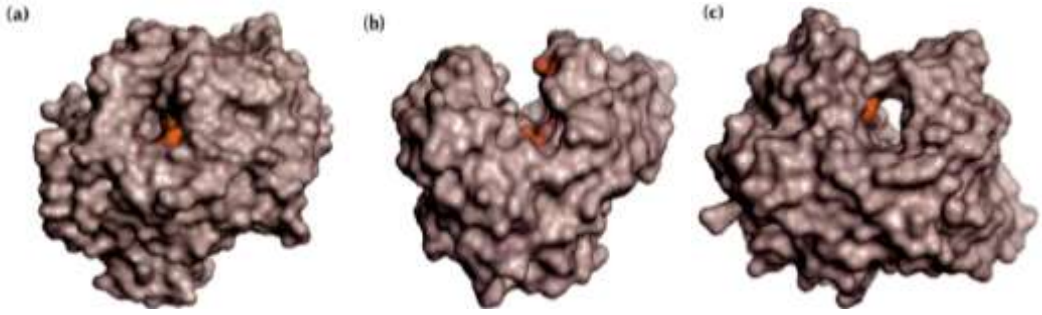


Examples of modular CAZymes. (a) cellobiohydrolase I from *Hypocrea jecorina* (SP P00725); (b) alginate lyase from *Sphingomonas* sp. A1 (GB BAB03312.1); (c) xylanase from *Cellulomonas fimi* (GB CAA54145.1); (d) xylanase D/licheninase from *Ruminococcus flavefaciens* (GB CAB51934.1). Adapted from Cantarel *et al*. (2009).

2.1.3.1.1. Glycoside hydrolases (GHs)

Glycoside hydrolases (EC 3.2.1.-) also referred as Glycosidases or Transglycosylases attack either β- or α-glycosidic bonds in di-, oligo and polysaccharides. The CAZy database categorizes GHs in 133 families (data collected on February 2015). They can be retaining or inverting enzymes, when catalyze transglycosylation or hydrolysis reactions with retention of configuration at the anomeric center, or when they catalyze hydrolysis reactions with inversion of configuration at the anomeric center, respectively (McCarter & Withers, 1994). Glycoside hydrolases also differ in the products they release when act on a particular substrate. Exo-acting enzymes remove units of one or more sugars from the ends of the polysaccharide chain. Endo-acting enzymes randomly hydrolyze glycosidic bonds within the chains, thereby producing more ends for the exoenzymes to act on (Warren, 1996). Exoenzymes and endoenzymes act in synergy, which increases the efficiency of polysaccharide hydrolysis. This is crucial for long linear substrates, such as cellulose, where the numbers of polysaccharide ends for exo-enzyme attack are limiting factors. However, it is not clear the distinction of these two types of enzymes since some exo-acting enzymes have some residual endo-acting activity (Ståhlberg, Johansson & Pettersson, 1993). The distinction can also be reflected by the architecture of the active sites, which fall into three general classes (Figure 2.4). Endoglucanases, for example, are commonly characterized by the presence of a groove or cleft into which any part of a cellulose chain can fit. In contrast, exoglucanases bear tunnel-like active sites, which can only accept a substrate chain via its terminus, being the hydrolysis processed in a sequential manner resulting the term of “processive enzymes” (Davies & Henrissat, 1995). Nevertheless, structural changes can convert endo-acting glycoside hydrolases into exo-acting enzymes (Gilbert, 2010). Finally, enzymes acting on the removal of decoration of the polysaccharide backbone contain pockets that recognize sugar side-chains. Glycoside hydrolases exhibit different degrees of substrate specificity; some enzymes have an exclusive target, while others act on different substrates (Warren, 1996). Recent observations in the bacterium *Clostridium thermocellum* suggest an evolutionary adaptation of some GHs to function as polysaccharide binding agents (like carbohydrate-binding modules) rather than enzymatic components, thus serving as extracellular carbohydrate sensors of the microorganism (Bahari *et al*., 2011).

Figure 2.4| The three types of active sites found in glycoside hydrolases.

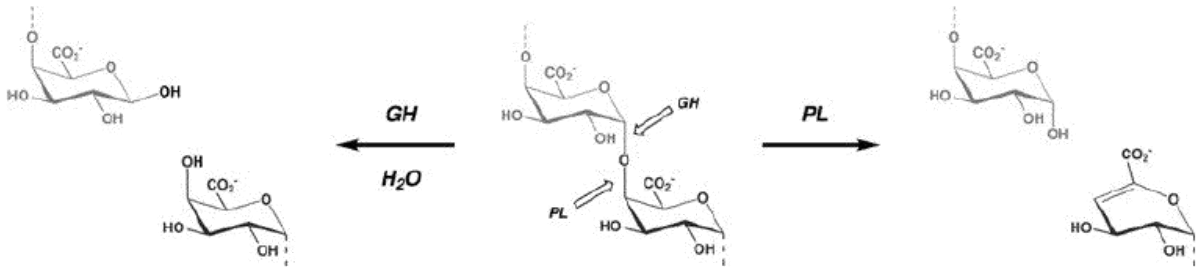


The catalytic residues are highlighted in red. a) The pocket or crater found in non processive exo-acting enzymes (glucoamylase from *Aspergillus awamori*); b) The cleft or groove found in endo-acting enzymes (endoglucanase E2 from *Thermononospora fusca*; c) The tunnel found in processive exo-acting enzymes (cellobiohydrolase II from *Trichoderma reesei*). Adapted from Davies & Henrissat (1995).

2.1.3.1.2. Polysaccharide lyases (PLs)

Polysaccharide lyases (EC 4.2.2.-) cleave the glycosidic bonds of uronic acid-containing polysaccharides by β-elimination instead of a hydrolytic mechanism and thus generate an unsaturated hexenuronic acid residue and a new reducing end at the point of cleavage (Yip & Withers, 2006) (Figure 2.5). The CAZy database categorizes PLs in 23 families (data collected on February 2015). As described for GHs, members of PLs are frequently polyspecific, including enzymes acting on different substrates or that generate different products. Concerning to folds and structures, PLs show a large variety of fold types (or classes), ranging from β-helices to α/α barrels (Lombard *et al*., 2010).

Figure 2.5| Comparison of the products generated by a polysaccharide lyase (PL) and a glycoside hydrolase (GH) exemplified by polygalacturonate (pectate) cleavage.



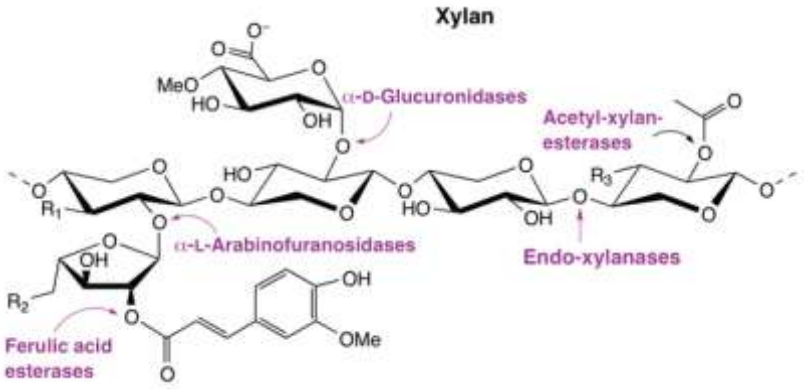
Both enzymes PL and GH generate a new reducing chain end (light grey). GHs cleave the glycosidic bond (C-1':O-4) by the addition of water, maintaining the 4-OH group at the new non-reducing chain end. PLs generate a hexeneuronic acid moiety at the new non-reducing end by elimination of the O-4:C-4 bond (Lombard *et al*., 2010).

2.1.3.1.3. Carbohydrate esterases (CEs)

Carbohydrate esterases remove ester- based modifications present in mono-, oligo- and polysaccharides and thereby facilitate the action of GHs on complex polysaccharides (Figure

2.6). Two classes of substrates for CEs are considered: those in which the sugar plays the role of "acid", such as pectin methyl esters, and those in which the sugar behaves as “alcohol”, such as in acetylated xylan (Cantarel *et al*., 2009). The CAZy database categorizes CEs in 16 families (data collected on February 2015).

Figure 2.6| Carbohydrate esterases and glycoside hydrolases involved in hemicellulose degradation, exemplified by xylan cleavage.



Feruloyl esterases or ferulic acid esterases (FAE) increase the release of sugars from xylan by removing the ferulic acid residues, thus destabilizing the structure and making it more susceptible to the action of hydrolytic enzymes. Acetyl esterases promote deacetylation of cellulose acetates by acting on carboxylic ester bonds. Adapted from DeBoy *et al*. (2008).

2.1.3.1.4. Glycosyl transferases (GTs)

Glycosyl transferases (EC 2.4.x.y) catalyze the transfer of a sugar moiety from an activated donor to a specific acceptor molecule to create a glycosidic bond. According to the stereochemistry of the substrates and reaction products, GTs can be classified as either retaining or inverting enzymes (Sinnott, 1990). The CAZy database categorizes these enzymes in 97 families (data collected on February 2015).

2.1.3.2. Non-catalytic modules and linker regions

As described above, the molecular architecture of most carbohydrate-active enzymes is frequently modular with different degrees of complexity, which may range from two to six or more domains bound through linker sequences of different lengths and amino acid composition. Linkers supply enzyme flexibility and provide the required spatial distance between modules to contribute to enhance enzyme-substrate interactions (Noach *et al*., 2009). Usually, linker regions are rich in serine and threonine (Coutinho & Reilly, 1994) and may be glycosylated conferring protection against proteolysis (Tomme *et al*., 1995). In addition to catalytic domain(s), CAZymes may contain one or more non-catalytic CBMs often organized in tandem. The CBMs’ main function is to recognize and bind specifically to carbohydrates (Boraston *et al*., 2004). Other non-catalytic modules identified in CAZymes

include, in addition to dockerins, thermostabilizing domains (Fontes *et al.*, 1995b), S-layer-homologous (SLH) domains for cell attachment (Sa & Sleytr, 2000), or fibronectin type III-like sequences that may have different functions, such as ligand-binding modules, peptide linkers or spacers between other domains, cellulose-disrupting modules or even help large enzyme complexes remain soluble (Alahuhta *et al*., 2010). In theory, the number of possible combinations of domains is enormous (Fernandes *et al*., 1999; Lombard *et al*., 2010). The modular nature of CAZymes constitutes an adaptation strategy to overcome the restrict access of the enzymes to the polysaccharides linkages within plant cell walls.

2.1.3.3. Carbohydrate-Binding Modules (CBMs)

CBMs, small sequences that contain from 30 to about 200 amino acids, constitute the majority of non-catalytic modules identified in CAZymes. They constitute auxiliary domains with autonomous folding and specific capacity to recognize heterogeneous and complex carbohydrates, thus promoting the association of the enzyme with their target substrates. CBMs can be located at the N- or C-terminal ends of CAZymes, between two catalytic modules, as a single unit or arranged in tandem (Guillé *et al*., 2010). CBMs can also be found independently from catalytic domains, such as the CBMs located in cellulosomal scaffolding proteins, such as the CBM3 of *C. thermocellum* scaffoldin CipA, which binds strongly to the crystalline cellulose (Bayer *et al*., 2004).

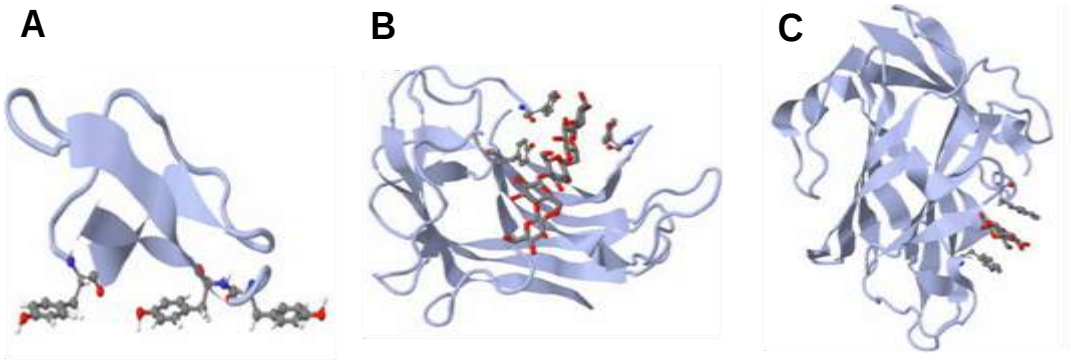
Initially, these modules were defined as cellulose-binding domains (CBDs) because they were first described to have a binding capacity to crystalline cellulose (Gilkes *et al.,* 1988). Subsequently, in order to reflect the diverse ligand specificities identified in these modules, the more inclusive term of CBM was proposed (Boraston, *et al*., 1999). Today, the ligand specificity of CBMs was recognized for a large variety of polysaccharides including crystalline cellulose, non-crystalline cellulose, chitin, β-1,3-glucans, β-1,3-1,4-mixed linkage glucans, xylan, mannan, galactan or starch (Boraston *et al*., 2004). As described for CAZymes, CBMs are divided into families based on amino acid sequence similarity on the continuously updated CAZy database (Cantarel *et al*., 2009). There are currently 71 defined families of CBMs (data collected on February 2015). With respect to nomenclature, the rules employed to describe CBMs are similar to the ones described for CAZymes. At its simplest, a CBM is named by its family but one may also include the organism and even the enzyme from which it is derived. If there are many modules of CBMs belonging to the same family in tandem, a number corresponding to the position of the CBM in the enzyme relative to the N-terminus is included. For example, *Clostridium stercorarium* contains an enzyme with a triplet of family 6 CBMs, being the first CBM referred as *Cs*CBM6-1, the second as *Cs*CBM6-2 and the third as *Cs*CBM6-3 (Boraston *et al*., 2004).

CBMs with similar fold are observed to present different ligand specificities (Guillén *et al*., 2010). The most common fold of CBMs is the β-sandwich followed by the β-trefoil. The β-

sandwich fold comprises two β-sheets, each of which consisting of three to six antiparallel β-strands. An example of a β-sandwich conformation is the *C. thermocellum* CBM11 (Carvalho *et al*., 2004). In contrast, CBMs with the β-trefoil fold contain 12 β-sheets, forming six hairpin turns. An example of a β-trefoil conformation is the *C. thermocellum* CBM42 (Ribeiro *et al*., 2010).

Based on the topology of CBM-ligand binding site, CBMs have been classified into three types: A ("surface binding"); B ("glycan chain binding") and C ("small sugar binding") (Boraston *et al*., 2004) (Figure 2.7). Type A CBMs have a flat or platform-like hydrophobic surface composed of aromatic residues. Due to complementary conformations, the flat type A binding sites interact with the flat surfaces of crystalline polysaccharides such as cellulose or chitin. The binding-sites of Type B CBMs have a cleft arrangement in which aromatic residues decorating the concave ligand-binding surface interact with free single polysaccharide chains. Aromatic side chains are oriented in such a way that forms twisted or sandwich platforms. The binding site architecture of Type B allows binding to amorphous cellulose or xylan. These CBMs also recognize substrates like β-1,3-glucans, mixed β-1,3-1,4-glucans, β-1,4-mannan, glucomannan, and galactomannan. Type C CBMs or lectin-like CBMs only bind mono-, di-, or trisaccharides due to steric restriction in the binding site (Boraston *et al*., 2004; Guillén *et al*., 2010). Recently, some refinements to the classification in Types A, B and C were proposed by Gilbert *et al*. (2013) whereby the Type B CBMs are classified as CBMs that bind internally on glycan chains (endo- type) and Type C CBMs are defined as CBMs that bind the termini of glycans chains (exo-type) (Gilbert, Knox & Boraston, 2013). The importance of the side chains of aromatic amino acids for carbohydrate recognition, in particular tryptophan but also tyrosine, is well known. They form stacking interactions with sugar rings resulting in strong Van der Waals interactions that stabilize the structure of the protein-carbohydrate complexes (Guillén *et al*., 2010). Also hydrogen bonds and calcium-mediated co-ordination play a key role in ligand recognition by CBMs (Boraston *et al*., 2004). According to binding affinity, proteins that bind to carbohydrates can be divided into two groups: group I include proteins which bind carbohydrates tightly ($Ka>10^6\ M^{-1}$), and group II comprises proteins that bind carbohydrates weakly ($Ka<10^6\ M^{-1}$). This group include all CBM-carbohydrate interactions (Quiocho, 1986).

Figure 2.7| Structures representatives of the three CBM types distinguished by the topology of carbohydrate binding site.



A) Type A - CBM1 from *Trichoderma reesei* cellobiohydrolase I (PDB code 1CBH); B) Type B - CBM4 from *Cellulomonas fimi* endo-1,4-glucanase C (PDB code 1GU3); and C) Type C - CBM9 from *Thermotoga maritima* xylanase 10A (PDB code 1I82). Adapted from Guillén *et al*. (2010)

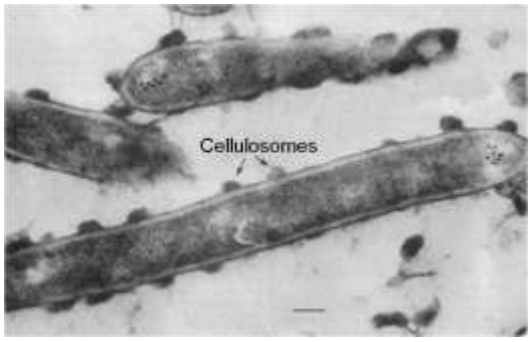
It is well established that the main function of CBMs is to recognize and bind specifically to carbohydrates. When appended to catalytic modules, CBMs fulfill the following three different roles to potentiate the efficiency of the associated enzymes: (1) proximity effect, (2) targeting function and (3) disruptive function (Boraston *et al*., 2004). CBMs can increase the concentration of the enzymes on the polysaccharide surface, which improves the proximity between the enzyme and the substrate, thus enhancing catalysis (Bolam *et al*., 1998). Therefore, it is clear that removal of CBMs from their appended enzymes, or from cellulosomal scaffoldins significantly decreases the activity of the associated catalytic modules on insoluble carbohydrates (Bolam *et al*., 1998; Boraston *et al*., 2003). However, resulting activity of truncated enzymes on soluble substrates is not frequently affected (Gilkes *et al*., 1988; Bolam *et al*., 1998; Boraston *et al*., 2003; Kleine & Liebl, 2006). Recent studies show that xylan- and cellulose-targeting CBMs can modulate the activity of appended catalytic modules against polysaccharides. Thus, CBMs may target either the substrate hydrolyzed by the catalytic module or non-substrate polysaccharides, promoting the enzyme activity. This capacity is greatly advantageous for the tightly packed cell walls in which the access to the polysaccharides is highly restricted. Nevertheless if the target for the CBM is absent or present at reduced levels, the impact of the CBM on enzyme activity is absent or significantly reduced (Hervé *et al*., 2010). CBMs may also enhance enzymatic activity by disrupting the interface between the substrate and other polysaccharides within the wall thereby improving substrate accessibility. Gao *et al*. (2001) suggested that binding of the cellulose-binding domains (CBD) to cotton fibers leads to structural changes in the polysaccharide and release of short fibers. Later, the same group reported that attachment of a CBM to cotton fibers promotes severe weakening of the cellulose-interchain hydrogen bonds (Wang, Zhang & Gao, 2008). It has been hypothesized that cellulose-specific CBMs might have a common mechanism of action with plant proteins termed expansins which

disrupt the structure of cellulose. However, in contrast to expansins that break hydrogen bonds between cellulose microfibrils and matrix hemicelluloses resulting in a decrease in the rigidity of plant cell walls, CBMs are not involved in a reduction of plant cell wall rigidity (Bolam *et al*., 1998).

2.1.3.4. Cellulosomes

In the early 1980s, a cell-bound, multicellulase complex designated “cellulosome” was first identified in the anaerobic thermophilic bacterium *Clostridium thermocellum*. At that time it was postulated that cellulosomes were responsible for both adherence and hydrolysis of the cellulose substrate (Lamed *et al*., 1983). In addition, it was suggested that arrangement of multiple cellulolytic enzymes (each of which containing intrinsic affinity for the substrate) into a large complex may serve to collectively enhance the catalytic efficiency of the entire complex (Bayer, Setter, & Lamed, 1985). More recently, cellulosomes have been identified in many anaerobic bacteria from the genera *Clostridium*, *Acetivibrio*, *Bacteroides* or *Ruminococcus*, which colonize different environmental niches, as well as in fungi (Fontes & Gilbert, 2010). It is known that besides cellulases, these complexes also contain hemicellulases, pectinases and other enzyme activities that include polysaccharide lyases, carbohydrate esterases, proteases and protease inhibitors (Doi *et al*., 2003; Tamaru & Doi, 2001). Cellulosomes were identified at the surface of *C. thermocellum* located in protuberances present on the cell envelope (Bayer & Lamed, 1986) (Figure 2.8). Since the primary sequences of all known cellulosomal polypeptides present a signal peptide (Béguin & Lemaire, 1996), it is believed that cellulosome assembly takes place in the extra-cellular media after protein sortage (Bayer & Lamed, 1986). It is well established that cellulosomes are more efficient to deconstruct plant structural polysaccharides than the “free” enzymes produced by aerobic bacteria and fungi (Fontes & Gilbert, 2010). This efficiency is justified by the potentiation in enzyme synergy afforded by enzyme proximity when enzymes are organized in cellulosomes together with the complexity of their multimodular enzymes.

Figure 2.8| Cellulosomes at the surface of *Clostridium thermocellum*.



Transmission Electron Microscopy (TEM) of cationized ferritin (CF)-labeled cellobiose-grown cells of *C. thermocellum* YS. Cells were grown on cellobiose. Cellulosomes correspond to the nodulous protuberances which appear in large numbers over the entire cell surface. Adapted from Bayer & Lamed (1986)

Integration of cellulosomal components occurs via highly ordered protein:protein interactions established between non-catalytic dockerin domains located on the cellulosomal enzymes with cohesin domains located on a molecular scaffold. In some cases, two or more types of scaffoldins may be involved in cellulosome assembly: primary scaffoldins and cell-surface anchoring scaffoldins. Cellulosome assembly promotes enzymes synergism, due to spatial proximity, and enzyme-substrate targeting. Cohesins are ~150-amino-acid-residues modules typically presented in tandem in scaffolding proteins, while the dockerins are modules comprehending two ~22-amino-acid duplicated segments usually located as a single copy at the C-terminus of the enzymes (Bayer *et al*., 2004). Calcium is required for dockerin stability and function and consequently for the cellulosomal integrity; treatment of the cellulosomes with EDTA, which chelates calcium, decreases the hydrolytic capacity of the cellulosome as dockerins are unable to bind cohesins (Lytle *et al*., 2000; Choi & Ljungdahl, 1996). Both cohesins and dockerins are highly homologous within the same species and residues supporting the protein:protein intractions are highly conserved (Fontes & Gilbert, 2010). It is important to note that these modules are also present in non-cellulosome-producing microorganisms, where they can be components of non-degrading carbohydrates enzymes (Peer *et al*., 2011).

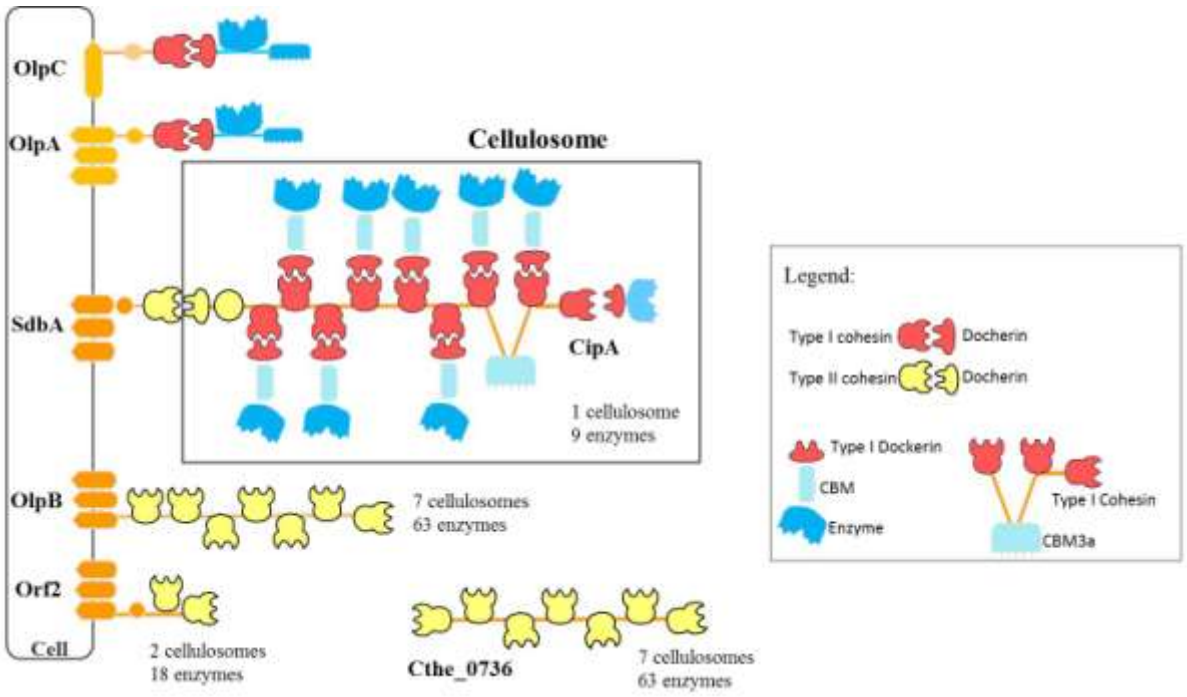
Assembly and polypeptide composition of cellulosomes vary among microorganisms as well as with the nature of carbon source available, making these higly complex nanomachines extraordinary dynamic structures. Bacterial cellulosomes can be classified in two types: one that presents multiple types of scaffoldins (both primary and anchoring as described in *C. thermocellum*, for example), and a second type which contains a single primary scaffoldin (most of them from mesophilic microorganisms). Thus, these simplest cellulosomes do not interact with the bacterial cell surface (Fontes & Gilbert, 2010).

2.1.3.4.1. The cellulosome of *Clostridium thermocellum*

The cellulosome of the anaerobic thermophilic bacterium *C. thermocellum* is one of the best characterized and presents one of the highest rates of cellulose degradation known to date (Demain, Newcomb, & Wu, 2005). The molecular base of *C. thermocellum* cellulosome assembly depends on the presence of a primary scaffoldin, called CipA, which bear up to nine catalytic subunits (Figure 2.9). The attachment of a given catalytic subunit is mediated by the interaction of its type I dockerin domain with one of the nine type I cohesin domains of the primary scaffoldin CipA (Kruus *et al*., 1995). Between the second and third cohesin repeats of CipA there is a family 3 CBM which targets the entire complex to crystalline cellulose (Poole *et al*., 1992). CipA is, in turn, attached to the cell surface through the interaction of its type II C-terminal dockerin domain with the type II cohesin domain of one of three S-layer anchoring scaffoldins: SdbA (contains one type II cohesin), Orf2 (contains two type II cohesins), or OlpB (contains seven type II cohesins) (Leibovitz & Béguin, 1996;

Demain *et al*., 2005). A fourth anchor protein containing seven type II cohesins, Cthe_0736, is believed to be exclusively extracellular. Type I cohesins modules were identified in two cell surface proteins (OlpA and OlpC) suggesting that, in addition to CipA-attachment, cellulosomal enzymes can also adhere directly onto the cell surface through their type I dockerins (Fontes & Gilbert, 2010)

Figure 2.9| Organization of *C. thermocellum* cellulases and hemicellulases in cellulosomes.



Schematic drawing depicting *C. thermocellum* cellulosomes attachment to the cell surface through the type II dockerin-cohesin interaction with cell surface proteins SdbA, OlpB and Orf2 (in orange). The cellulosome complex is boxed. On the other hand, OlpA and OlpC, both of which contains a type I cohesin, are presumed to anchor an enzyme or protein containing a type I dockerin. The type II dockerin from CipA can also binds specifically to type II cohesins of the extracellular Cthe_0736. Adapted from Fontes & Gilbert (2010)

The presence of contiguous repetitions of type II cohesins in anchoring scaffoldins allows the formation of polycellulosomes. In *C. thermocellum*, three polycellulosomes can be assembled containing 9, 18 and 63 catalytic units if all type II cohesins of SdbA, Orf2 and OlpB, respectively, are bound to type II dockerins of CipA (Fontes & Gilbert, 2010). The structural organization of the cellulosome and its attachment to the cell surface depends on protein-protein interactions of similar type, which retain the spatial flexibility required to optimize the catalytic synergy within the enzyme complex (Carvalho *et al.*, 2003). Nevertheless type I and type II cohesins-dockerins pairs do not interact, ensuring a clear distinction between the interactions involved in cellulosome assembly and the cell surface attachment, respectively (Leibovitz & Béguin, 1996). It should be noted that, in *C.*

thermocellum, cellulosomes are released to the extracellular medium in the latter phase of the growth, which may indicate the requirement for a more mobile enzyme complex to target more recalcitrant forms of cellulose (Fontes & Gilbert, 2010). The complexes are constructed extracellularly, probably at the cell surface maybe mediated by Ca^{2+} present outside the cell (Demain et al., 2005). Despite various studies dedicated to *C. thermocellum* cellulosome, several questions concerning the structure, function and its importance for polysaccharide hydrolysis still remain to be elucidated.

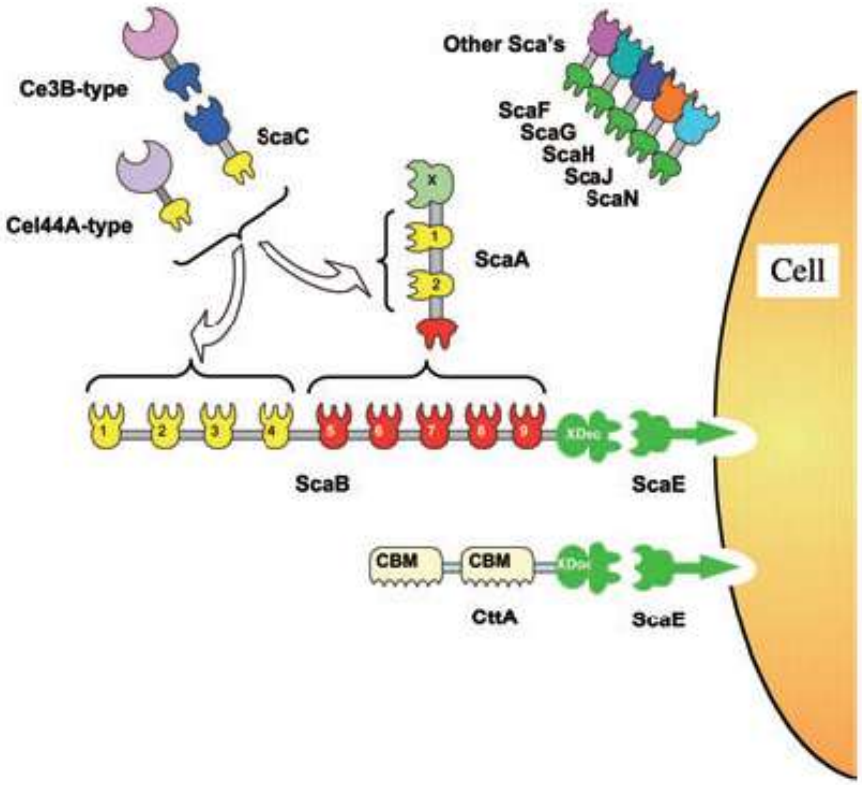
2.1.3.4.2. The cellulosome of *Ruminococcus flavefaciens*

The cellulosome of the ruminal cellulolytic bacterium *R. flavefaciens* represents the most elaborate and potentially versatile multi-enzyme complex known to date, as judged by the recent genome sequencing and subsequent biochemical and bioinformatics analysis of the bacterium genome (Bayer *et al*., 2008). Initial observations suggested that its cohesin modules clearly diverge in sequence and structure from the previously described type I and type II systems of *C. thermocellum*, so cohesin-dockerin pairs of *R. flavefaciens* were termed type III (Ding *et al*., 2001). Several studies suggest that cellulosome structural organization varies amongst different strains of *R. flavefaciens*, which may reflect the complexity of the rumen anaerobic environment and the heterogeneity of lignocellulosic substrates (Jindou *et al*., 2006).

Work on *R. flavefaciens* strain 17 revealed four scaffoldins (ScaA, ScaB, ScaC and ScaE) carrying one or more cohesin domains (Ding *et al*., 2001; Rincon *et al*., 2003). ScaE is covalently attached to the bacterial cell envelope and provides the anchoring point of the largest structural protein ScaB. In contrast to the anchoring scaffoldins from *Clostridia*, in which cell-surface attachment is mediated via the nonconvalent binding of SLH modules to the S-layer of the host cell, ScaE is covalently attached to the cell wall through a sortase-mediated mechanism (Rincon *et al*., 2005). ScaB carries seven homologous cohesin modules that exclusively interact with the dockerin of ScaA. Both ScaA and ScaC contain cohesins (three and one, respectively) capable to bind to a wide range of enzyme subunits as well as many yet uncharacterized proteins (Rincon *et al*., 2003). The single cohesin of ScaE shows significant divergence from ScaA, ScaB, and ScaC cohesins and its binding with ScaB occurs via a novel cohesin-dockerin interaction (Rincon *et al*., 2005). In *R. flavefaciens* strain FD1 (Figure 2.10), which encodes more than 200 dockerin-containing proteins (Bayer *et al*., 2008), the equivalent ScaB scaffoldin differs in the type and number of cohesins. It carries nine cohesins that present two different specificities: four bind directly to dockerins appended to catalytic units and five bind to the C-terminal dockerin of ScaA (Jindou *et al*., 2006). ScaA carries only two cohesins, in contrast to the three that are present in ScaA from *R. flavefaciens* 17. In addition, ScaC, a small dockerin-bearing protein with a single divergent cohesin module, is able to bind both ScaA and ScaB dockerins and also

other dockerins from catalytic units or proteins of unknown functions. Thus, ScaC has been proposed to function as an adaptor protein that expands the dockerin-binding repertoire of the cellulosome (Rincón *et al*., 2004). There is a fifth protein reported in both strains 17 and FD1, called CttA, of which the C-terminus resembles that one of ScaB. This protein is also able to bind to the ScaE anchoring scaffoldin, thereby becoming attached to the bacterial envelope and carries two putative CBMs that may mediate the primary anchorage to cellulosic substrates. CttA likely provides a mechanism for substrate binding, perhaps compensating for the absence of an identified cellulose-binding module in the major cellulosomal scaffolding proteins identified in this specie (Rincon *et al*., 2007).

Figure 2.10 | The complexity of *R. flavefaciens* strain FD-1 cellulosome.



The single cell-surface scaffoldin ScaE may bind CttA or ScaB, which contains cohesins with two different specificities. One cohesin type (red) exclusively interacts with the adaptor scaffoldin ScaA. The other cohesin type of ScaB (yellow) binds cellulosomal enzymes or ScaC. In addition, ScaA contains two cohesins that present a similar specificity to the second set of cohesins of ScaB. Like ScaA, ScaC is an adaptor scaffoldin that recognizes a different set of dockerin-containing proteins. Other adaptor scaffoldins, presenting a similar structure to ScaC but displaying a yet unknown specificity, exist in *R. flavefaciens*. Adapted from Bayer *et al*. (2008)

2.2. Biotechnological use of cellulases and hemicellulases

2.2.1. Enzymatic supplementation of cereal-based diets for poultry

In recent years we have observed an increased usage of exogenous enzymes in animal nutrition. It is known that dietary soluble non-starch polysaccharides, such as β-glucans and arabinoxylans, which are present in cereals such as barley and wheat, respectively, display an anti-nutritive effect for poultry and pigs. Solubilization of these non-starch polysaccharides after feed ingestion leads to an increase in digesta viscosity, which affects the digestive process by decreasing nutrient availability and, ultimately, animal performance (Choct, 1997).

Non-starch polysaccharides (NSPs) are a large variety of polysaccharides molecules which differ in composition and structure from starch (the predominant carbohydrate in cereals) and comprise some of the most representative compounds of the plant cell wall, such as cellulose, hemicelluloses and pectins (Morgan & Bedford, 1995; Williams *et al*., 1997). Different cereals contain different levels of NSPs. Choct & Annison (1992) classified different plants based on their total NSP content from low to high as follows: rice, sorghum, maize, wheat, triticale, rye and barley. Nevertheless, the NSP content also varies within the same plant species due to genotype and the geographical location where cereals are grown (Williams *et al*., 1997; Khattak *et al*., 2006). As mentioned before, plant cell wall polysaccharides have complex physical structures and different chemical inter-linkages. Single-stomach animals, such as poultry, do not possess endogenous enzymes capable of cleaving the β(α) linkages of NSPs and thus these polysaccharides are not well digested by these animals (Adams & Pough, 1993). Water-insoluble NSPs, which include cellulose, can be considered practically undigested by poultry and pigs and have the ability to absorb large amounts of water and to maintain the normal motility of the gut. In contrast, soluble NSPs are more susceptible to biological hydrolysis and are partially digested by birds in the last compartments of the gastrointestinal (GI) tract, particularly in the caecum (Carré, 1993). However, soluble NSPs display an anti-nutritive effect for poultry due to the resulting increase in digesta viscosity in the upper regions of the GI tract (Choct, 1997). An increase in digesta viscosity decreases feed passage rate leading to a reduction in feed intake. The efficiency of nutrient absorption through the intestinal wall also decreases as soluble NSPs form a viscous gel that reduces the capacity of digestive enzymes to interact with their substrates (Johnson & Gee, 1981; Edwards, Johnson, & Read, 1988). In addition, modifications in gut physiology include enlargement of digestive organs and an increased secretion of digestive juices which increases the energy maintenance needs of these animals. The lower passage rate of high viscous digestas also result in the proliferation of a fermentative anaerobic microflora in the upper compartments of the GI tract that can lead to the production of toxins and to the deconjugation of bile salts, which are essential for the digestion of fat (Choct, 1997). All these effects result in wet beds and outbreaks of

coccidiosis due to the production of wet feaces (Ribeiro *et al*., 2008). The conjugation of all these digestive problems lead to a poor growth performance as a result of the reduction of both feed intake and feed digestibility. Other factors that influence NSP digestion by poultry include age, solubilty and amount of NSPs in the diet (Choct *et al*, 1996). Young chicks are affected to a greater degree by anti-nutritional compounds than older birds due to a poor development of the repertoire of endogenous digestive enzymes (Marquardt *et al*., 1996)

Monogastric animals like poultry lack the repertoire of enzymes required to depolymerize the anti-nutritive NSPs, and thus addition of exogenous enzyme to animal diets becomes necessary. Therefore, exogenous enzymes are used either to supplement an endogenous deficiency or to supply a digestive capacity nonexistent in the host animal (Creswell, 1994). Enzymes catalyze depolymerisation of NSPs and thus decrease their viscosity when in aqueous solutions. Consequently, decreased intestinal viscosity leads to an improvement of the digestibility of nutrients by improving gut performance (Williams *et al*., 1997; Khattak *et al*., 2006). To reduce the viscosity of digesta, exogenous enzymes need to randomly hydrolyse the polysaccharide chain thus reducing the degree of polymerization of the carbohydrates (Williams *et al*., 1997). As a result of the decrease in digesta viscosity, feed enzymes were shown to enhance feed digestibility and absorption of nutrients. In addition, as a result of the action of exogenous enzymes there is an improvement in the Apparent Metabolizable Energy (AME) value of the diet, an increase in feed intake, weight gain and an improvement of feed conversion ratios. Moreover, feed enzymes have physiological impacts in animals as the reduction in digesta viscosity leads to a decrease in the size of the GI tract and a change in the population of microorganisms colonizing the hind gut. Finally, enzymes contribute to reduce water intake, water content of excreta and the output of excreta, including the reduction of N and P (Khattak *et al*., 2006). Over the years, CAZymes such as β-cellulases, β-glucanases, β-xylanases, β-galactosidases, β-mannanases or pectinases have been extensively used by the animal feed industry to reduce the detrimental effects associated with the ingestion of NSPs. Furthermore, complementary activities including phytases, proteases and lipases are known to be effective when associated with CAZymes. The use of a combination of CAZymes with different specificities acting in synergy has been proved to be highly advantageous in some circunstances. For example, the supplementation of a poultry diet containing rye and wheat using a combination of xylanase and β-glucanase results in a significant increase in body-weight and feed intake (Pettersson & Aman, 1989) when compared with the use of a single exogenous enzyme. However, a considerable number of cases have been described showing a lack of response to enzyme supplementation due, in some circunstances, to the low efficiency of the exogenous enzyme mixtures. For barley- and oat-based diets for non-ruminant animals, β-glucanases or cellulases appear to be beneficial enzymes, while the xylanases, or more specifically the endoxylanases, are more appropriate for diets rich in wheat, triticale and rye. It is crucial to

ensure that the exogenous enzyme preparations have the appropriate activities to target the hydrolysis of the predominant NSPs present in the diet. The effect of the exogenous enzymes is influenced by the type and concentration of NSPs as well as by the origin of the cereals and age of poultry. The evolution of the digestive capacity with the animal age influences the response of birds to enzymatic supplementation (Marquardt *et al.*, 1996). Furthermore, the activity of exogenous feed enzymes must not be affected by feed processing, such as expansion or pelleting, which may involve raising feed temperatures above 95°C, or by the low pH (<4) of the GI tract. Moreover, exogenous enzymes need to resist to inactivation by endogenous proteases present in the GI tract. The development of highly thermotolerant CAZymes that retain significant activity at mesophilic temperatures would be significantly advantageous in the majority of industrial processes involving feed supplementation (Bedford, 2000; Munir & Maqsood, 2012). Therefore, there are several aspects related with the activity of exogenous enzymes *in vivo* that await clarification, including those related with their mode of action, the selection of the best activities to different feedstuffs and the understanding of the mechanisms of resistance of the exogenous enzymes to the variable environment of the animal's gut. Exploring correct enzyme cocktails is crucial in order to formulate more efficient and economically viable diets that use a wider range of ingredients in feed formulation (Khattak *et al*., 2006). Following the above discussion, it is also believed that cellulosomes may open novel perspectives for feed supplementation with exogenous enzymes. Cellulosome architecture provides a biological blue print to design more efficient enzymatic complexes that synergistically combine multiple cellulases, hemicellulases and CBMs with different specificities in order to achieve higher activities against plant cell wall carbohydrates when compared with the “free” enzyme systems (Fontes *et al*., 2004; Guerreiro *et al*., 2008).

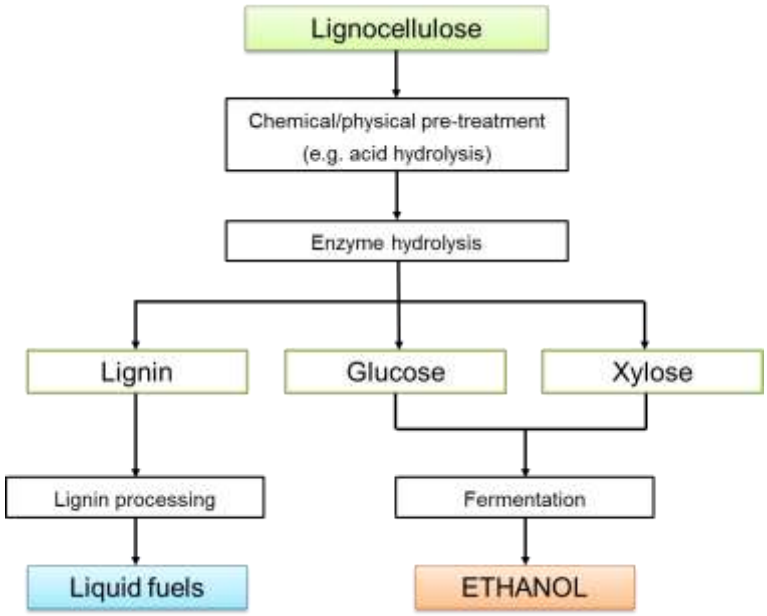
2.2.2. Bioenergy production from lignocellulosic materials

Plant cell wall biodegradation, an important step of the carbon cycle, is also important in several agricultural and waste treatment processes. CAZymes are the key to produce first (maize, sugar cane) and second-generation (cellulosic materials) biofuels and other bioproducts to replace depleting fossil fuels (Percival *et al*., 2006). Various cellulosic feedstocks, such as agricultural residues, wood residues, specifically grown crops, as well as municipal and industrial wastes may be used as substrates for the production of bio-fuels (Aristidou & Merja, 2000). These waste residues are usually highly abundant and, in some regions, may be available at low costs.

The production of second-generation biofuels from plant cell-wall materials requires the degradation of structural polysaccharides (cellulose and hemicelluloses) to monosaccharides and their subsequent conversion to ethanol in a fermentation process mediated by yeast or other ethanol producing organism (Schubert, 2006) (Figure 2.11). Nevertheless, plant cell

walls are highly recalcitrant to hydrolysis either by physical means or by the action of microbial hydrolases. As described previously, cellulose forms a tightly packed crystalline structure resistant to degradation and, in addition, it is protected by hemicelluloses and lignin that reduce the access of cellulases to their target substrates. As such, many strategies have been developed to overcome the recalcitrance of natural lignocellulosic materials (Pauly & Keegstra, 2008). During the last decades, there has been increasing interest in enhancing the enzymatic hydrolysis of these materials, with several studies exploring novel strategies to achieve an efficient and cost-effective pre-treatment and develop novel enzyme mixtures to lignocellulosic conversion (Mosier *et al*., 2005). Cheaper and/or more efficient cellulases and hemicellulases mixtures that can transform lignocellulosic materials to reduce the harshness of chemical pre-treatment and/or enzyme costs would be advantageous to the emerging biofuel industry (Aden *et al*., 2002). Cellulase production at large scale is relatively expensive and to generate a more efficient biorefinery process, the strategies to develop cellulase-based mixtures must include: increased volumetric productivity of commercial enzymes, the production of enzymes using cheaper substrates, the production of enzyme preparations with greater stability for the specific processes, and production of cellulases with higher specific activity on solid substrates (Percival *et al*., 2006).

Figure 2.11 | The different steps involved in the bioconversion of lignocellulose to ethanol.



Lignocellulose is hydrolyzed by a combination of acid and enzymatic processes involving delignification to liberate cellulose and hemicelluloses from lignin. The monosaccharides produced are subsequently converted to ethanol by appropriate microorganisms. Adapted from Aristidou & Merja (2000)

With the current increase in the sequencing information and the observed advances in metagenomics, microbial communities and their genomes have become a major resource for bioprospecting. This is the case of thermophilic cellulolytic microorganisms, which produce

highly thermostable enzymes ideal for a large range of biotechnological applications such as the production of biofuels. For this application the anaerobic bacterium *C. thermocellum* may have a particular interest for several reasons. Besides having a highly efficient cellulase system, resulting from the production of cellulosomes, *C. thermocellum* anaerobiosis is advantageous since the provision of oxygen for cellulase production using the current fungal systems is very expensive. In addition, growth at high temperatures facilitates the recovery of ethanol (Lynd *et al*., 1991). A co-culture of *C. thermocellum* (cellulase producer) and *Clostridium thermosaccharolyticum* (cellobiose user and ethanol producer) was shown to have great potential for the production of ethanol from a cellulose based substrate (Demain *et al*., 2005). In addition, cellulolytic and/or saccharolytic microorganisms could be engineered to improve their capacity to produce ethanol from lignocellulose biomass (Shaw *et al*., 2008; Fontes & Gilbert, 2010).

Different strategies related with the manipulation of plant cell wall composition have also been proposed to overcome the recalcitrant nature of lignocellulose, which could sustain the creation of a cost-effective biofuel industry. Pauly & Keegstra (2008) reported several strategies to manipulate cell-wall carbohydrate composition to increase its solubility and hence the access of enzymes to the polysaccharide linkages which could result in an increased number of monomer units available for fermentation by yeast (hexoses, such as glucose and mannose, rather than pentoses). Genetic manipulation of plants to express cellulases and other plant-cell-wall-degrading enzymes has also been reported. Tayler and co-workers (2008) reported several examples of microbial glycoside hydrolases genes expressed in plants for applications related with the conversion of biomass to biofuels (Taylor *et al*., 2008).

2.2.3. Mini-cellulosomes for biotechnological applications

The artificial design of multiprotein complexes with selected enzymes appropriately directed to a particular biotechnological application is a feasible challenge using the cellulosome concept as template. Several studies have been reported aiming the construction of cellulosome chimeras, or mini-cellulosomes, in which appropriate dockerin-containing cellulolytic enzymes are precisely assembled into a mini-scaffolding protein (Fierobe *et al*., 2001; Fierobe *et al*., 2002). These mini-cellulosomes, containing an optional cellulose-binding module, were shown to exhibit enhanced synergetic activities on crystalline cellulose when compared with mixtures of “free” cellulases (Fierobe *et al*., 2001). Nevertheless, they still have lower capacity to degrade cellulose when compared with naturally occurring cellulosomes (Fierobe *et al*., 2002). Therefore, future research should improve the efficiency of these artificial nanomachines by incorporating a larger repertoire of cellulolytic and hemicellulolytic enzymes. Many reports indicate that, within the same species, cohesins present in a primary scaffoldin do not differentiate the dockerin connected to the cellulosomal

catalytic components and thus different cellulosomes have a different composition (Yaron *et al*., 1995; Lytle *et al*., 1996). However, the cohesin-dockerin interaction is apparently highly species-specific, i.e. *C. thermocellum* dockerins do not bind to *C. cellulolyticum* cohesins while *C. cellulolyticum* dockerins do not interact with *C. thermocellum* cohesins *(*Pagès *et al*., 1997). Thus, incorporation of dockerin-containing enzymes from different microorganisms into a chimeric scaffoldin requires appropriate divergent cohesins. This premise should be taken into account when incorporating selected enzymes in precise locations of the mini-cellulosomes (Fierobe *et al*., 2002). The presence of suitable CBMs in these chimeras could enhance the efficiency of selected enzymes by targeting the entire complex to its predominant substrate (Boraston *et al*., 2004).

In theory, a mini-cellulosome can be optimized for any particular biotechnological application that might benefit from enzyme proximity. For instance, supplementation of barley-based diets for simple stomach animals could benefit from the organization of glucanases and cellulases in highly effective mini-cellulosomes. In addition, designing cellulosomal chimeras by employing novel and improved cellulases and hemicellulases for breaking biomass into fermentable sugars could significantly reduce the cost of fuels production from lignocellulosic biomass. It is clear that the molecular building blocks of the cellulosome (cohesins and dockerins) can be used to integrate different types of enzymes, with different specificities, into a macromolecular complex protein in order to orchestrate a particular function(s). Besides cellulases or other plant cell wall-degrading enzymes, cellulosomes could also be used to assemble enzymes expressing other activities, which might benefit from close proximity. It is also possible to introduce the genetic capacity to synthesize cellulosomes into non-cellulosomal microorganisms (Fontes & Gilbert, 2010).

2.3. High-throughput protein expression in the post-genomic era

2.3.1. Genomic and metagenomic sequence information

DNA sequencing technology has undergone tremendous progresses since the establishment of the Human Genome Project (HGP) initiated in 1990 (http://www.genome.gov/10001772). Scientists started applying computing solutions to biology and specifically to genetics and DNA sequencing, developing innovative methods that made possible the task of generating and handling an enormous amounts of information, such as the Whole-genome shotgun (WGS) sequencing (Weber & Myers, 1997; Staden, 1979; Sanger *et al*., 1982). By the time the HGP was being finished, novel sequencing technologies were developed which improved the sequencing efficiency tremendously. Next generation sequencing (NGS) technologies are becoming routine techniques which allow the inexpensive characterization of genomes and metagenomes. Several genomic sequencing projects are presently ongoing, including

the sequencing of viruses, bacteria, fungi, plants and animals. Bacteria continue to be the most attractive microorganisms to develop several biological functions, as they represent a major proportion of life's genetic diversity. It is estimated that there are $4\text{-}6\times10^{30}$ prokaryotes on Earth today occupying all ecological niches and playing critical roles in all ecosystems (Whitman, Coleman, & Wiebe, 1998). Advances in sequencing technologies have made bacterial genetic information more accessible (Mavromatis *et al*., 2012), with data being applied to both research and clinical studies and allowing to understand how bacteria have evolved to play a diversity of physiological functions. In addition, advances in metagenomics allows the analysis of bacterial genetic material recovered directly from environmental samples, which can be directly accessed overcoming the need of microbial cultivation (Pace *et al*., 1986). Significant efforts are being made to sequence metagenomes (i.e. collective genomes from all microorganisms colonizing a given habitat) of different ecological niches, usually employing NGS approaches (Li *et al*., 2009). Thus, recent high-throughput sequencing technologies allowed the development of several metagenomic projects that are either still ongoing (Genomes OnLine Database) or already completed (Integrated Microbial Genomes with Microbiome Samples).

Both genomes and metagenomes can be targeted for prospecting novel enzymes, such as biocatalytic enzymes (e.g. GHs) for industrial use and biofuel production. Target genes can be discovered from genomic and metagenomic databases using bioinformatics tools, subsequently amplified by the Polymerase Chain Reaction (PCR) with specific primers, or synthetically produced, and cloned into appropriate expression systems. Alternatively, two complementary approaches can be used to discover novel enzymes from metagenomes: function-based screening of expression libraries, in which expression libraries are constructed and screened for target enzyme activities, and sequence-based gene searches, in which target genes are directly amplified from metagenomic DNA (Li *et al*., 2009).

2.3.1.1. Sequencing cellulosomal microorganisms: organization of cellulosomal genes in the genomes

The genome sequences of *C. thermocellum*, *C. acetobutylicum*, *R. flavefaciens* and *C. cellulolyticum* are already available, providing a complete and global analysis of all cellulosomal molecular components, including the characterization of cellulosome modular architecture, organization, regulation and assembly (Fontes & Gilbert, 2010; Munir *et al*., 2014). Taking in consideration the biotechnological use of cellulosomes in a variety of applications that could benefit from improved stable enzymes and/or multi-enzymatic complexes, the availability of genetic information related with cellulosomal organisms is of highly importance. The complete sequencing of cellulosomal microorganisms allows the discovery of novel CAZymes and also CBMs, which could contribute to our understanding on the mechanisms of complex carbohydrate hydrolysis. In addition, understanding how

cellulosomal genes are organized and expressed will allow modulating the remarkable cellulosomal complexes through different genetic engineering approaches.

Initial studies on *C. thermocellum cipA* DNA sequence, which encodes the primary scaffolding protein produced by this bacterium, revealed that this gene is part of an operon containing other genes. Gene walking allowed the cloning and subsequent sequencing of all the genes located in the *cipA* operon (Fujino, Beguin & Aubert, 1993; Lemaire *et al*., 1995). Notably, all the proteins encoded by these genes have cohesin domains. For bacteria expressing a single primary scaffoldin, such as *C. thermocellum*, the primary scaffolding gene (*cipA*, in the case of *C. thermocellum*) is clustered together on the genome with one or more anchoring scaffoldins. Genes encoding the catalytic cellulosomal units are distributed elsewhere on the genome either alone or in small clusters (Lemaire *et al*., 1995). *C. thermocellum cipA*, *olpB*, *orf2* and *olpA* are located in tandem, whereas *sdbA*, a highly expressed gene encoding an anchoring scaffoldin, is located in another place in the genome (Bayer *et al*., 2004).

The scaffoldin gene cluster of *R. flavefaciens* consists of genes encoding four cohesin-containing scaffoldins of different sizes (*scaC*, *scaA*, *scaB* and *scaE*) and one additional gene, *cttA*, encoding a distinct cellulose-binding protein. Although the general design of *sca* gene cluster from strains FD-1 and 17 are identical, there are considerable differences in the modular architecture of ScaA and ScaB, implying strain-specific divergence in cellulosome organization in *R. flavefaciens*. This strain heterogeneity may reflect the intricacy and diversity of the lignocellulosic substrate found in the rumen (Jindou *et al*., 2006). More extensive analysis of the *R. flavefaciens* FD-1 genome indicated that it harbours the largest number of dockerin-containing components known so far in a single organism - at least 223, (Rincon *et al*., 2010). For example, the genome of *C. thermocellum* contains only about 73 dockerin-carrying cellulosome proteins (Bayer *et al*., 2004), while the mesophilic *C. cellulolyticum* contains 71 putative dockerins (Rincon *et al*., 2010). As the presence of a dockerin module in a protein may serve as a signature sequence for cellulosomal proteins (Zverlov, Kellermann, & Schwarz, 2005), the number of dockerin-containing units found in *R. flavefaciens* can predict the complexity, and potential versatility, of its cellulosome. Of the over 200 dockerin-containing proteins annotated from *R. flavefaciens* genome, most of them are still of unknown function. Furthermore, Fontes & Gilbert (2010) suggest that there is a variety of potential substrate specificities displayed by the enzymes bearing homology with glycoside hydrolases, carbohydrate esterases, and polysaccharide lyases of known function identified in cellulosomes.

According to Davies, Gloster & Henrissat (2005), approximately 1-3% of the proteins encoded by the genomes of most organisms correspond to CAZymes. GH-coding genes are the most abundant, corresponding to almost half of the enzymes classified in the CAZy database, followed by glycosyl transferases representing about 41% of the CAZy entries.

Polysaccharide lyases and carbohydrate esterases represent a minor portion of the CAZy content, roughly 1.5% and 5% of the total numbers, respectively (Cantarel *et al*., 2009). It is evident that the sequence-based classification of CAZymes constitutes an excellent guide for the efficient annotation (e.g. prediction of the general function, fold and catalytic mechanism) of open reading frames (ORFs) found during genome sequencing.

2.3.1.2. Understanding cellulosomal genes: expression and regulation

In most cellulolytic organisms, expression of cellulase genes is repressed in the presence of easily metabolized soluble carbon sources and induced when cellulose is present. Significantly, induction of those genes appears to be due to the occurrence of soluble products originated during the initial steps of cellulose hydrolysis by enzymes synthesized at low levels through constitutive expression. These products are probably converted into true inducers by transglycosylation reactions (Béguin & Aubert, 1994).

In the simplest cellulosome systems, such as the one found in *Clostridium cellulovorans,* the scaffoldin gene is located in an operon with a series of genes encoding dockerin-containing enzymes, being several of these genes co-expressed as polycistronic mRNA (Han *et al*., 2003). It was shown that most of the cellulosomal genes of *C. cellulovorans* are expressed at high levels when cells are grown on natural polymers, such as cellulose, xylan and pectin. In contrast, expression of these genes is repressed during growth on cellobiose or fructose, and repressed even further on mannose and lactose (Han *et al*., 2004).

Regarding to complex cellulosome systems, such as the ones found in *C. thermocellum* and *R. flavefaciens*, the scaffoldin genes are organized into scaffolding gene clusters, as reported above. It was shown that mRNA levels of cellulosomal components from *C. thermocellum* are regulated by carbon source and/or growth rate and changes in one or the other factor will be reflected in the overall level of cellulase production (Zhang & Lynd, 2005), as well as in the cellulosomal subunit profile (Bayer *et al*., 1985; Freier, Mothershed, & Wiegel, 1988; Morag, Bayer & Lamed, 1990). Recently, a proteomic analysis was carried out for two strains of *C. thermocellum*, ATCC 27405 (Gold & Martin, 2007) and F7 (Zverlov *et al.*, 2005). Some common proteins were identified as highly abundant in both strains (e.g., CipA, CelS, CelK, XynC, XynZ, CelA, CelR, and CbhA), although there are some relevant differences concerning the expression of cellulosomal proteins in these two bacteria. CelE, CelJ and Cthe_0821 (see Chapter 5 of this thesis) are major components of the cellulosome strain ATCC 27405 but not in strain F7. The overall cellulase genes expression is higher when cells are grown on cellulose rather than cellobiose. Analysis of *R. flavefaciens* FD-1, revealed that *scaA*, *scaB* and *scaC* are co-expressed either as a polycistronic mRNA or sharing the same regulator with similar affinity for these genes. *cttA* and *scaE*, do not appear to be co-expressed with the other components of the *sca* gene cluster. Comparing the expression patterns of several putative enzymes from *R. flavefaciens* FD-1 with those from *C.*

thermocellum, it is suggested that differences in the expression profiles found are likely due to the different environmental conditions to which these two bacteria are exposed, including oxygen concentrations and the type of substrate available for hydrolysis (Miller *et al*., 2009).

The genome sequence of *C. thermocellum* allowed the identification of various proteins that may be involved in the regulation of the cellulosomal genes on the basis of sequence homology. For instance, a gene homologue to that encoding the catabolite control protein CcpA was identified in the genome of this bacterium (Bayer *et al*., 2004). CcpA is a negative regulator that mediates catabolite repression in gram-positive microorganisms and thus allow the bacterium to utilize carbon sources in a strictly controlled hierarchical manner. Other potential negative regulator, GlyR3, was reported by Newcomb *et al*. (2007). GlyR3 displays sequence homology with LacI of *Escherichia coli* and binds to the non-cellulosomal *celC* gene cluster by preventing its expression. The binding is inhibited by laminaribiose, suggesting that the sugar serves as an inducer of *celC* and/or the gene cluster encoding enzymes acting on β-1,3-glucans (Newcomb, Chen, & Wu, 2007). On the other hand, at the level of protein protection and stabilization, Kang *et al*. (2006) showed that members of the serpin superfamily of serine proteinase inhibitors, which have pivotal regulatory functions in eukaryotes, also reside within the cellulosome of *C. thermocellum* ATCC 27405. Two of the three serpins found in this bacterium contain a dockerin module for location in the cellulosome (Kang *et al*., 2006). The role of serpins was suggested to be to protect *C. thermocellum* from the proteolytic degradation by endogenous or exogenous serine proteases (Eggers *et al*., 2004; Schwarz & Zverlov, 2006). Schwarz & Zverlov (2006) speculate that protease inhibitor/protease pairs in cellulosomes play hitherto unknown roles in protein stability and regulation. All these studies expand the cellulosome paradigm of protein complex assembly beyond CAZymes and CBMs.

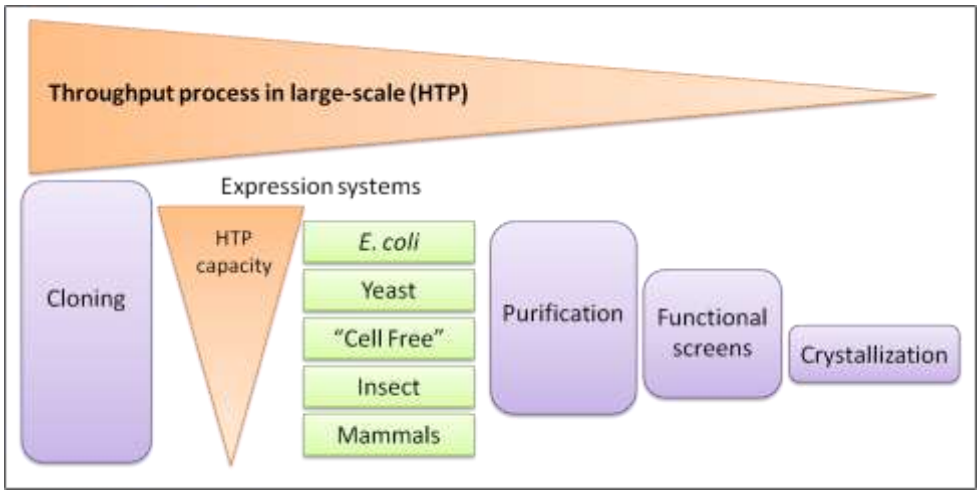
2.3.2. Post-genomic strategies: protein research

As more and more genome and metagenome sequencing projects become complete, scientists are faced with the task of functionally analyze a multitude of gene products. Computational analysis of genomes using the basic local alignment search tool (BLAST) (Altschul *et al*., 1997) attempts to assign functions for the majority of predicted proteins encoded by sequenced genomes. Detection of amino acid sequence similarities to proteins of known function allowed the annotation of 40-70% of novel genome sequences by homology; the larger percentage being for the well-studied prokaryotes (Eisenberg *et al*., 2000). Nevertheless, the proportion of proteins without reliable functional annotation is large. Furthermore, extensive analysis of the functional annotated proteins is necessary to establish a well-characterized model as similarity studies are not sufficient to establish a clear identification of the biological role of genes and proteins. Thus, structural and biochemical characterization of proteins (e.g. mode of action, ligand/substrate specificity,

ligand/substrate binding, regulation and factors affecting their function, interaction with other molecules inside or outside the cell and metabolic or signaling pathways in which they are involved) is required even for a large proportion of already annotated enzymes. Notably, there is an exponential increase in the identification of CAZymes, which has resulted in a significant discrepancy between the number of enzymes in CAZy families and the biochemical/structural data of these proteins. Indeed, only around 3% of proteins in the CAZy database have a characterized biochemical activity, and only mere 0.3% of these enzymes have three-dimensional structural information available (Cantarel *et al*., 2009). Thus, functional, biochemical and structural analysis is urgently required in order to integrate information obtained from the genomic technologies and to understand the biology of plant cell wall deconstruction in a diversity of ecological niches.

Structural genomics and proteomics offer the promise of assigning a biological function to all the proteins encoded by the genome of an organism. The optimal utilization of genomic sequence data by genome-scale protein work requires, however, rapid and efficient methods for the generation of expression clones and the evaluation of protein production, thus leading to the rapid protein characterization and structure determination. High-throughput (HTP) approaches of the post-genomic era require the implementation of novel methods for cloning, expression, protein purification and detection that allow working with large numbers of genes and proteins, and at the same time to analyze a multitude of data and results. Even for laboratories studying a single protein target, these steps are usually expensive and time-consuming. Solutions to overcome these problems have emerged from structural genomics projects, which use a standard experimental workflow and a "funnel" approach (Figure 2.12). Each stage of the process needs to be optimized on parameters like throughput, automation, speed, cost-effectiveness and scalability (Mancia & Love, 2011).

Figure 2.12| A typical "funnel" scheme for high-throughput structural studies on proteins.



The throughput of the entire process decreases from cloning to crystallization. The last step (not showed) involves progression from crystallization to structure determination and, in general, cannot be automated. Throughput also diminishes as the complexity of the expression systems increase. Adapted from Mancia & Love (2011)

2.3.2.1. High-throughput methods for gene cloning

Methods for the rapid and inexpensive cloning of large numbers of open reading frames (ORFs) into expression vectors are of critical importance for challenging structural and functional genomics projects. According to Marsischky and LaBaer (2004), an ideal HTP cloning method should be reliable, ease to use, flexible and inexpensive. Thus, several properties should be contemplated by these innovative methods: (1) when transferring the cloned DNA from master clones to expression plasmids, the transfer must be 100% (or almost) efficient, conservative, thereby avoiding mutations, and should result in the correct orientation of ORFs, (2) validation of the cloned products should be simple; ideally only a single clone for each target gene would be required to be sequenced and the cloning vector must be suitable for DNA sequencing reactions, (3) the cloning system should be able to support the transfer of genes into virtually any type of expression vector, (4) the addition of fusion tags or cloning related sequences to an ORF should be minimal since they often affect the subsequent expression and crystallization of the recombinant protein (Marsischky & LaBaer, 2004). Furthermore, such a HTP cloning method should be independent of the sequence of the target gene, primers should be easy to design, and a single PCR amplification should be sufficient for cloning into several vectors (Geertsma & Dutzler, 2011). Conventional cloning methods based on DNA cleavage by restriction endonucleases and then ligation by DNA ligase ("cut and paste") are, in this respect, unsatisfactory. The traditional cloning strategies based in restriction enzymes and ligases are relatively inefficient, time-consuming and labour-intensive. Moreover, if an uniform cloning strategy for a large number of different ORFs is required, the use of this method is prevented by the frequent occurrence of restriction sites in the sequences of the target genes. To overcome the limitations of ligation-based cloning, alternative approaches have been developed based on ligation-independent cloning (LIC) procedures (Table 2.1). The majority of LIC methods developed so far are based on a recombination reaction occurring between the insert and the destination vector. These technologies include, for example, the Gateway (Hartley, Temple & Brasch, 2000), the Creator (Colwill *et al*., 2006), and the MAGIC (Li & Elledge, 2005) systems, that make use of site-specific recombination, and the In-Fusion (Berrow, Alderton, & Owens, 2009) or sequence- and ligation-independent cloning (SLIC) (Li & Elledge, 2007), that rely on homologous recombination. Other LIC technologies use complementary single-strand overhangs on the vector and insert, which allows cloning by base complementation without the need of a ligation step. These methods include, for instance, the LIC method based on T4 DNA polymerase (Tachibana *et al*., 2009).

Table 2.1| Cloning methods for protein expression.

Cloning method	Cloning sites	Size of cloning sites	Comments	Ref.
Gateway (Invitrogen)	*attB* sites	21-25 bp	Large cloning sites Initial cloning into non-expression vector Trivial to subclone into many available vectors 1-hour room temperature cloning reaction	1
Topo (Invitrogen)	Trapped topoisomerase	9-16 bp	For cloning of PCR products (5'-OH required) Cloning of long DNA inserts Vectors must be purchased 5-min cloning reaction	2
Creator (Clontech)	*loxP* sites	34 bp	Large cloning sites Initial cloning into non-expression vector Easy to subclone into available vectors C-terminal fusions must undergo splicing 15-min room temperature cloning reaction	3
In-Fusion (Clontech)	Homologies to vector ends	14-15 bp	For cloning of PCR products Compatible with any vector Vectors must be linearized 15-min at 50°C cloning reaction	4
LIC (Novagen)	Single-stranded homologies to vector ends	12-15 bp	Sequence-dependent Many available vectors Prior insert treatment with T4 DNA polymerase 10-min cloning reaction	5
MAGIC	Homologies to vector	~50 bp	Reactions occur inside host cells – *in vivo* Uses bacterial mating	6

References: 1) Hartley, Temple & Brasch (2000); 2) Shuman (1994); 3) Colwill *et al*. (2006); 4) Berrow, Alderton, & Owens (2009); 5) Tachibana *et al*., (2009); 6) Li & Elledge (2005).

The Gateway® recombinational cloning system (available from Invitrogen™ Life Technologies) exploits the *in vitro* site-specific recombination method based on bacteriophage λ to accomplish the directional cloning of PCR products and the subcloning into new vector backbones. Many DNA segments can be transferred in parallel from entry clones (or donor plasmids) to various destination vectors (or expression plasmids), providing an approach to HTP cloning and subsequent protein expression. In the Gateway® system, the orientation and reading frame register of cloned DNA are maintained through vector transfers by the use of two nearly identical (but not compatible) versions of the λ *att* recombination site (Hartley, Temple, & Brasch, 2000). However, the long extraneous coding regions (around 21 bp) may affect downstream applications (Koehn & Hunt, 2009).

The In-Fusion™ system (available from Clontech) is a simple and reliable method that enables the rapid cloning of PCR products using the In-Fusion™ enzyme. The mechanism of this reaction has not been fully reported but relies on the presence of homology between

extensions on the PCR product (around 15 bp) and the ends of a linearized vector. One advantage of the In-Fusion™ system is that it is not sequence-dependent. Thus, it can be used as a general method for inserting DNA fragments into any vector.

The LIC system based on T4 DNA polymerase (available from Novagen) takes advantage of the controlled exonuclease activity of T4 DNA polymerase to produce long single-stranded complementary overhangs (roughly 12 to 15 bp) at the ends of a PCR-amplified DNA fragment and a linearized vector. These overhangs hybridize after mixing the vector and insert together, producing a stable hybridization product that can be readily used to transform host organisms without ligation. After transformation, the nicks are sealed by the cellular machinery (Aslanidis, de Jong, & Schmitz, 1994). However, LIC based on T4 DNA polymerase is strictly sequence-dependent as it requires the presence or absence of specific nucleotides at certain positions in the overlapping regions, which restricts its widespread application to different vectors (Tachibana *et al*., 2009). In general, sequence-dependent methods are less convenient in a HTP practical setting because they require unique and specific sites in both the insert and the vector. Thus, the more flexible sequence-independent cloning methods are preferred. Other critical points reported in LIC methods relate to the additional steps for engineering or preparation of insert and/or vector DNA fragments that are sometimes time-consuming and usually use expensive enzymes (Quan & Tian, 2009).

One important issue to take into account in HTP cloning relates to the selection of the correct clones, which is usually labour-intensive and often limited by parental vector background. Many positive selection vectors have been developed to address this problem. In these vectors, the successful cloning of a target gene results in a change of the phenotype, which can be achieved either by the inactivation or the replacement of a gene marker (e.g. the lethal gene *ccdB*) through insert cloning (Haag & Ostermeier, 2009; Hu *et al*., 2010). Despite the reasonable effectiveness of the currently available LIC systems for HTP cloning, there are considerable pressures for protocol simplification and reduction in costs that could make them a more appropriate alternative for the rapid and efficient creation of large ORF clones resources. Thus, the research community has been improving the current methods or developing cheaper and more reliable approaches than those today commercially available (Bryksin & Matsumura, 2010; Yang *et al*., 2010; Geertsma & Dutzler, 2011). This thesis reports the development of an innovative HTP platform for gene cloning, expression and protein purification that was used to explore genes encoding cellulosomal components of unknown function (see Chapter 3). The selection of the method used to clone a gene into an expression vector is of great importance since proteins are highly diverse in their properties and it is not possible to predict how they will express or whether they will be soluble, easy to purify, display biological activity or crystallize in the recombinant form (Hartley, 2006).

2.3.2.2. High-throughput methods for protein expression

There are a large variety of expression systems available for the high-throughput production of recombinant proteins. *Escherichia coli* is the simplest host for recombinant protein expression and has been the most readily adapted expression system to the HTP format. Expression in *E. coli* is fast, cost-effective and scalable (Swartz, 2001; Chambers, 2002; Mancia & Love, 2011). Furthermore, there are many available engineered strains and vectors for heterologous expression, in which high levels of controllable gene expression are usually achieved. The most common promoter used to favour gene expression is the T7 promoter that is usually regulated by lactose or in the majority of the situations by the non-hydrolysable lactose analogue isopropyl β-D-1-thiogalactopyranoside (IPTG). The use of *lac*I^Q, T7 lysozyme (provided by a compatible plasmid, pLysS or pLysE) (Moffatt & Studier, 1987) and *lacO* operators, positioned downstream of the T7 promoter sequence, supports the establishment of the hybrid T7/*lac* promoter, (Dubendorf & Studier, 1991) and provide extra levels of expression control, to address problems related with the production of toxic proteins (i.e. ribosome destruction, cell death and plasmid or expression instability) (Koehn & Hunt, 2009). Nevertheless, for many complex and large eukaryotic proteins subjected to complex post-translational modifications, expression in *E. coli* is usually impaired and eukaryotic expression systems arise as a good choice for the production of biologically active polypeptides. However, in this case, usually the throughput of protein expression decreases, accompanied by low yields and high costs (Hartley, 2006). Despite technical constraints, several mammalian (Heyman *et al*., 1999), yeast (Schuster, 2000) and insect cells (Albala *et al*., 2000; Coleman *et al*., 1997) largely used as expression systems have been successful adapted and implemented for HTP protocols. Cell-free expression systems have also emerged as an alternative to the cell-based expression systems for recombinant protein expression, especially for producing toxic and insoluble proteins. This approach involves *in vitro* protein expression using cellular lysates that contain all the necessary biochemical machinery for the transcription and translation of DNA templates. Many features make this system useful in a HTP setting as it allows expressing proteins directly from PCR-generated transcripts with no need to generate expression vectors or manipulate cells in culture (Martemyanov, Spirin, & Gudkov, 1997). There are several commercial cell-free systems based on *E. coli* lysate, among others, that can be used to produce proteins *in vitro*. The main disadvantages of cell-free systems are the low yield of target proteins and the limited post-translational modifications if those are based in *E. coli* (Carlson *et al*., 2011). Alternatively, *E. coli* can be adapted for recombinant protein production through genetic engineering.

Implementation of an HTP platform for cloning and expression of hundreds to thousands of targets at the same time, requires significant modifications to the *modus operandi* of traditional cloning and expression protocols. There are increasing numbers of reports

suggesting methodologies for the automated generation of protein expression clones using microwell plates, multichannel pipettors and, in some cases, commercially available liquid handlers (Figure 2.13), with a series of interlinked protocols based on the 96-or 384-well plates formats (Dieckman *et al*., 2002; Scheich, Sievert, & Büssow, 2003; Abdullah, Joachimiak, & Collart, 2009; Bruni & Kloss, 2013). The conventional cell cultivation systems using shake-flasks, even small, is not practical and the solution is to use deep-well plates available in different formats (24-, 48- and 96-, wells) and volumes (1, 2 and 5 mL per well) covered with breath seals which allow efficient gaseous exchanges. Chambers (2002) showed that growth profiles of cell cultures from *E. coli* and insect expression systems are the same over time using a shake-flask or a commercial deep-well block (Chambers, 2002). With this approach it is possible the simultaneous and parallel cultivation over a wide range of conditions (e.g. temperature, time, media, host) to screen for optimum expression conditions of soluble recombinant proteins. With optimal growth and induction conditions established, these protocols can easily be scaled up to increase the capacity of protein production (Bruni & Kloss, 2013).

Figure 2.13| Liquid-handling workstations equipped with vacuum manifolds.

An increasing number of molecular biology laboratories cultivate *E. coli* cells during protein expression protocols in auto-induction media, first described by Studier (2005). These media allow the simultaneous induction of expression of multiple recombinant proteins under the control of a T7/*lac* promoter. These media contain specific components that after an initial period of tightly regulated growth allow fully automated induction of target protein at high optical densities without the need for either monitoring growth or induction with IPTG (Studier, 2005). Even using 96-well plates, high cell densities can be achieved (Lesley *et al*., 2002). Different components of complex media are reported to support or suppress growth to high cell density of a wide range of *E. coli* strains with different nutritional requirements (Studier, 2005). Lactose can support growth of *E. coli*, but several restrictions are reported when it is used as a carbon and energy source for high-level production of target protein (e.g. lactose's catabolism leads to the production of galactose which is not used as a carbon

source by BL21, the most common *E. coli* protein expression strain) (Studier, 2005). Thus, cultures growing in media containing glucose and lactose will utilize all glucose before starting to metabolize lactose (following a catabolite repression mechanism). To prevent lactose metabolism, a control of gene expression relies on the binding of the CAP-cAMP complex to the promoter required for the transcription of the *lac* operon. The presence of glucose is associated with the presence of the complex. Once glucose concentration increases inside the cell, the cAMP decreases, as the amount of the complex also decreases. In this situation, the complex does not bind to the *lac* promoter and the *lac* operon is turned off (Epstein, Rothman-Denes, & Hesse, 1975). When lactose is present, it binds to the *lac* repressor protein (encoded by *lacI*), making it unable to bind to the operator. Therefore, lactose acts as the inducer of the transcriptional expression controled by the T7 promoter (Horton, Lewis, & Lu, 1997). There are some auto-induction media solutions sold commercially; being the most used the Overnight Express™ Autoinduction system (Novagen). This thesis reports the development of two novel auto-induction media systems to use in *E. coli* as expression host: NZY Auto-Induction LB medium (powder) and NZY Auto-induction kit, both already available from NZYTech (see Chapter 3). Auto-induction media provide great convenience as recombinant strains only have to be inoculated and grow to saturation, which addresses the difficulty in large-scale screening to obtain all of the cultures in a comparable state of growth for simultaneously induction. Furthermore, this system allows achieving higher yields of soluble protein production typically several-fold higher than those obtained by standard IPTG induction (Studier, 2005).

For an initial and rapid assessment of levels of protein expression, whole and soluble protein extracts can be analysed directly by SDS-PAGE and Coomassie staining without prior purification of the recombinant expression products (Dieckman *et al*., 2002; Abdullah, Joachimiak, & Collart, 2009). However, additional information of the expressed protein is usually required, which will involve cell lysis, protein purification, and analysis of recombinant protein biological activity.

2.3.2.3. High-throughput methods for protein purification

Affinity chromatography is one of the most robust and efficient methods for protein purification. Current recombinant expression systems involve engineering specific affinity tags in the recombinant proteins that allow the rapid and efficient implementation of purification protocols. Purification based on histidine tags (His-tag) is an universal solution for purifying proteins in parallel through Imobilized Metal Affinity Chromatography (IMAC). Thus, His-tags (which bind to immobilized divalent metal ions, Ni^{2+}) are widely used for protein purification and usually provide, in a single-step affinity purification protocol, high levels of purity that are satisfactory for most downstream applications. Furthermore, His-tag sequences require the addition of usually only six amino acids to the recombinant protein,

thus reducing the negative effects that could derive from the fusion of large polypeptide sequences in protein folding/activity (Lesley, 2001). In addition, they also have the advantage of being easily recognized by commercially available antibodies. So, it is not surprising that many metal chelating resins with high affinity for His-tag have been adapted to 96-well format protocols for protein purification using automated systems (Chambers, 2002).

In general, IMAC automatable protocols involve multiple steps. Thus, after cultivation on deep-well plates, cells are collected by centrifugation and are lysed using either mechanical disruption or chemical lysis. Mechanical disruption can be achieved in a HTP procedure using a deep-well microplate-horn sonicator commercially available. Alternatively, chemical lysis has been successfully applied in many HTP protocols since it is fast, robust and less labour-intensive than the mechanical lysis (Chambers, 2002). There are many reagents specially formulated for bacterial cell walls disruption and most of them include lysozyme and treatments with nucleases to degrade DNA and thus decreasing the viscosity of the cell extracts. Also, non-ionic and zwitterionic detergents can be used for nondenaturing lysis (Lesley, 2001). BugBuster® (available from Novagen), with proprietary formulation (Grabski, McCormick, & Mierendorf, 1999) and B-PER® (from Thermo Scientific, Pierce Protein Biology Products) are two of the most common detergent-based reagents available commercially. This thesis reports the development of a reagent for the gentle disruption of *E. coli* cell walls: NZY Bacterial Cell Lysis Buffer, already available from NZYTech, genes & enzymes (see Chapter 3).

For automatable protein purification, large-pore 96-well filter plates are usually required. Thus, after cell lysis, the crude cell lysate is mixed with the nickel charged resin to capture the recombinant protein and then transferred to filtration plates. The system requires the use of vacuum pressure to allow the passage of cell extracts and wash buffers through the plate columns (Scheich, Sievert, & Büssow, 2003). Depending on the expression level and the volumes used, roughly 1.5 µg to 6 mg of target protein can be purified from small-scale *E. coli* cultures, which is enough for expression analysis, evaluating the solubility, and for implementation of initial functional studies (Chambers, 2002). For many screening protocols, on average, 10 mg of protein are typically required, which can be achieved by large-scale (~70 mL) protein purification in an automated way provided by custom robots (Lesley, 2001). Automation in crystallography, with implementation of liquid dispensing workstations (Stewart & Baldock, 1999) and the development of micro-crystallization protocols (Pebay-Peyroula *et al*., 1997), has reduced the amount of concentrated, homogeneous and soluble protein required to crystallize, being 10 mg usually sufficient for these procedures (Lesley, 2001).

2.4. Enhancing recombinant protein expression, folding and solubility in *E. coli*

2.4.1. Recombinant protein expression in *Escherichia coli*

There are several features that make *Escherichia coli* the most widely used host for the production of heterologous proteins: (1) rapid growth at high cell density on inexpensive substrates; (2) cells have short generation times; (3) cells do not require specialized equipment for cultivation; (4) easy manipulation due to its well characterized genetics; (5) and the availability of a large number of molecular tools and protocols, such as cloning vectors with different N- and C-terminal tags, engineered strains and cultivation approaches. In addition, the resulting yields of recombinant proteins can be very high (Baneyx, 1999b; Mancia & Love, 2011; Rosano & Ceccarelli, 2014). Despite the strengths of the *E. coli*-based expression system, there are some difficulties to address when expressing heterologous proteins, mainly from eukaryotic origin, in this host. The main difficulty lies on the production of inclusion bodies that derive from high levels of recombinant protein expression and unappropriate conditions for correct protein folding. Further problems include low levels of expression, protein degradation, toxicity, and production of non-functional protein (Peleg & Unger, 2012). Thus, several different strategies have been developed to produce suitable amounts of recombinant proteins in soluble and biologically active form in *E. coli* on a large scale by increasing both yield and solubility.

In cases where eukaryotic recombinant proteins need post-translational modifications (e.g. glycosylation and phosphorylation) in order to become active and/or adopt the proper structure, expression in *E. coli* may not be suitable. However, several *E. coli* strains have been genetically modified to allow for some post-translational modifications, such as disulfide-bond formation in the cytoplasm by providing oxidizing conditions due to mutations in thierodoxin reductase (*trxB*) or/and glutathione reductase (*gor*) genes in AD494 and Origami™ (Novagen) strains, or by co-production of DsbC proteins in SHuffle® strains (Novagen) (Derman, Prinz & Beckwith, 1993; Bessette *et al*., 1999). Other issue that must be taken into account for recombinant protein expression is the codon usage of the recombinant gene and its adequacy to the expression host. *E. coli*, as all other organisms, uses a specific pool of the 61 available amino acids codons for the production of most mRNA molecules (Wadal *et al*., 1992). Differences in codon usage can cause problems during translation (e.g. interrupted translation, frameshifting and misincorporation of amino acids during translation) due to the request for one or more tRNAs that may be rare or lacking (Kane, 1995; Kurland & Gallant, 1996). Changing growing conditions, such as temperature, media composition and growth rate may vary the tRNA abundance and the codon usage bias. However, most tRNAs corresponding to rare codons (i.e. those that occur in genes expressed at a low level) remain unchanged at different growth rates (Dong, Nilsson, & Kurland, 1996). In addition, there are many strategies available to increase the expression levels of heterologous proteins containing rare codons: (1) gene synthesis to replace the rare codons by major codons that

encode for the same amino acid, and (2) co-expression of genes encoding rare tRNAs. There are several commercial *E. coli* strains, such as the Rosetta™ series (Novagen), that contain a plasmid carrying tRNA genes for a number of rarely used codons in *E. coli*. Finally, some recombinant proteins may be toxic for *E. coli* as they may carry out a detrimental function in the host cell, by compromising the normal proliferation and homeostasis of the microorganism (Rosano & Ceccarelli, 2014). Thus, in these cases, basal levels of expression need to be tigher controlled and different approaches have been developed such as the addition of glucose to the growth medium (Studier, 2005), co-expression of T7 lysozyme (Moffatt & Studier, 1987), such as in BL21pLysS and BL21pLysE strains, or by using low copy number plasmids, such as the pETcoco vectors (Novagen) (Wild, Hradecna, & Szybalski, 2002).

2.4.2. Protein precipitation in the form of inclusion bodies

One of the major bottlenecks for recombinant protein expression in *E. coli* is the production of inclusion bodies (IB). Overproduction of heterologous proteins in the cytoplasm of *E. coli* caused by strong promoters may be accompanied by their misfolding and precipitation into insoluble, biologically inactive form of inclusion bodies (Williams *et al*., 1982). Protein precipitation *in vivo* occurs because the microenvironment of *E. coli*, respecting to pH, osmolarity, redox potential, cofactors, and folding mechanisms differs from that of the endogenous host causing protein instability. Nevertheless, even endogenous proteins can accumulate as IB if overexpressed, what suggests that IB formation is a consequence of an increase in the level of expression of the recombinant protein (Gribskov & Burgess, 1983). Limiting amounts of chaperones when genes are expressed at high levels probably contributes for IB formation (Rinas & Bailey, 1993; Thomas & Baneyx, 1996a; Lorimer, 1996). It has been reported that IB formation is not associated to particular protein sequences. However several attempts to enhance protein solubility by directed mutation or gene fusion have resulted successfully mainly when affecting the extent of hydrophobic regions (Hartley & Kane, 1988; Carrió & Villaverde, 2002). Thus, it cannot be predicted if a certain protein has propensity to form inclusion bodies. However, if a protein containing disulphide bonds is produced in the cytoplasm, the formation of IB can be expected since disulphide bonds usually are not formed in this cell compartment that has a highly reduced environment (Lilie, Schwarz, & Rudolph, 1998). Much effort has been made to overcome the insolubility problem in order to promote properly folded protein, reduce the percentage of inclusion bodies and obtain biologically active proteins. Renaturation of inclusion body proteins is possible by solubilisation of inclusion body material with strong denaturants, followed by the removal of the denaturant through dialysis or dilution. The effectiveness of renaturation depends on the balance between protein aggregation and correct folding. Following these approaches, several conditions must be optimized (e.g. protein

concentration, temperature, pH or ionic strength), which make the process time consuming and unreliable, and should be adapted for each specific protein (Lilie *et al*., 1998). Thus, methods that allow the production of soluble recombinant proteins when overproduced in *E. coli*, and thus preventing the formation of inclusion bodies, are preferable. These strategies can be applied either at the cell-culture, cellular or molecular level.

A traditional approach to reduce protein aggregation is to decrease the cultivation temperature to attenuate protein production, thus providing more time to allow the newly produced recombinant proteins to fold properly. Lower cultivation temperatures also avoid protein degradation due to the reduced activity of bacterial heat shock proteases (Chesshyre & Hipkiss, 1989; Spiess, Beil, & Ehrmann, 1999; Hunke & Betton, 2003) and may enhance the expression of a number of chaperones (Ferrer, Chernikova, and Yakimov, 2003). Temperatures of 15-20 ºC during gene expression induction usually improve the solubility of many eukaryotic and prokaryotic recombinant proteins (Baneyx, 1999a; Gräslund *et al*., 2008; Sahdev, Khattar, & Saini, 2008). Successful expression of eukaryotic proteins in soluble form at 4 ºC during 48-72h in *E. coli* was described by San-Miguel, Pérez-Bermúdez, & Gavidia (2013). However, those extremely low temperatures cause some problems related to replication, transcription and translation rates, as well as with bacterial growth, which can lead to low protein yields (Rosano & Ceccarelli, 2014). The engineered ArcticExpress strains (Agilent Technologies) may in turn overcome this problem due to the expression of cold-adapted chaperonins. Furthermore, some promoters have been engineered for stronger expression of recombinant protein at low temperatures (Vasina & Baneyx, 1997). Another alternative method to obtain large amounts of active soluble proteins may involve the supplementation of the cell culture with various additives. Blackwell & Horgan (1991) showed that growing and inducing cells under osmotic stress in the presence of sorbitol and glycyl betaine caused a great increase in the yield of active recombinant protein and the disappearance of the protein in the form of IB.

Increasing protein solubility may also be approached by taking advantage of the host cellular chaperones, which facilitate proper protein folding by binding and stabilizing unfolded or partially folded proteins and are implicated in driving proteins to their sub-cellular compartments. The periplasm is a particularly attractive destination for eukaryotic proteins since it provides oxidizing conditions for the formation and rearrangement of disulfide bonds which usually contribute to the stabilization of a folded protein conformation (Missiakas & Raina, 1997; Creighton, 1986). It was postulated that the misfolding and degradation of several heterologous proteins which destination is the periplasm can occur because the necessary chaperone(s) becomes limiting (Baneyx, 1999b). Therefore, it is necessary to develop strategies to increase the availability of these folding modulators in *E. coli*. Co-overproduction of the well characterized cytoplasmic ATP-dependent DnaK-DnaJ-GrpE and GroEL-GroES chaperone systems can greatly increase the soluble yields of passenger

proteins by preventing aggregation, both in the cytoplasm and periplasm (Thomas & Baneyx, 1996a; Gragerov *et al*., 1992). The natural set of chaperones can be induced by the addition of benzyl alcohol or heat shock (de Marco *et al*., 2005). Alternatively, there are chaperone systems commercially available, being the most usual the chaperone co-expression plasmids from Takara (Nishihara, 1998; Nishihara *et al*., 2000).

2.4.2.1. Fusion protein technology

At the molecular level it has been shown that the fusion of recombinant proteins with peptide or protein partners to form a chimeric protein, promotes both protein yield and solubility (Terpe, 2003). Examples of popular fusion partners used to enhance protein solubility include: glutathione S-transferase (GST), maltose binding protein (MBP), N-utilization substance protein A (NusA), thioredoxin (Trx), small ubiquitin-like modifier (SUMO), and ubiquitin (Ub) (Table 2.2). More recently, other fusion partners have been proposed, such as the Fh8 (Costa *et al*., 2013). It is well established that besides allowing for a convenient single-step purification approach, several fusion tags can also enhance solubility of the target protein. The mechanism of how fusion tags operate as solubility enhancers remains, however, unclear. Probably, fusion of a stable or conserved molecule to an insoluble recombinant protein may stabilize and promote proper folding of the recombinant protein. On the other hand, fusion tags may act as a nucleus of folding ("molten globule hypothesis") (Englander, 2000; Creighton, 1997). For example, it was shown that MBP possesses an intrinsic chaperone-like activity (Kapust & Waugh, 1999; Rais-Beghdadi *et al*., 1998; Raran-Kurussi & Waugh, 2012). Fusion technology was also shown to be an useful tool for the production of many recombinant proteins that are components of multiprotein assemblies and that often require an interacting protein for correct folding and stability (Gräslund *et al*., 2008). Other advantage of fusion technology lies on the protection of the recombinant protein from degradation; fusion can promotes translocation of the "unwanted" recombinant protein to different cellular compartments avoiding the exposition to proteases. For instance, MBP may be involved in the translocation of proteins to the membrane (Nikaido, 1994). Disulfide isomerases (e.g. DsbA and DsbC) have also been proposed as fusion partners. They may enhance solubility and proper fold of proteins in the non-reducing periplasmic environment or in the cytoplasm, if expressed without their signal peptide (Baneyx, 1999b; Nozach *et al*., 2013). Some fusion partners, such as NusA, SUMO, Trx and Ub require an affinity tag, such as the poly-histidine for protein purification. Otherwise, MBP and GST can serve to purify the recombinant protein by affinity chromatography, as MBP binds to amylose-agarose and GST to glutathione-agarose, respectively (Rosano & Ceccarelli, 2014).

Table 2.2| Main properties of the most common protein fusion tags used in enhancing protein solubility in *Escherichia coli*.

Tag	Residues/ Size (KDa)	Matrix/ Elution	Comments	Ref.
Fh8	69/ 8.0	An affinity tag must be added (usually His-tag)	Small tag; Ca^{2+}-dependent binding to phenyl-Sepharose	1
Trx	109/ 11.7	An affinity tag can be added (usually His-tag)		2
SUMO (Smt3)	101/ 11.6	An affinity tag must be added (usually His-tag)	Cleavage by SUMO Protease 1	3
Ub	128/ 14.73	An affinity tag must be added (usually His-tag)		4
DsbA	208/ 23.1	An affinity tag must be added (usually His-tag)	Introduces disulfide bonds; enables protein solubilization in the periplasm or in the cytoplasm	5
DsbC	216/ 23.4	An affinity tag must be added (usually His-tag)	Isomerization of disulfide bonds; enables protein solubilization in the periplasm or in the cytoplasm	6
GST	211/ 26.0	Glutathione-agarose/ glutathione	GST dimerization and glutathione elution may affect fusion protein properties	7
MBP	396/ 42.5	Cross-linked amylase/ maltose	Large tag; Matrix compatible with nonionic detergents and high salt, but not reducing agents	8
NusA	495/ 54.87	An affinity tag must be added (usually His-tag)	Large tag, may affect properties of fusion protein	9

References: 1) Costa *et al*. (2013); 2) LaVallie *et al*. (1993); 3) Butt *et al*. (2005); 4) (Baker, 1996); 5) Collins-Racie *et al*. (1995); 6) Nozach *et al*. (2013); 7) Smith & Johnson (1988); 8) di Guana *et al*. (1988); 9) Davis *et al*. (1999).

There are several studies comparing the effects of various fusion tags on protein yields and levels of soluble recombinant protein obtained. Hammarström *et al*. (2002), in a study with 27 human proteins, ranked the fusion tags in terms of ability to promote protein solubility as Trx~MBP~Gb1>ZZ>NusA>GST>His_6. In addition, Braun *et al*. (2002), in a study with 32 human proteins, ranked the fusion tags in terms of increased expression and soluble yield as GST~MBP>CBP>His_6. Shih *et al*. (2002) cloned 40 different proteins with eight fusion tags and observed that MBP gave the best overall results in terms of total and soluble expression. Moreover, Dyson *et al*. (2004), in a study with 20 mammalian proteins, ranked the fusion tags in terms of increased soluble expression as Trx~MBP>His_{10}>GST>GFP. In contrast, De

Marco *et al*. (2004) observed that NusA is better than GST at enhancing the solubility and stability of recombinant proteins, while Marblestone *et al*. (2006), in a study with three recombinant proteins, ranked the fusion tags in terms of increased total expression as Trx>SUMO~NusA>Ub~MBP~GST and increased protein solubility as SUMO~NusA>Ub~GST~MBP~Trx. More recently, Bird (2011), in a study with 20 protein targets, proposed the following rank order: SUMO > Trx = MBP > TF > NusA > GST = MsyB > His_6 > HaloTag, while the rank of the fusion tags for protein solubility according to Costa *et al*. (2013) was MBP>NusA>Fh8~Trx>GST~H>His. The inconsistency of the data from these comparative studies suggests that each protein or class of proteins has unique optimal conditions and that fusion tags vary greatly in efficiency.

The correct choice of an apropriate fusion partner should take into account its size, since this parameter plays a critical role in the total yield of the target protein, as well as its effects on the tertiary structure or biological activity of the fused protein (Balbás, 2001). When using small tags, these effects depend on the location and on the amino acids composition of the tag (Bucher, Evdokimov, & Waugh, 2002). Therefore, the removal of the fusion tag must be considered when structural or biochemical studies on the target recombinant protein are required (Balbás, 2001). Cleavage may be performed either chemically (Chong *et al*., 1997) or through enzymatic approaches (Jenny, Mann, & Lundblad, 2003). In the former the tag is removed by treatment of the fusion protein with a chemical reagent in harsh conditions, so its use is mainly limited to purified recombinant proteins obtained from IBs (Hwang, Pan, & Sykes, 2014). The last strategy involves the insertion of a unique amino acid sequence that is susceptible to cleavage by a highly specific protease, such as tobacco etch virus (Tev) protease, factor Xa, thrombin protease, and the SUMO protease (Jenny *et al*., 2003; Blommel & Fox, 2007; Satakarni & Curtis, 2011). In opposition, unfold and precipitation can occur in some proteins after tag removal (Koehn & Hunt, 2009). Additionally, complete cleavage rarely occurs leading to reduction of the target protein yield (Baneyx, 1999b).

In summary, ideally a fusion partner should not compromise the tertiary structure and biological activity of the fused protein, be easy to remove without affecting protein structure after removal, allow simple purification procedures of the recombinant protein, and be applicable to a range of proteins (Terpe, 2003). To fulfill these requirements, new tag-protein fusion systems are constantly emerging for the improvement of the soluble production of recombinant proteins, which is highly relevant to high-throughput protocols. This thesis reports the development of novel gene cloning and expression strategies for the efficient production of soluble recombinant proteins in *E. coli* by investigating novel fusion tags to improve protein solubility (see Chapter 4).

2.5. Objectives

The work presented here aims to exploit the cellulosomal proteomes of *R. flavefaciens* and *C. thermocellum* to discover novel CAZymes that may increase our understanding of the fundamental mechanisms involved in plant cell wall hydrolysis. In addition, these novel CAZymes may constitute innovative tools to explore novel biotechnological applications, in particular in the area of animal nutrition, biofuel production or mini-cellulosome construction. Moreover, the discovered and characterized CAZymes will expand and diversify the portfolio of plant cell wall degrading enzymes that NZYTech offers at a commercial level. To support these goals, the development of novel strategies for high-throughput cloning, expression and purification of cellulosomal proteins in *E. coli* in soluble form is an underlying and crucial aim of this work. The novel protocols could be used to seek the expression of different heterologous proteins with industrial relevance and are in agreement with the development of innovative molecular biology products that NZYTech pursues. Specifically, the main goals of this project may be summarized as follows:

- To develop a high-throughput platform for cloning, expression and protein purification for the rapid discovery of novel cellulosomal enzymes, bearing in mind that modules of unknown function appended to dockerins may comprehend important biocatalysts for carbohydrate deconstruction (Chapter 3).
- To develop novel strategies to enhance protein solubility in *E. coli*, including an approach based on the pivotal protein:protein interactions that sustain cellulosome complex and by creating fusion quimeras with heterologous cellulosome proteins (Chapter 4).
- To functionally characterize a major cellulosomal enzyme from *Clostridium thermocellum* when this bacterium is cultured on cellulose, the cthe_0821, here named as Man5A. Also to study the binding preferences of the associated CBM32 (Chapter 5).
- To functionally characterize three cellulosomal pectinolytic enzymes, PL1A, PL1B and PL9 from *Clostridium thermocellum,* and elucidate the pectinolytic activity displayed by this bacterium (Chapter 6).
- To evaluate the capacity of two β-glucanases from *Clostridium thermocellum* (β-1,3-1,4-glucanase 16A and β-1,4-glucanase 8A) to improve the nutritive value of barley-based diets for broilers, and thus elucidate which of these two distinct enzyme activities performs better to improve the nutritive value of those diets for monogastric animals (Chapter 7).
- To construct GH16 β-glucanase mini-cellulosomes applied to improve the nutritive value of barley-based diets for broilers (Chapter 7).

3. HIGH-THROUGHPUT CLONING AND EXPRESSION OF CELLULOSOMAL ENZYMES

3.1. A high-throughput platform for gene cloning, expression and purification of recombinant proteins in *Escherichia coli*: discovery of novel enzymes in cellulosomes

Vânia O. Fernandes[1,2], Ana Sofia Luís[3], Harry J. Gilbert[3], Joana L. A. Brás[2], Catarina I. P. D. Guerreiro[2], Luís M. A. Ferreira[1,2], Carlos M. G. A. Fontes[1,2]

[1] CIISA-Faculdade de Medicina Veterinária, Universidade de Lisboa, Av. da Universidade Técnica, 1300-477 Lisboa, Portugal; [2] NZYTech Lda, Estrada do Paço do Lumiar, Campus do Lumiar, Edifício E-R/C, 1649-038 Lisboa, Portugal; [3] Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, United Kingdom

Adapted from a manuscript in preparation

Abstract

Increased concerns over generating green chemicals have triggered interest in the efficient conversion of biomass. Production of biofuels from lignocellulosic materials requires the extensive hydrolysis of the recalcitrant plant cell wall polysaccharides. This process involves the concerted action of a vast repertoire of carbohydrate-active enzymes (CAZymes), which act individually or associated in large molecular mass multi-enzyme complexes termed cellulosomes. The proportion of these enzymes which are functionally annotated is, however, scarce. In addition, extensive analysis of cellulosomal polypeptides revealed a significant number of proteins of unknown function that are assembled in multi-enzyme complexes. Since cellulosomes play a key role in carbohydrate deconstruction they comprise an extremely interesting source for the discovery of novel CAZymes. Here we describe the development of a high-throughput method for gene cloning, expression and protein purification that was applied to produce recombinant cellulosomal proteins of unknown function. The platform was designed to clone multiple inserts into a novel prokaryotic expression vector through an innovative ligation-independent cloning method and to express the microbial polypeptides at small-scale but at high levels in *Escherichia coli*. Purified recombinant proteins were screened for α-L-arabinofuranosidase activity using a colorimetric assay, which allowed the discovery of a novel CAZyme family in the cellulosome of the ruminal bacterium *Ruminococcus flavefaciens*. Distant similarities with members of Glycoside Hydrolases families 43 and 62 (GH43 and GH62) suggest that this novel family belongs to Clan GH-F and should display the 5-bladed β-propeller fold.

3.1.1. Introduction

The decrease in fossil fuel reserves and the growing environmental consciousness increased the economic and social interest in the production of biofuels from lignocellulosic biomass. In this context, research towards the efficient conversion of biomass is increasing. However, plant cell walls are highly recalcitrant to biological degradation, which difficult the development of a biorefinery based on lignocellulosic biomass at an economical scale (Li *et al*., 2009). Therefore, a lot of emphasis is being put on the discovery and development of more efficient carbohydrate-active enzymes (CAZymes) and in innovative mechanisms to potentiate their activity and stability (Lynd *et al*., 2002; Lynd *et al*., 2005; Demain, Newcomb, & Wu, 2005).

The major constituent of plant cell walls is cellulose, mainly in the form of crystalline micro fibrils (Somerville, 2006). Other abundant structural components of cell walls are hemicelluloses (such as xyloglucans, xylans, mannans or glucomannans), pectins, lignin and proteins with diverse roles (Cosgrove, 1997; Scheller & Ulvskov, 2010). The physical association between polysaccharides and between polysaccharides and lignin restrict their accessibility to the degrading enzymes (Gilbert, 2010). In addition, only a limited number of microorganisms have acquired the capacity to deconstruct structural carbohydrates such as cellulose and hemicellulose (Fontes & Gilbert, 2010). Reflecting the intricacy of plant cell walls, microorganisms that degrade these structures produce an extensive arsenal of carbohydrate-active enzymes, such as glycoside hydrolases (GH), polysaccharide lyases, carbohydrate esterases and glycosyl transferases. These enzymes are frequently modular, containing a catalytic module connected through flexible linker sequence to non-catalytic domains, such as Carbohydrate Binding Modules (CBMs) involved in protein-carbohydrate interactions (Cantarel *et al*., 2009). CAZymes synthesized by some anaerobic microorganisms, particularly from the genus *Clostridia* and *Ruminococcus*, are assembled into a large molecular mass multienzyme complex, termed the cellulosome (Gilbert, 2007; Fontes & Gilbert, 2010). Cellulosomes promote enzyme synergism, afforded by spatial proximity, and stability (Bayer *et al*., 2004). It is well established that cellulosomes are more efficient than the free enzyme systems produced by aerobic microorganisms in the degradation of recalcitrant substrates (Fontes & Gilbert, 2010). Integration of cellulosomal components occurs via highly ordered protein:protein interactions established between non-catalytic dockerin domains located on the cellulosomal enzymes with cohesin domains located on a molecular scaffold. Cohesin-dockerin interactions also mediate cellulosome attachment to the bacterial cell surface (Bayer *et al*., 2004; Carvalho *et al*., 2005; Fontes & Gilbert, 2010; Brás *et al*., 2012).

Several genomes of cellulosome producing bacteria, such as *Clostridium thermocellum*, *Clostridium acetobutylicum*, *Ruminococcus flavefaciens* and *Clostridium cellulolyticum* have recently been sequenced, providing the capacity to perform a detailed analysis of all

encoded cellulosomal molecular components (Fontes & Gilbert, 2010; Munir *et al*., 2014). Cellulosomes are exciting resources for mining novel CAZymes that could present novel biochemical properties adequate for several biotechnological applications. In addition, cellulosomal analysis revealed several modules of unknown function that could constitute novel enzymes presenting critical substrate specificities. Moreover, although there is an exponential increase in the annotation of novel CAZymes in recently sequenced genomes, there is a significant discrepancy between the number of enzymes organized in families in the constantly updated CAZy database (www.cazy.org) and the availability of biochemical/structural data for these proteins (Cantarel *et al*., 2009).

Reflecting the complexity of plant cell walls, the number of CAZymes that remain to be discovered may be remarkably large. According to Davies, Gloster, & Henrissat (2005), approximately 1-3% of the proteins encoded by the genomes of most organisms correspond to CAZymes, being the GH-coding genes the most abundant. Among GHs, α-L-arabinofuranosidases (EC 3.2.1.55) have recently received special attention due to their potential for biotechnology. Alone or in combination with other CAZymes, these hemicellulases can be used in the pretreatment of lignocellulose materials for biofuel production (Das *et al*., 2012), as well as in the biobleaching of paper pulp (Bothast, 1998; Saha, 2003; Fridjonsson *et al*., 2004). α-L-arabinofuranosidases hydrolyze the terminal α-1,2-, α-1,3- and/or α-1,5 α-L-arabinofuranosyl residues from mixed linkage polysaccharides such as arabinoxylans and arabinans, thus facilitating the access of synergistic enzymes to the polysaccharide backbones (Saha, 2000; Numan & Bhosle, 2006). In the CAZy database, α-L-arabinofuranosidases are mainly grouped in families GH2, GH3, GH10, GH43, GH51, GH54 and GH62.

Characterization of enzymes revealed by the enormous data generated by sequencing projects requires efficient high-throughput methods for gene isolation and cloning, protein expression and purification, coupled with efficient enzyme screening assays that may reveal novel enzyme specificities that may be subsequently studied at a structural level and potentiate the development of novel biotechnological applications. Here we describe the development of a novel high-throughput platform to efficiently clone and express a large number of recombinant cellulosomal proteins of unknown function that can be rapidly used in subsequent enzyme discovery projects. The platform was adapted to operate in a Tecan robot (Switzerland). The genes encoding cellulosomal modules of unknown function from *R. flavefaciens* strain FD-1 and *C. thermocellum* strain ATCC 27405 were cloned and their encoded enzymes expressed at high levels in *Escherichia coli*. The recombinant proteins of unknown function were purified in a high-throughput platform and screened for α-L-arabinofuranosidase activity. This work allowed the discovery of a novel enzyme which displays activity against 4-nitrophenyl-*α*-arabinofuranoside (*p*NPA*f*). This enzyme is the

founder member of a novel GH family which is distantly related with GH43 and GH62 families.

3.1.2. Materials and Methods

3.1.2.1. Construction of a novel prokaryotic ligation-independent cloning vector for protein expression

Confidential

3.1.2.2. Automated PCR primer design

We have developed a dedicated software to the large scale design of primers based in multiple DNA sequences. The algorithm used for primer design allows optimizing melting temperatures, primer lengths, GC content and introducing engineered 5´sequences. The software uses a Microsoft Excel interface containing the variable number of genes that are required to be amplified and selects the appropriate pair of forward and reverse primers for each gene. Vector complementary single-strand overhangs for the ligation-independent cloning reaction are automatically introduced in the primers. Primers pairs were designed using this automated method to amplify 239 genes encoding dockerin-bearing modules with unknown function from *R. flavefaciens* FD-1 and *C. thermocellum* ATCC 27405. The primers were designed to have a Tm of ~60-70°C and included additional 16-bp pHTP homologous sequences at 5'- terminus of both forward and reverse primers in order to allow the ligation-independent cloning.

3.1.2.3. PCR amplification

R. flavefaciens FD-1 and *C. thermocellum* ATCC 27405 genomic DNA (obtained from DSMZ, Germany) were used as templates to amplify 239 genes (223 from *R. flavefaciens* and 16

from *C. thermocellum*) encoding cellulosomal modules of unknown function. Primer dilutions and PCR amplifications were carried out in a 96-well plate format. NZYProof DNA polymerase (NZYTech genes & enzymes, Portugal) was used for PCR using optimized conditions to minimize primer-dimer formation and nonspecific amplifications. PCR reactions were performed in 50 µL total volume and each primer was added to a final concentration of 0.14 µM. The cycling parameters were as follows: 1 cycle of 95°C for 10 min; 30 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 3 min; followed by 1 cycle of 72°C for 10 min. After PCR, reaction products were visualized by agarose gel electrophoresis and purified using the NZYDNA Clean-up 96 well plate kit (NZYTech genes & enzymes, Portugal) in a Tecan robot (Switzerland).

3.1.2.4. A novel method for ligation-independent cloning

High-throughput cloning is facilitated by the use of ligation-independent cloning (LIC) methods, which are not limited by the number of targets and occurrence of restriction sites in selected genes. In the LIC method described here, no previous preparation of the vector DNA fragment is required and selection of correct clones is achieved by inactivation of a lethal gene present in the vector. Inserts prepared by PCR are cloned directly into the expression vector, without the need to go through an intermediate cloning vector. Purified PCR fragments were mixed with 240 ng of circular pHTP vector in a 96-well PCR plate format. Different vector to insert molar ratios (1:2–1:20) were mixed with 1 µL of the cloning enzyme mix and 2 µL of 10x reaction buffer and then the final reaction volume was adjusted to 20 µL with water. The cloning reactions proceeded for 1 h at 37°C on a thermo cycler to generate DNA nicks in both vector and insert sequences. The mixtures were then incubated at 80°C for 10 min followed by 10 min at 30°C to stimulate the denaturing and the single-strand annealing of both vector and insert sequences. The cloning reaction mixtures were used to transform *E. coli* DH5α competent cells (NZYTech genes & enzymes, Portugal). From each transformation reaction, colonies were picked and the presence of the insert was confirmed through colony PCR using two vector specific primers (the universal T7 promoter and pET24a primers). NZYLong 2× Green Master Mix (NZYTech genes & enzymes, Portugal) was used for the PCR amplification. Cultures containing recombinant plasmids were grown in 5 mL of LB medium supplemented with 50 µg/mL of kanamycin in 24-deep-well plates (5 mL) sealed with gas-permeable adhesive seals. After growth at 37°C for 16 h, cells were harvested at 1000 ×*g* for 10 min. Plasmids were purified from the bacterial pellet using NZYMiniprep 96 well plate kit (NZYTech genes & enzymes, Portugal) in a Tecan robot (Switzerland).

3.1.2.5. Comparison of the efficacy of different cloning methods

The efficiency of the cloning method described above was compared with other alternative commercially available solutions for the cloning of DNA fragments of 0.4, 1.2 and 2.6 kb. Four commercial systems were selected for this experiment: In-Fusion® HD Cloning Kit (Clontech); ClonEZ® PCR Cloning Kit (GenScript); GeneArt® Seamless Cloning and Assembly Kit (Invitrogen™ by Life Technologies); and Choo-Choo Cloning™ Kit (MCLab). For all these four methods, pET28a was PCR linearized and used for the cloning reactions. Vector to insert molar ratios of 1:4 (for fragments of 0.4 and 1.2 kb) and 1:2 (for 2.6 kb fragment) were mixed with the appropriate enzyme mixtures, following the manufacturers' protocols. Percentage of recombinant colonies was calculated for each condition in test, by analyzing 32 colonies from each transformation. The competency of *E. coli* DH5α cells was 1.54×10^8 cfu/μg for circular pET28a. DH5α cells are killed due to the presence of the lethal genesalthough reduced background can result from its disruption by nucleases or spontaneous mutagenesis (Bernard, 1996).

3.1.2.6. Testing the efficacy of different auto-induction media

We have developed two new *E. coli* auto-induction media systems: an LB-based auto-induction medium (NZY Auto-Induction LB medium) and the NZY Auto-Induction Kit, both now available from NZYTech, genes & enzymes, Portugal. The NZY Auto-Induction LB medium is a rich culture medium formulated with tryptone, yeast extract, nitrogen, phosphate, magnesium, and trace metals in adequate proportions. The method is based on the presence of different carbon sources that are metabolized differentially to promote culture growth to high cell densities and subsequently induce protein expression from *lac*-based promoters. This medium is prepared by simply adding 50 g/L of powder and 10 mL/L of glycerol to sterile distilled water. The NZY Auto-Induction Kit contains two concentrated sterile solutions (induction and buffering) that can be added to traditional complex media, such as Luria-Bertani (LB) broth or Terrific Broth (TB). Like the NZY Auto-Induction LB medium, this method is also based on medium components that promote culture growth to high cell densities and subsequent induction of protein expression from *lac*-based promoters. The kit is used by adding 25 mL/L of induction solution and 50 mL/L of the buffering solution to sterile glucose-free complex media.

The capacity of these and other auto-induction media for expressing bacterial recombinant proteins at high levels in *E. coli* was compared. Initially, a recombinant protein from *R. flavefaciens* was tested for expression in either NZY Auto-Induction LB medium or LB medium supplemented with solutions aiming at increasing cell densities provided by the NZY Auto-Induction Kit, against the standard auto-induction medium ZYM-5052 (Studier, 2005). A pre-inoculum of *E. coli* BL21(DE3) cells containing the recombinant plasmid grown from a

fresh isolated colony was prepared and used to inoculate 50 mL of each auto-induction media in 500 mL erlenmeyer flasks The recombinant cultures were grown at 37°C with vigorous aeration and collected at different times during ~ 24 h of incubation. The OD_{600nm} of the cultures at different time points was monitored, and protein expression was accessed by analyzing total protein through denaturing gel electrophoretic analysis (SDS-PAGE). Subsequently, different auto-induction media were tested for the expression of 24 different *R. flavefaciens* recombinant proteins in both *E. coli* BL21(DE3) and BL21(DE3)pLysS (Novagen®). The media under investigation were: NZY Auto-Induction LB medium (NZYTech genes & enzymes, Portugal) against the equivalent Overnight Express™ Instant TB Medium (Novagen®), and NZY Auto-Induction Kit (NZYTech genes & enzymes, Portugal) which was probed against the Overnight Express™ Autoinduction System 1 (Novagen®). Pre-inoculums from fresh colonies were prepared and used to inoculate 5 mL of each auto-induction medium supplemented with kanamycin (50 µg/mL). Recombinant strains were grown in 24-deep-well plates at 37°C at 400 rpm in an orbital incubator shaker to stationary phase (OD_{600nm} > 10). The bacteria were collected by centrifugation of the deep-well plates at 2.500 ×*g* for 10 min. The His_6-tagged recombinant proteins were purified from cell-free extracts by immobilized metal-ion affinity chromatography (IMAC), as described below, and analyzed by SDS-PAGE. Protein concentration was determined using a NanoVue spectrophotometer (GE Healthcare).

3.1.2.7. Small-scale protein expression

Based on the previous comparison studies of the growing conditions, recombinant plasmids encoding *R. flavefaciens* or *C. thermocellum* cellulosomal modules of unknown function were used to transform *E. coli* BL21(DE3)pLysS competent cells. The 239 recombinant strains generated were subsequently grown in 5 mL of NZY Auto-Induction LB medium (NZYTech genes & enzymes, Portugal) supplemented with kanamycin (50 µg/mL). Growth was carried out in 24-deep-well plates at 37°C for four hours in a microplate shaker. The temperature was then dropped to 30°C and cells were left to grow for 16-20 hours (OD_{600nm} > 10). Cells were harvested by centrifugation of the deep-well plates at 2.500 ×*g* for 10 min.

3.1.2.8. Using different bacterial lysis methods for extracting recombinant proteins

Initial studies were performed to compare the capacity of different chemical lysis methods that can easily be integrated in HTP protocols to disrupt *E. coli* cells. Thus, *E. coli* BL21(DE3) or BL21(DE3)pLysS cells expressing the *C. thermocellum* cellulase 8A (*Ct*Cel8A) were used to test three different lysis buffers: NZY Bacterial Cell Lysis Buffer (a previously developed detergent-based buffer available from NZYTech genes & enzymes, Portugal), supplemented with 0.1 mg/mL lysosyme and 0.004 mg/mL DNase I, B-PER® Bacterial Protein Extraction

Reagent (Thermo Scientific) and BugBuster® Protein Extraction Reagent (Novagen®). *E. coli* cells expressing *Ct*Cel8A were cultured in 100 mL of LB-Broth media supplemented with kanamycin (50 µg/mL) at 37°C to mid-exponential phase (OD_{600nm} = 0.6) and recombinant protein overproduction was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) at 1 mM final concentration. Cells were further incubated for 16 h at 19°C. The following day, 5 mL of each culture (BL21(DE3) and BL21(DE3)pLysS) were distributed into 9 wells of a 24-deep-well plate and bacteria collected by centrifugation at 2.500 ×*g* for 10 min. Cells were ressuspended in 1 mL of buffer in test (in triplicates) and lysis proceeded in a microplate shaker at room temperature for ~15 min, when an homogeneous cell-lysed extract appeared. The His_6-tagged recombinant *Ct*Cel8A was purified from cell-free extracts by immobilized metal-ion affinity chromatography (IMAC) as described below and separated by SDS-PAGE. Protein concentration was determined by the Bradford Assay.

The efficiency of the NZY Bacterial Cell Lysis Buffer was compared to the classical mechanical ultra-sonication method in the disruption of 22 recombinant *E. coli* BL21(DE3) strains expressing different *R. flavefacien*s proteins. Cells were cultured in 40 mL of NZY Auto-Induction LB medium (NZYTech genes & enzymes, Portugal) supplemented with kanamycin (50 µg/mL) in 500 mL-flasks. After growth at 37°C during 16 h, 2 x 5 mL of each cell culture were distributed into two 24-deep-well plates, respectively. Bacteria were collected by centrifugation and pellets were either ressuspended in 1 mL of 50 mM NaHEPES, 1 M NaCl, 10 mM Imidazole, 5 mM $CaCl_2$, pH 7.5 followed by sonication (1 min at 70% for each well) (1 plate) or lysed in 1 mL of NZY Bacterial Cell Lysis Buffer, supplemented with 0.1 mg/mL lysosyme and 0.004 mg/mL DNase I (~15 min at room temperature) (1 plate). Recombinant proteins extracted using these two strategies were purified through IMAC as described below and separated by SDS-PAGE. Protein concentration was determined using a NanoVue spectrophotometer (GE Healthcare).

3.1.2.9. Automated small-scale protein purification

Based on the previous comparison studies of the cell lysis conditions, the cell-pellets from the 239 recombinant cultures encoding unknown cellulosomal proteins were disrupted in 1 mL of NZY Bacterial Cell Lysis Buffer (NZYTech genes & enzymes, Portugal), supplemented with 0.1 mg/mL lysosyme and 0.004 mg/mL DNase I. All recombinant proteins contained an engineered N-terminal His_6-tagged. Recombinant proteins were purified by immobilized metal-ion affinity chromatography (IMAC) using a Tecan robot (Switzerland). After cell lysis, the crude cell lysates were incubated at room temperature with 200 µL of Ni^{2+} Sepharose beads to capture the recombinantly expressed proteins. The deep-well plates were shaken for 10 min in the robot plate shaker. Sepharose beads were prepared by combining 4x Chelating Sepharose Fast Flow (GE Healthcare) and 0.04 M of nickel in 50 mM NaHEPES, 1

M NaCl, 10 mM Imidazole, 5 mM $CaCl_2$, pH 7.5. To separate the beads from the cell extract, the whole samples were transferred into a 96-well receiver filter plate (Macherey-Nagel). Filtration was achieved by applying vacuum. The beads were washed twice with 1 mL of 50 mM NaHEPES, 1 M NaCl, 10 mM Imidazole, 5 mM $CaCl_2$, pH 7.5, followed by 2 x 1 mL washes with 50 mM NaHepes, 1 M NaCl, 35 mM Imidazole, 5 mM $CaCl_2$, pH 7.5. The target proteins were eluted from the beads with 0.15 mL of 50 mM NaHEPES, 1 M NaCl, 300 mM Imidazole, 5 mM $CaCl_2$, pH 7.5 in 96-deep-well plates.

3.1.2.10. Analysis of levels of gene expression and enzyme discovery screen

Levels of protein expression were estimated by SDS-PAGE analysis with 14% gels according to Laemmli (1970). Purified protein samples were screened for α-L-arabinofuranosidase activity in a colorimetric assay. Reactions were carried out in 96-well plates by mixing 25 µL of sodium phosphate buffer, pH 7.0 containing 1 mM 4-nitrophenyl-*α*-arabinofuranoside (*p*NPA*f*) (Sigma) with 5 µL of each protein sample. A negative reaction control was performed using 5 µL of 50 mM NaHEPES, 1 M NaCl, 300 mM Imidazole, 5 mM $CaCl_2$, pH 7.5 buffer. The positive control consisted in the addition of a GH51 arabinofuranosidase (CR0024, NZYTech genes & enzymes, Portugal). Plates were incubated at 37°C in a microplate shaker incubator. Enzyme activity was qualitatively accessed by the formation of an intense yellow soluble product.

3.1.2.11. Thin layer chromatography (TLC)

Reaction products of the enzyme identified in this study were analysed by TLC. Briefly, the enzyme reaction was performed in 200 mM HEPES buffer, pH 7.0, containing 150 mM NaCl, 0.2% wheat arabinogalactan, 0.1 mg/mL BSA, 2 mM $CaCl_2$, and 1 µM of enzyme in a final volume of 200 µL. Reactions were incubated for 16 h at 37°C. Each reaction was spotted 2 times (3 µL) in a thin layer chromatography plate and run two times in 1-butanol/acetic acid/water (2:1:1, v/v). Visualization of the sugars was done by incubation with orcinol sulphuric acid reagent (sulphuric acid/ethanol/water 3:70:20 v/v, orcinol 1%) followed by incubation at 120°C until sugars appeared. Galactose and arabinose were used as monosaccharide standards.

3.1.3. Results and Discussion

3.1.3.1. Modules of unknown function in *R. flavefaciens* and *C. thermocellum* cellulosomes

Genome sequencing revealed more than 70 genes encoding dockerin-containing proteins, which constitute cellulosomal proteins, in the genome of *C. thermocellum* ATCC 27405, while over 200 of these genes are present in *R. flavefaciens* FD-1. The majority of the cellulosomal proteins identified in these two bacteria are glycoside hydrolases, carbohydrate esterases or

polysaccharide lyases. In addition, putative proteases and protease inhibitors have also been identified in the cellulosome of *C. thermocellum* (Demain *et al*., 2005; Raman *et al*., 2009; Bayer *et al*., 2008). These large arrays of multi-modular enzymes work in a concerted and synergistic manner to deconstruct structural polysaccharides. However, a large number of cellulosomal modules, mainly from the recent genome-sequenced *R. flavefaciens*, are still of unknown function, i.e. do not bear homology with any of the families catalogued in the CAZy database. Furthermore, even for CAZymes classified into sequence-based families, substrate specificity cannot be directly predicted based on *in silico* analysis as within the same CAZyme family there is usually no conservation in substrate preference (Mizutani *et al*., 2012). Here, a high-throughput platform was designed to recombinantly produce unknown modules identified in bacterial cellulosomes to be used in subsequent functional screenings. Thus, 239 genes, 223 from *R. flavefaciens* and 16 from *C. thermocellum*, encoding cellulosomal modules of unknown function, were isolated, cloned, expressed and purified in a high-throughput platform. The proteins of unknown function were then screened for α-L-arabinofuranosidase activity.

3.1.3.2. An automatable high-throughput ligation-independent cloning method

To clone 239 cellulosomal genes in a high-throughput manner we developed a ligation-independent-based method that allows cloning the nucleic acid inserts (PCR products) directly into a novel prokaryotic expression vector by a base complementation strategy. The efficacy of the novel system (here reported as Easy cloning method) was tested by cloning DNA fragments of different sizes (164-bp, 335-bp and 2.6-kb) and using different vector to insert molar ratios (1:2, 1:4, 1:6, 1:8, 1:10, and 1:20). The data, presented in Table 3.1, revealed that the largest number of colonies was obtained using 1:4-1:6 ratios for the smallest insert and 1:2-1:4 ratios for the 335-bp fragment. The number of colonies generated when cloning the largest DNA fragment was similar for all molar vector:insert ratios tested (Table 3.1). Regarding the recombinant clones, the percentage obtained for the smallest-sized insert was high (100%) and similar to all the ratios tested, except for 1:20 (which resulted in the lowest number of colonies). The percentage of recombinant clones varied slightly for the 335-bp fragment, with the lowest number obtained for ratio 1:4 (90%). A lower percentage of recombinant clones was observed for the largest-sized DNA fragment, which also decreased when high molar insert ratios were used. Nonetheless, in general, the data correspond to high cloning efficiencies (considering vector:insert ratios from 1:2-1:10). Molar ratios of vector to insert could, however, be adjusted according to the insert length.

Table 3.1| Recombinant colony counts using different molar ratios of vector to insert. A total of 20 colonies from each transformation was analyzed.

Insert	Molar ratio of vector to insert	Nº. colonies	Recombinant colonies (%)
164 bp	1:2	64	100
	1:4	120	100
	1:6	94	100
	1:8	56	100
	1:10	46	100
	1:20	2	50
335 bp	1:2	120	95
	1:4	95	90
	1:6	76	100
	1:8	68	95
	1:10	44	100
	1:20	10	100
2.6 kb	1:2	115	85
	1:4	130	85
	1:6	140	85
	1:8	140	70
	1:10	115	75
	1:20	100	75

The efficacy of different ligation-independent methods to clone DNA fragments of 0.4, 1.2 and 2.6 kb was also investigated. Cloning reactions were assayed following the manufacturers protocols using the vector pET28a (Novagen®) for In-Fusion® HD Cloning Kit, ClonEZ® PCR Cloning Kit, GeneArt® Seamless Cloning and Assembly Kit, and Choo-Choo Cloning™ Kit. The efficiency of the different protocols was tested by randomly analyzing 32 colonies generated by each protocol through colony PCR. The data, summarized in Table 3.2, revealed that the Easy cloning method displays a considerable high cloning efficiency, with percentages of recombinant colonies of 100% for the 0.4-kb insert, 97% for the 1.2-kb insert, and 81% for the 2.6-kb insert, which are in the same order of magnitude to the other methods tested. The challenge occurred when trying to clone the largest DNA fragment; relatively low percentage of recombinant clones was obtained for all the methods in study when comparing with small inserts. As stated above, it is possible that the vector to insert molar ratio requires an adjustment for larger inserts to maximize cloning efficiencies.

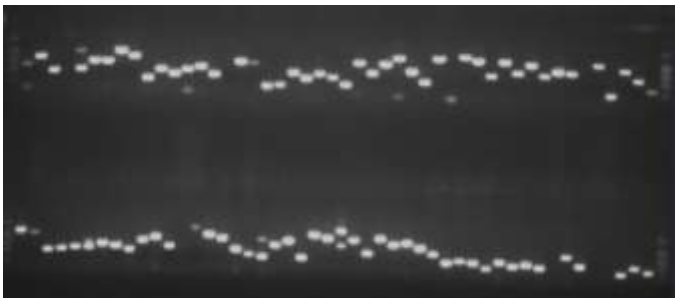
The Easy cloning method described above was used to clone 239 cellulosomal genes in the pHTP plasmid. DNA amplifications and cloning reactions were performed in three batches of 47, 96 and 96 genes, respectively, performed in three 96-well plates. In general, PCR amplifications were very efficient (Figure 3.1) although for 42 genes optimization of PCR

conditions were needed to improve yield (data not shown). After cloning, all 239 genes were sequenced to confirm that no mutations accumulated during PCR amplification.

Table 3.2| Cloning efficiency for inserts of different lengths using different ligation-independent cloning methods.

Insert	Cloning method	Nº. colonies	Recombinant colonies (%)
0.4 kb	Easy cloning method	820	100
	In-Fusion® HD Cloning kit	280	100
	CloneEZ® PCR Cloning Kit	9	100
	GeneArt® Seamless PLUS Cloning and Assembly Kit	1200	100
	Choo-Choo Cloning™ Kit	1150	100
1.2 kb	Easy cloning method	850	97
	In-Fusion® HD Cloning kit	240	100
	CloneEZ® PCR Cloning Kit	8	75
	GeneArt® Seamless PLUS Cloning and Assembly Kit	1300	81
	Choo-Choo Cloning™ Kit	480	88
2.6 kb	Easy cloning method	375	81
	In-Fusion® HD Cloning kit	440	97
	CloneEZ® PCR Cloning Kit	11	55
	GeneArt® Seamless PLUS Cloning and Assembly Kit	2000	94
	Choo-Choo Cloning™ Kit	410	78

Figure 3.1| Representative PCR amplification for 96 of the targets in study.



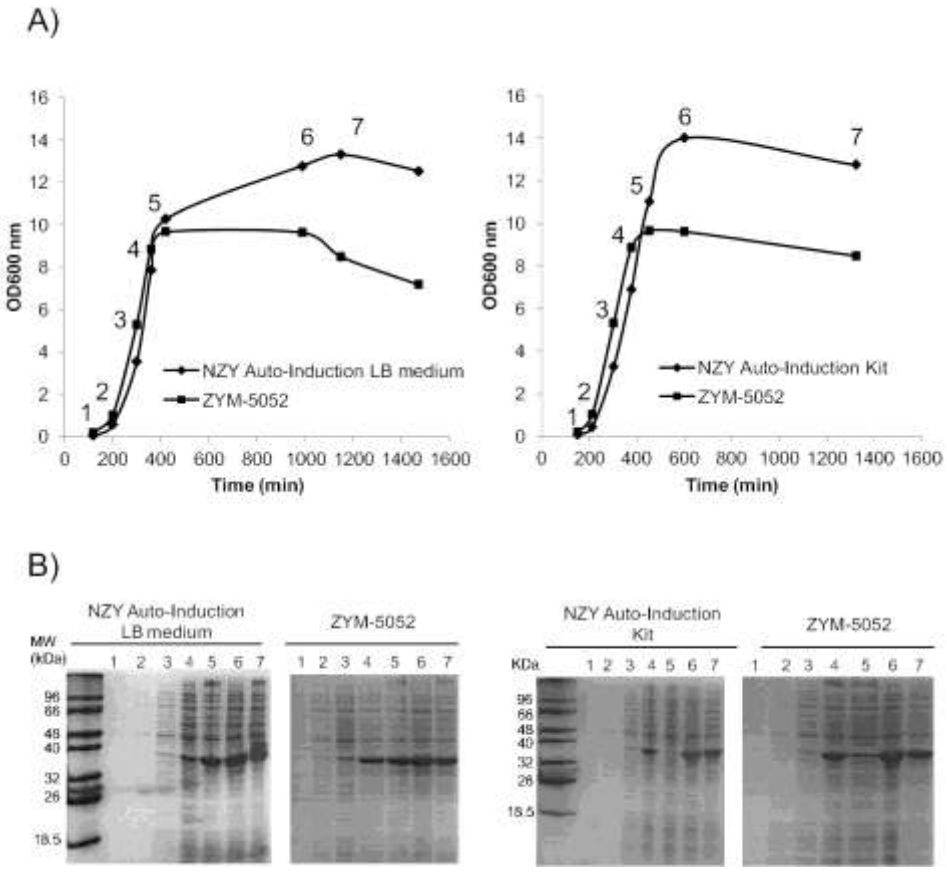
PCR was repeated when absence of the intended band was observed, as well as in the presence of nonspecific bands and/or primer-dimers formation.

3.1.3.3. Levels of expression of cellulosomal modules in *E. coli*

Two auto-induction systems were developed to support growth of *E. coli* to high cell densities and to obtain high-levels of recombinant protein expression from *lac*-based promoters: NZY Auto-Induction LB Medium and NZY Auto-Induction Kit (both now available from NZYTech, genes & enzymes, Portugal). The capacity of these auto-induction media to promote the expression of a *R. flavefaciens* protein was compared against the standard ZYM-5052 auto-induction medium (Studier, 2005). The data, presented in Figure 3.2, revealed that for the

NZY Auto-Induction LB medium, induction started at OD_{600nm} ~ 8 (after around 6 h of incubation in 50 mL cultured flasks at 37 ºC) and maximal levels of recombinant proteins were obtained at OD_{600nm} values of 13-14, which corresponded to 16-20 h of incubation. Levels of protein expression obtained using NZY Auto-Induction LB medium were equivalent or slightly higher when compared with the ZYM-5052 medium. For cells grown in LB medium supplemented with the NZY Auto-Induction Kit, the time and OD_{600nm} from which induction started was similar to that observed for the NZY Auto-Induction LB medium. Levels of protein expression obtained using the kit were similar when compared with ZYM-5052. Both NZY Auto-Induction LB medium and NZY Auto-Induction Kit allowed cells to reach higher OD_{600nm} values when compared with the standard auto-induction medium. After 22 h of incubation, a slightly decrease of protein expression was observed. Indeed, low or even no protein production has been reported at high cell densities mainly as a result of plasmid loss (Baneyx, 1999b), pH drop or restricted oxygen availability (Jana & Deb, 2005).

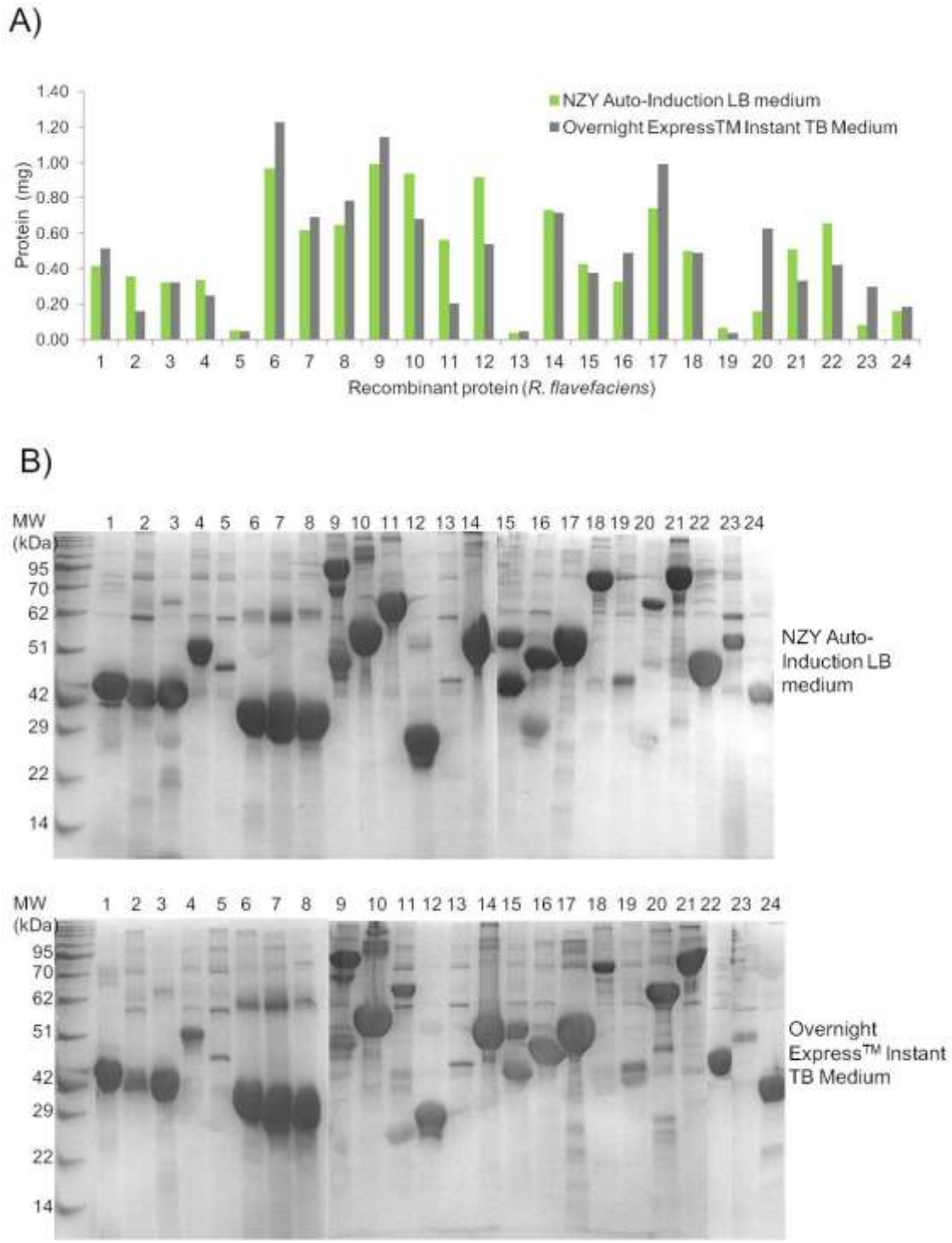
Figure 3.2| Levels of expression of a recombinant protein from *Ruminococcus flavefaciens* obtained using the NZY Auto-Induction LB medium and the LB medium supplemented with the solutions provided by the NZY Auto-Induction Kit.



Both auto-induction systems available from NZYTech, genes & enzymes were compared with a standard auto-induction medium – ZYM-5052. Samples were taken at different time points (1-7) during growth to construct growth curves (**A**) and the corresponding total cell extracts were separated through SDS-PAGE to analyze protein expression profiles (**B**). *E. coli* BL21(DE3) strain was used as host.

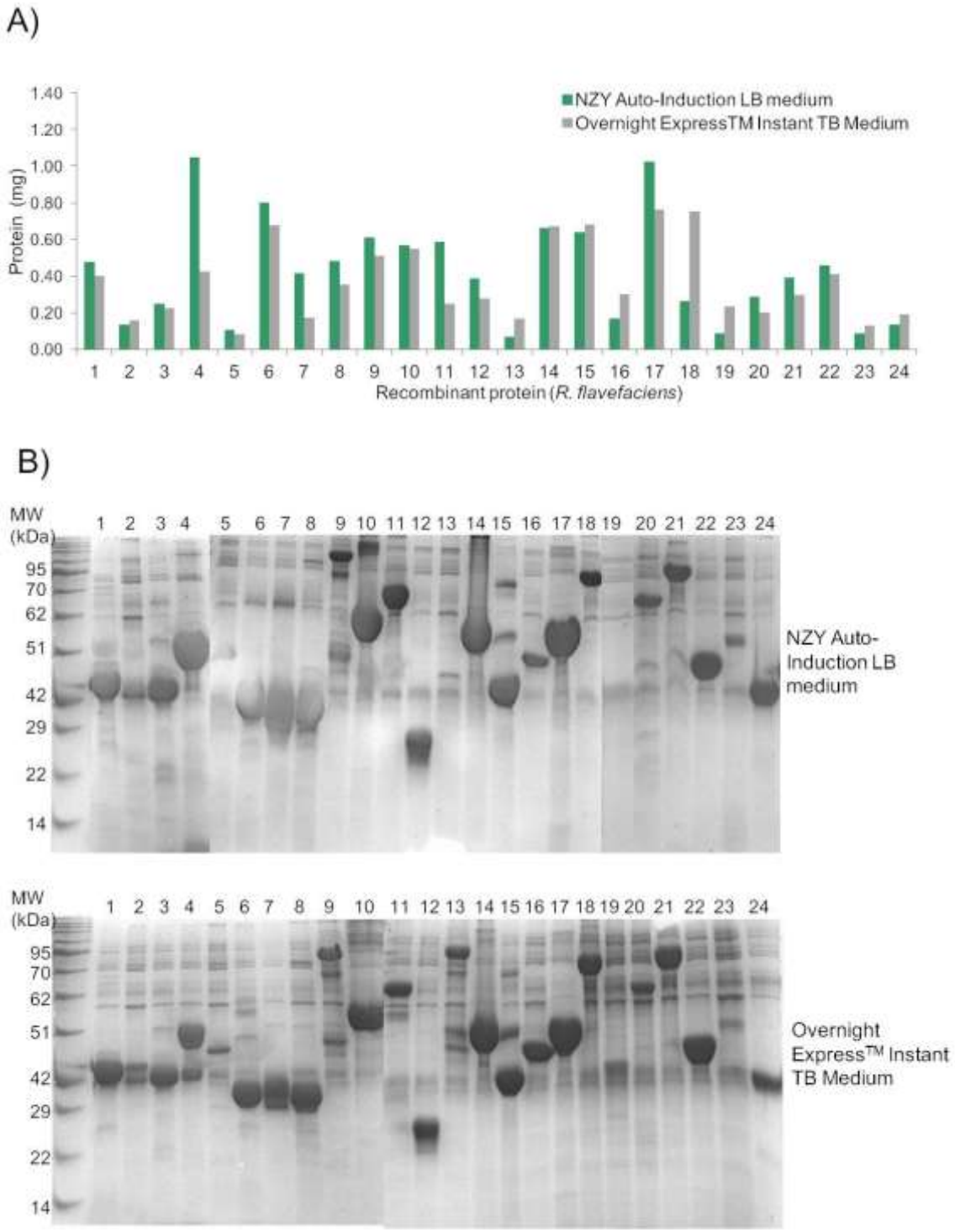
The efficacy of NZY Auto-Induction LB medium and Overnight Express™ Instant TB Medium for the production of 24 different recombinant proteins from *R. flavefacien*s in two expression *E. coli* strains, BL21(DE3) and BL21(DE3)pLysS, was also compared. Recombinant proteins were purified through IMAC and separated by SDS-PAGE. Data, presented in Figures 3.3 and 3.4, revealed that NZY Auto-Induction LB medium led from similar to higher levels of pure protein when compared with Overnight Express™ Instant TB Medium. For BL21(DE3) cells (see Figure 3.3), a detailed data analysis revealed that for 12 proteins tested, NZY Auto-Induction LB medium performed better than the Overnight Express™ Instant TB Medium. Only one protein, protein number 20, was expressed at higher levels by the Overnight Express™ Instant TB Medium. For the remaining 10 proteins, the differences observed between media were not significant. Using BL21(DE3)pLysS cells (see Figure 3.4), the levels of expressed proteins were, in general, slightly lower than those obtained in BL21(DE3), with some exceptions. Since pHTP uses a T7/lac promoter, the combination of the T7 lysozyme expressed by the pLysS plasmid (Moffatt & Studier, 1987) and the *lac* repressor (encoded by *lac*I) provides a tighter control of gene expression (Dubendorf & Studier, 1991), which may lead to significant reduced protein expression upon induction (Studier, 1991; Pan & Malcolm, 2000). Nevertheless, NZY Auto-Induction LB medium performed better for 15 proteins when compared with the Overnight Express™ Instant TB Medium. A similar experiment was performed comparing the NZY Auto-Induction Kit with the Overnight Express™ Autoinduction System 1 (Novagen®). The results are presented in Figures 3.5 and 3.6. Although the level of protein purity may have affected the A_{280nm} measurements of the purified samples through the NanoVue, looking for the SDS-PAGE gels (Figures 3.5B and 3.6B), the data suggested that levels of protein expression were similar, in both BL21(DE3) and BL21(DE3)pLysS strains.

Figure 3.3| Levels of purified protein obtained from 24 different recombinant *E. coli* BL21(DE3) strains grown in the NZY Auto-Induction LB medium or in the Overnight Express™ Instant TB Medium.



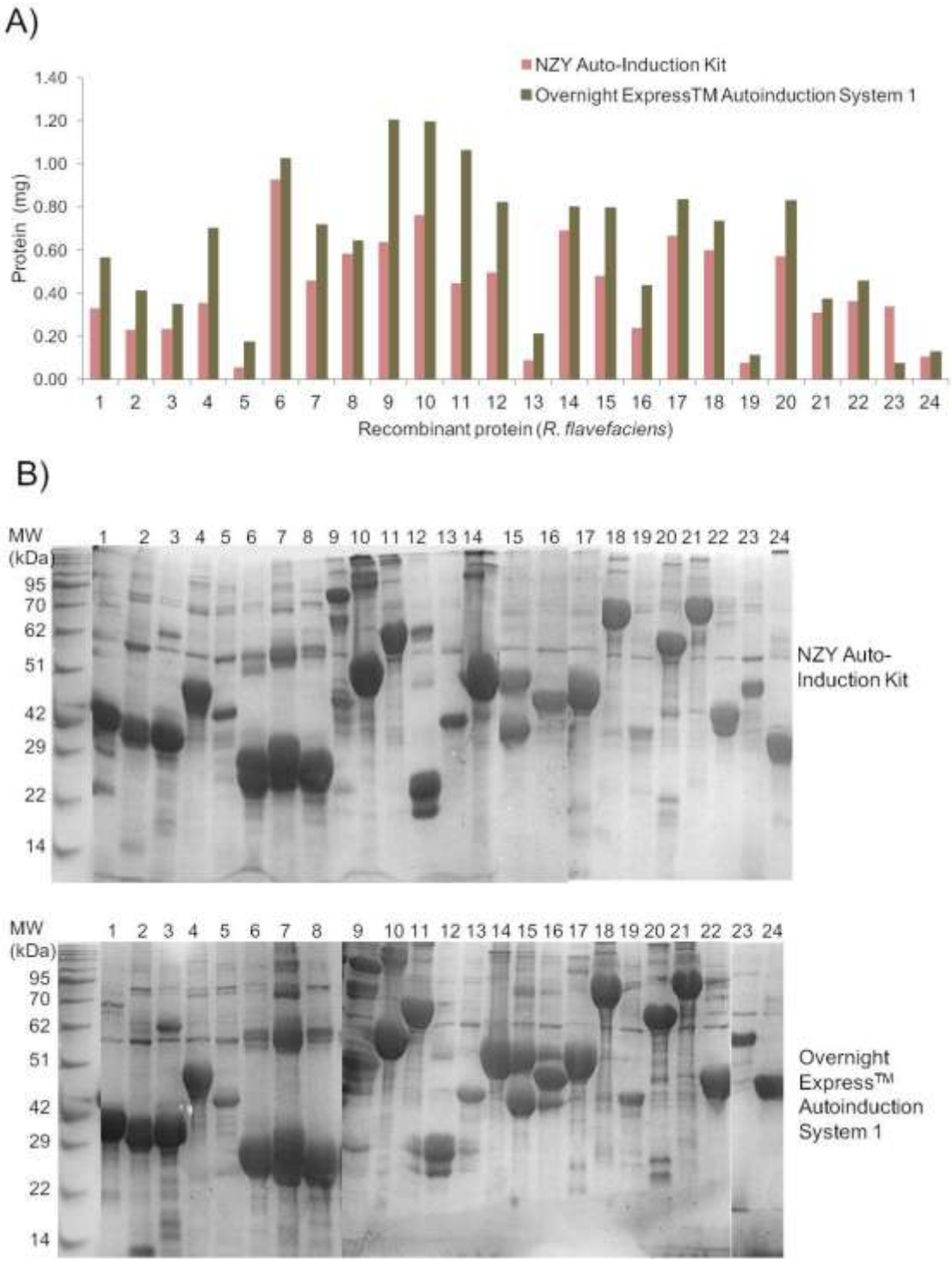
The 24 recombinant *R. flavefaciens* proteins obtained from *E. coli* BL21(DE3) grown in the NZY Auto-Induction LB medium (NZYTech, genes & enzymes) and in the Overnight Express™ Instant TB Medium (Novagen®) were purified through IMAC and levels of protein obtained evaluated (**A**) while the degree of purification was confirmed through SDS-PAGE (**B**).

Figure 3.4| Levels of purified protein obtained from 24 different recombinant *E. coli* BL21(DE3)pLysS strains grown in the NZY Auto-Induction LB medium or in the Overnight Express™ Instant TB Medium.



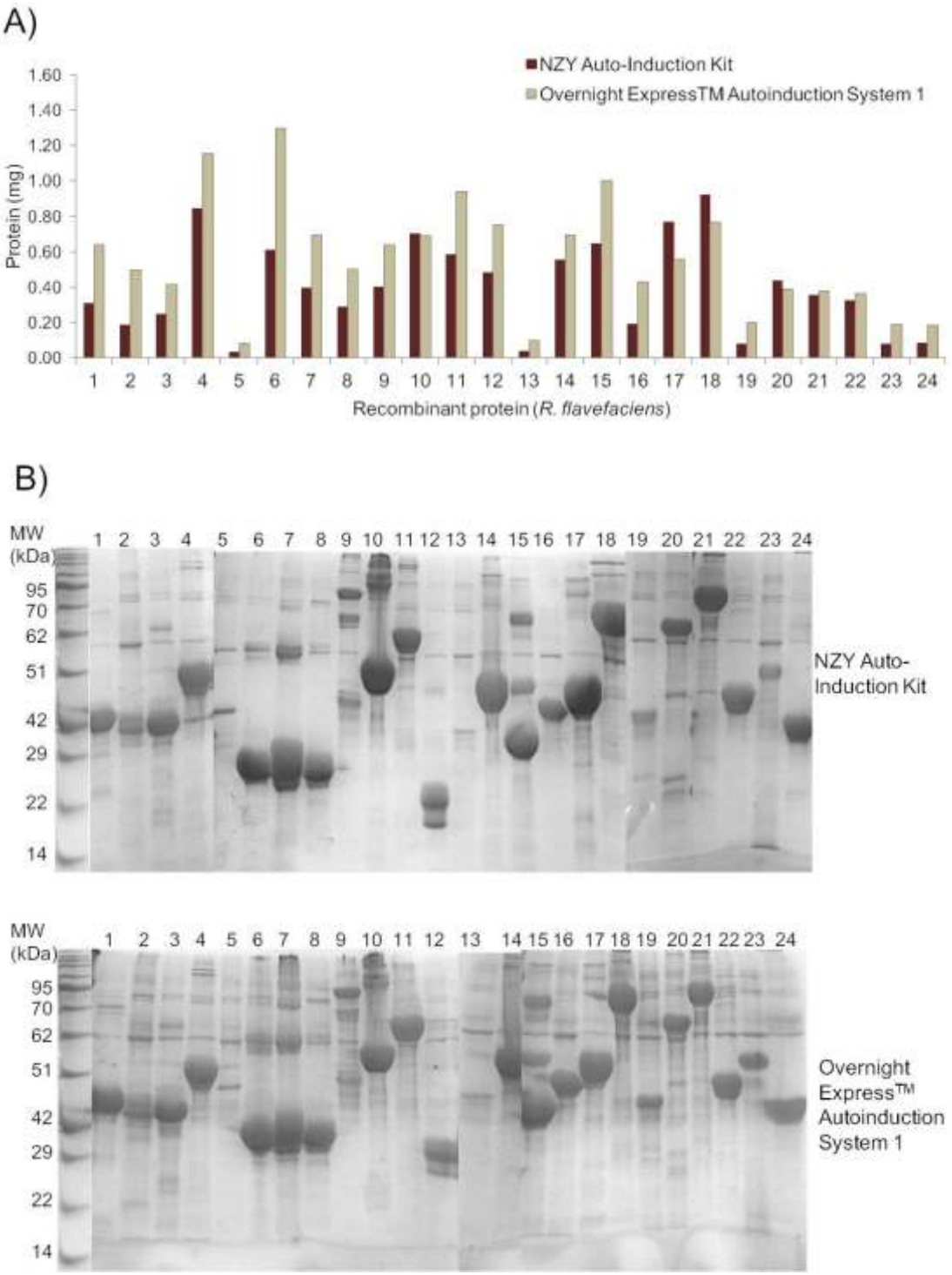
The 24 recombinant *R. flavefaciens* proteins obtained from *E. coli* BL21(DE3)pLysS grown in the NZY Auto-Induction LB medium (NZYTech, genes & enzymes) and in the Overnight Express™ Instant TB Medium (Novagen®) were purified through IMAC and levels of protein obtained evaluated (**A**) while the degree of purification was confirmed through SDS-PAGE (**B**).

Figure 3.5| Levels of purified protein obtained from 24 different recombinant *E. coli* BL21(DE3) strains grown in LB medium supplemented with the NZY Auto-Induction Kit or the Overnight Express™ Autoinduction System 1.



The 24 recombinant *R. flavefaciens* proteins obtained from *E. coli* BL21(DE3) grown in LB medium supplemented with the NZY Auto-Induction Kit (NZYTech, genes & Enzymes) and the Overnight Express™ Autoinduction System 1 (Novagen®). The 24 recombinant *R. flavefaciens* proteins were purified through IMAC and levels of protein obtained evaluated (**A**) while the degree of purification was confirmed through SDS-PAGE (**B**).

Figure 3.6| Levels of purified protein obtained from 24 different recombinant *E. coli* BL21(DE3)pLysS strains grown in LB medium supplemented with the NZY Auto-Induction Kit or the Overnight Express™ Autoinduction System 1.



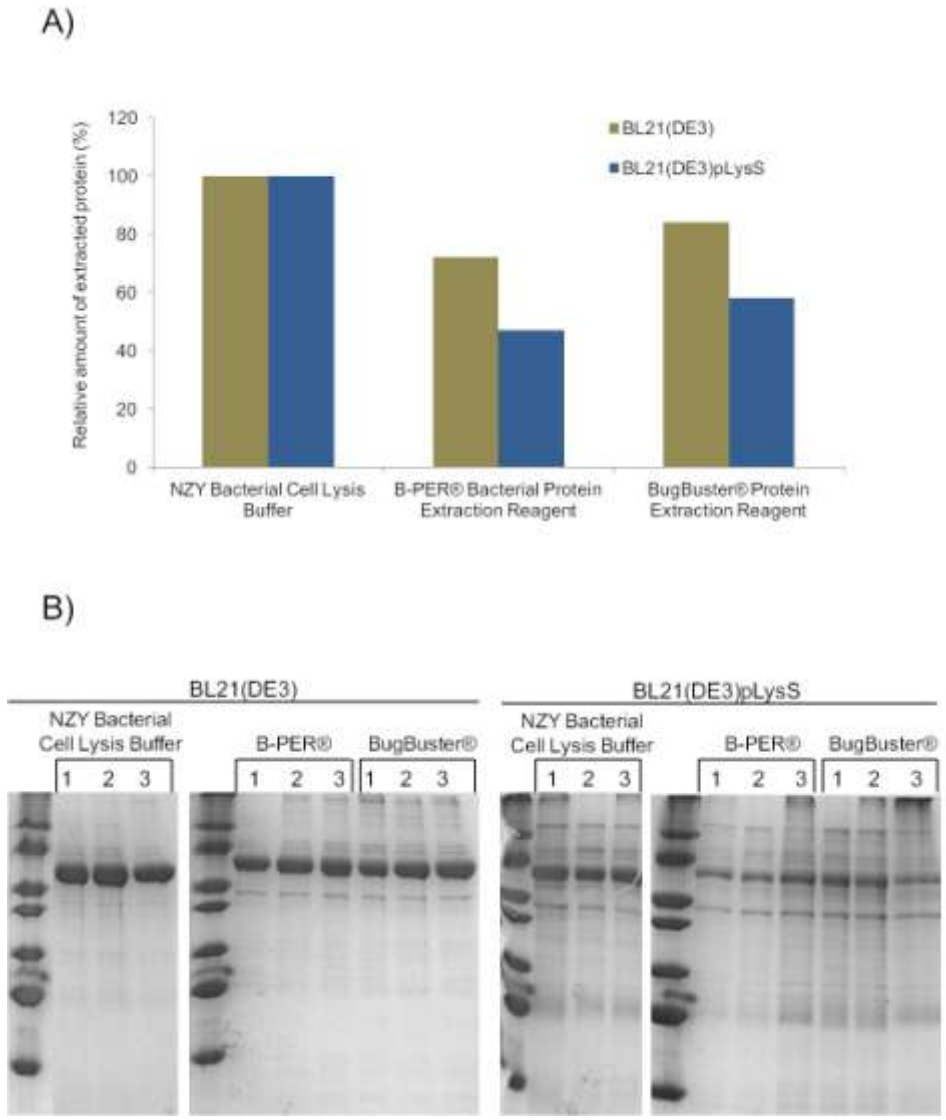
The 24 recombinant *R. flavefaciens* proteins obtained from *E. coli* BL21(DE3)pLysS grown in LB medium supplemented with the NZY Auto-Induction Kit (NZYTech, genes & Enzymes) and the Overnight Express™ Autoinduction System 1 (Novagen®). The 24 recombinant *R. flavefaciens* proteins were purified through IMAC and levels of protein obtained evaluated (**A**) while the degree of purification was confirmed through SDS-PAGE (**B**).

Data presented above suggest that NZY Auto-Induction LB medium is the most reliable medium to produce high levels of cellulosomal protein in *E. coli* and can be easily adaptable to a HTP approach. Thus, following the protocols described in the Materials and Methods section, 239 recombinant BL21(DE3)pLysS strains transformed with the plasmids encoding the cellulosomal modules of unknown function were grown in a small-scale (5 mL) in this medium. The cells were recovered by centrifugation and stored at -20 ºC before proceeding with lysis and purification (see below).

3.1.3.4. Efficient *E. coli* cell lysis and high-throughput protein purification

Three different cell lysis reagents were tested for the disruption of a recombinant *E. coli* strain expressing the *C. thermocellum* cellulase 8A (*Ct*Cel8A). The data, presented in Figure 3.7, revealed that yields of extracted protein using the NZY Bacterial Cell Lysis Buffer were higher than those obtained by the other two products for the BL21(DE3)pLysS cells, and slightly higher for BL21(DE3) cells. The use of pLysS or pLysE hosts can enhance the cell wall lysis procedure due the action of T7 lysozyme with amidase activity endogenously produced (Zhang & Studier, 2004). Thus, addition of lysozyme during extraction provides extra disruption level in BL21pLysS cells. However, excess amidase activity can cause spontaneous *E. coli* lysis during growth resulting in reduced expression levels (Studier, 1991). Nevertheless, yields of extracted protein after treatment with NZY Bacterial Cell Lysis Buffer were very similar for both BL21pLysS and BL21 expressing strains, even adding extra lysozyme. When comparing all three cell lysis detergents, complete cell lysis occurred after 10-15 minutes incubation at room temperature with gentle shaking in an orbital incubator (data not shown). To verify if the detergent extraction had not affected the biologic activity of the recombinant cellulase, the enzyme was purified and tested in the hydrolysis of hydroxyethylcellulose. The data (not shown) revealed that *Ct*Cel8A obtained after treatment with the three different chemical lysis buffers have similar activity to the recombinant enzyme obtained after cell disruption by sonication.

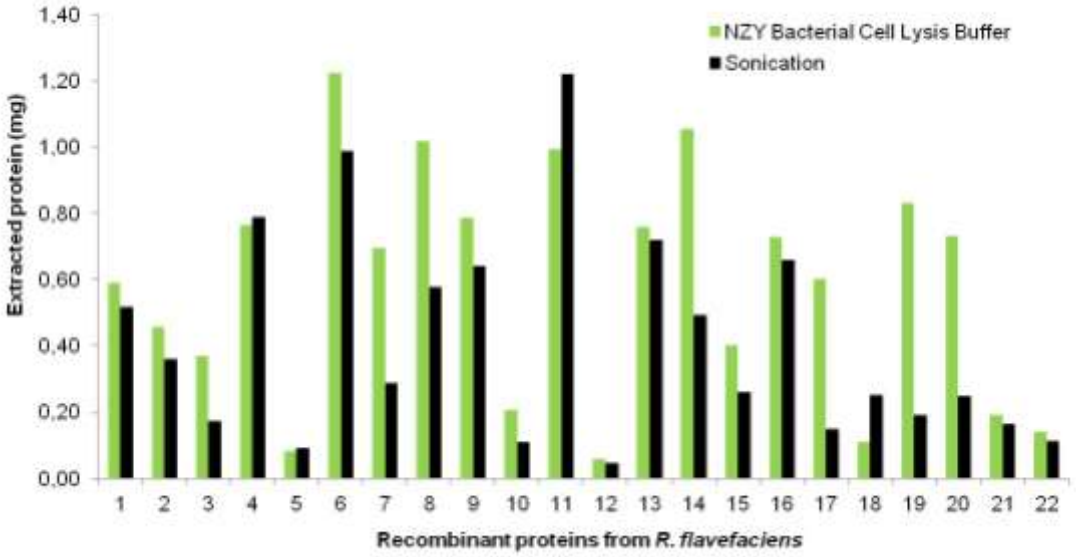
Figure 3.7| Comparing the efficiency of cell lysis using NZY Bacterial Cell Lysis Buffer, B-PER® Bacterial Protein Extraction Reagent and BugBuster® Protein Extraction Reagent.



E. coli BL21(DE3) and BL21(DE3)pLysS cells harvested from 5 mL of cultured media were lysed (in triplicates) using three different protein extraction chemicals: NZY Bacterial Cell Lysis Buffer (NZYTech, genes & enzymes), B-PER® Bacterial Protein Extraction Reagent (Thermo Scientific) and BugBuster® Protein Extraction Reagent (Novagen®). (**A**) Levels of extracted protein obtained were evaluated. (**B**) The recombinant protein was purified through IMAC and separated through SDS-PAGE.

In addition, the efficacy of detergent lysis (NZY Bacterial Cell Lysis Buffer) was compared with the standard mechanical procedure (sonication) in the extraction of 22 different recombinant *R. flavefaciens* proteins produced in *E. coli* BL21(DE3). Overall, the data, presented in Figure 3.8, revealed that cell-free extracts from the chemical lysis showed higher to similar levels of target protein extracted when compared with sonication. A detailed analysis of data collected revealed that only for four proteins in test, sonication performed slightly better than the cell lysis extraction reagent.

Figure 3.8| Comparing the efficiency of cell lysis using the NZY Bacterial Cell Lysis Buffer or a mechanical procedure (sonication).



Levels of extracted protein obtained from 22 different recombinant *E. coli* BL21(DE3) cultures using the NZY Bacterial Cell Lysis Buffer (NZYTech, genes & enzymes, Portugal) or sonication.

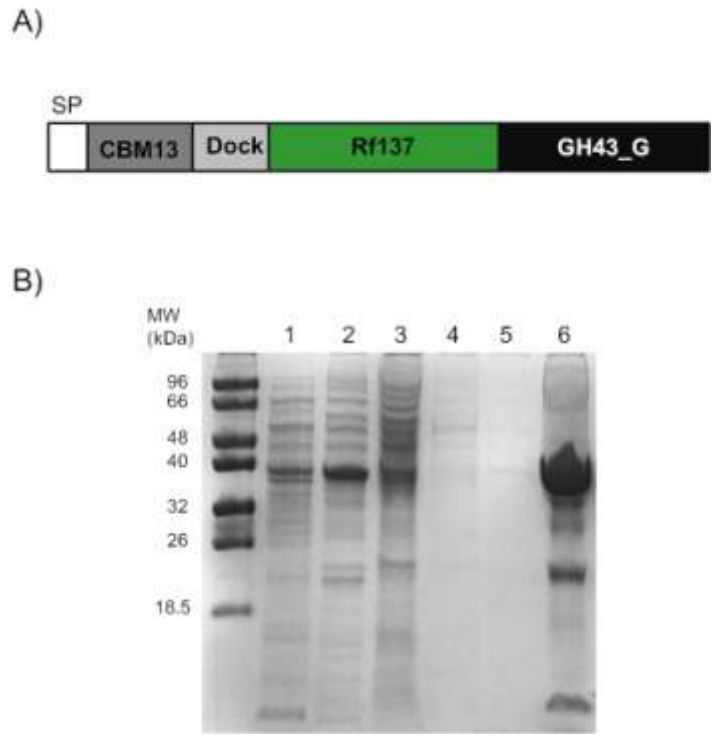
Data presented above confirm that detergent lysis is as efficient as mechanical lysis to disrupt *E. coli* cells while retaining the biological activity of the recombinant proteins. Detergent lysis is also more amenable in HTP approaches. Thus, the 239 recombinant BL21(DE3)pLysS cell pellets from cultures grown previously were lysed in 1 mL of NZY Bacterial Cell Lysis Buffer, as described in Materials and Methods section. Recombinant proteins were purified by IMAC and eluted in 0.15 mL of elution buffer. In total, 184 (77%) out of 239 proteins were obtained in significant levels in a purified form (> 90% pure as determined by SDS-PAGE analysis). The primary sequences and the molecular architecture of these 184 expressed cellulosomal proteins from which the modules of unknown function were isolated are available in Table S3.2 (see in Annex). The remaining proteins presented none or low production levels as a consequence of their accumulation in inclusion bodies, lower RNA stability or improper protein folding within the *E. coli* cytoplasm (data not shown).

3.1.3.5. Discovery of novel α-L-arabinofuranosidases in cellulosomes

The 184 purified cellulosomal modules of unknown function were screened for α-L-arabinofuranosidase activity (EC 3.2.1.55) using 4-nitrophenyl-*α*-arabinofuranoside (*p*NPAf) as the substrate in a 96-well plate format. Small-scale affinity purification resulted in 60 µg to 2 mg of recombinant proteins in 0.15 mL final volume, which was sufficient for the activity screening protocol. The activity screen revealed that protein Rf137, an unknown domain located in the cellulosome of *R. flavefaciens* (see Table S3.2, in Annex), displays α-L-arabinofuranosidase activity. Primary sequence analysis on the enzyme containing the Rf137

module revealed the presence of an N-terminal signal peptide followed by a family 13 CBM, an internal type I dockerin, the Rf137 domain and a C-terminal GH43_G catalytic domain (Figure 3.9A). The recombinant protein displays molecular size of 37.78 kDa which is in close agreement with the observed molecular mass on SDS-PAGE analysis (Figure 3.9B).

Figure 3.9| Properties of the protein Rf137 from *R. flavefaciens* cellulosome.

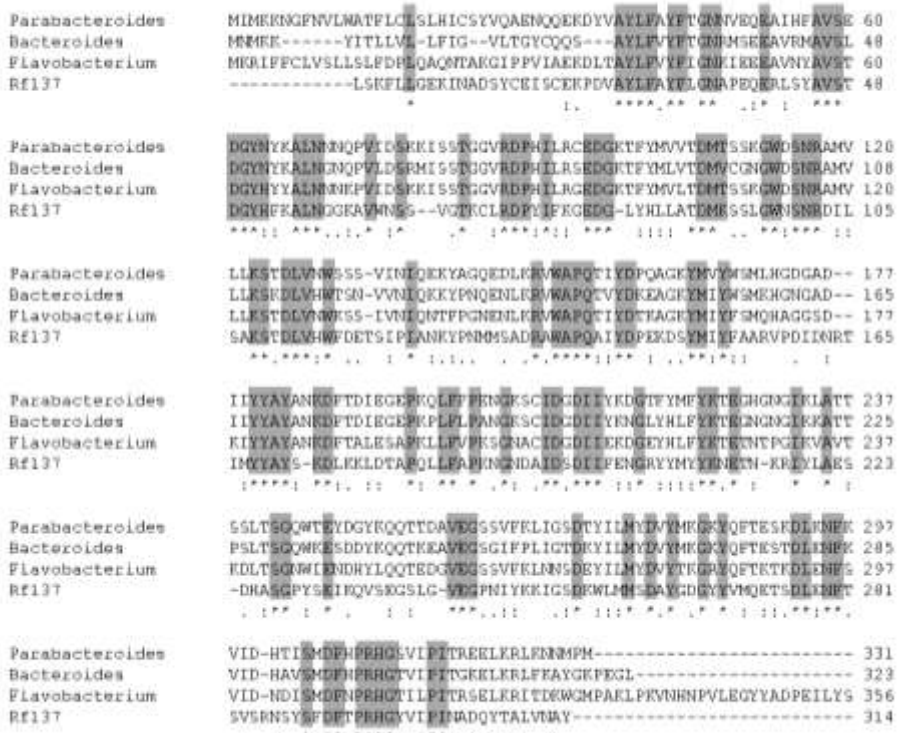


(**A**) Molecular architecture of modular Rf137. SP, signal peptide; Dock, dockerin module. (**B**) Expression and purification of Rf137. The recombinant protein was expressed in *E. coli* BL21(DE3)pLysS strain and purified through IMAC. Lane 1: insoluble protein cellular extract (pellet); Lane 2: soluble protein cellular extract (supernatant); Lane 3: sample collected after supernatant filtration and passed through the affinity column; Lane 4: sample collected after a first wash step with 50 mM NaHEPES, pH 7.5, 1 M NaCl, 10 mM Imidazole, 5 mM $CaCl_2$; Lane 5: sample collected after a second wash step with 50 mM NaHEPES, pH 7.5, 1 M NaCl, 35 mM Imidazole, 5 mM $CaCl_2$; Lane 6: sample of purified Rf137.

Sequence alignment using the BLAST tool (www.ncbi.nlm.nih.gov/BLAST), revealed that there are more than 30 proteins of unknown function and not previously classified in any GH family, which share >30% identity with the Rf137, suggesting that this enzyme is the founding member of a novel GH family (Figure 3.10A). Nevertheless, the novel family is distantly related with GH43 (Figure 3.10B) and GH62 (data not shown) suggesting that the novel family is part of Clan GH-F and should display the characteristic 5-bladed β-propeller fold.

Figure 3.10| Multiple Sequence Alignment of Rf137 performed by CLUSTALW program.

A)



B)

Identical amino acid residues common in all sequences are represented by grey shade. (**A**) Rf137 alignment was with the following proteins sharing >30% identity: protein from *Parabacteroides gordonii* (sequence ID WP_028729310.1); protein from *Bacteroides finegoldii* CL09T03C10 (sequence ID gb|EKJ90516.1); protein from *Flavobacterium* sp. F52 (sequence ID WP_008467609.1). (**B**) Rf137 alignment was with the appended C-terminal GH43 module.

Analysis of Rf137 degradation products when the enzyme is attacking arabinogalactan (from wheat) revealed the presence of arabinose as the sole reaction product, suggesting that Rf137 acts in the removal of arabinose side chains from this complex polysaccharide (Figure 3.11).

Figure 3.11| Thin layer chromatography (TLC) showing the enzymatic degradation products of wheat arabinogalactan.

TLC analysis of hydrolysis products from wheat arabinogalactan by Rf137. Lane 1: Galactose standard; Lane 2: Arabinose standard; Lane 3: Rf137 reaction products; Lane 4: Rf138 reaction products (negative control).

3.1.4. Conclusions

As a result of the dramatic increase in the number of genome and metagenome sequences available, there is a significant growing pressure to develop appropriate high-throughput methods to uncover the functional and structural novelty of the proteomic information revealed by genomics. Particularly, CAZymes are being discovered at a breakneck speed, especially due to the availability of complete genome sequences of many cellulosomal organisms, coupled to current interest on the efficient conversion of biomass for biofuels production. As a result, the number of proteins without functional characterization in the CAZy database has been increasing. Here we report the development of a high-throughput (HTP) platform for the efficient cloning, expression and purification of a large number of cellulosomal modules of unknown function in *Escherichia coli* at a scale that can be applied in an enzyme functional screen. The cloning approach developed uses simple methodologies and is easily adaptable to an automatic system to test many different genes simultaneously. Furthermore, it allows directional cloning and high efficiencies for the generation of expression clones, even using large DNA fragments. Thus, the labor required for the selection and validation of recombinant clones is much reduced. The designed protocol for growth and expression has been found to be optimal to produce recombinant cellulosomal proteins. Using auto-induction media, cultures can reach OD_{600nm} of 13–14, with corresponding high yields of recombinant protein production. In addition, chemical lysis offers an efficient, rapid and an automatable procedure when purifying multiple proteins in parallel. Routinely, we can obtain yields up to 2 mg of purified proteins in 0.15 mL final volume, from 5 mL cultures. It should be beared in mind that once the proteins have been expressed in soluble form they can be produced at large-scale using the same miniaturized growth conditions. Thus, it is evident that automation is essential in proteomic studies involving dozen to thousands of proteins by providing the required throughput. The platform developed

here uses a robot to perform cloning, expression and purification on the order of hundreds of targets. However, even without the convenience of automated liquid handling equipment, it is possible to carry out this method by using multichannel pipettes, multi-well plates and a manifold vaccum apparatus. The colorimetric screening protocol established here allowed the rapid discovery of a novel α-L-arabinofuranosidase from a pool of 184 expressed proteins of unknown function. The novel CAZyme is the founder member of a novel GH family belonging to the clan GH-F. Although centered in the discovery of novel α-L-arabinofuranosidases, other CAZyme activities can be screened by using the methodologies reported in this study. In addition, this platform could be applied for the production of proteins with different origins and functions, providing enough protein quantities to establish functional screens that could support the discovery of novel enzymes.

4. EXPRESSION AND SOLUBILITY TAGS FOR *Escherichia coli*

4.1. Identification of novel solubility tags for recombinant protein expression in *Escherichia coli*

Vania O. Fernandes[1,2], Ana Filipa Sequeira[1,2], Joana L. A. Brás[2], Catarina I. P. D. Guerreiro[2], Luís. M. A. Ferreira[1,2], Carlos. M. G. A. Fontes[1,2]

[1] CIISA-Faculdade de Medicina Veterinária, Universidade de Lisboa, Av. da Universidade Técnica, 1300-477 Lisboa, Portugal; [2] NZYTech Lda, Estrada do Paço do Lumiar, Campus do Lumiar, Edifício E-R/C, 1649-038 Lisboa, Portugal

Adapted from a manuscript in preparation

Abstract

Production of recombinant proteins at high yields and in the soluble and correctly folded form in *Escherichia coli* is essential in biochemistry studies. Fusion protein biotechnology, in which recombinant proteins are expressed with a fusion tag, opens the possibility to produce stable targets at significant levels and in the biologically active form. Fusion tags can enhance protein expression, stability and solubility. Here we describe the development of a battery of novel expression vectors containing a range of different solubility tags, including three novel ones: RF1, RF47 and CEL. The 12 expression vectors were used to express 8 different proteins displaying different degrees of insolubility when expressed in *E. coli*. The 96 (12 x 8) recombinant plasmids were used to transform *E. coli* BL21(DE3) strain, which was grown under five different culture conditions. The data suggest that tags RF1 and RF47, in contrast to CEL, are highly efficient to generate the production of high levels of soluble proteins. However, from the 12 tags analyzed there is no single tag that performs universally well in raising the percentage of protein solubility. Thus, considering that the efficacy of the solubility tags depends on protein type, data presented here suggest that the best tag for a specific protein needs to be selected from a battery of tags and expressing conditions. To achieve this aim we report the generation of a general cloning strategy, where the target gene is initially cloned into an entry cloning vector (pHTP28) from which it can be easily transferred to a range of different expression vectors for tag selection.

4.1.1. Introduction

In recent years, the availability of genomic and metagenomic sequence data increased dramatically. One of the major findings that results from the annotation of sequencing information is the presence of a large number of gene-encoded products of unknown function. Thus, it is clear that in the post-genomic era substantial efforts are required to investigate the function of increasing numbers of uncharacterized proteins revealed by genome sequencing projects. Recently, a series of novel methodological developments have come forward to allow the analysis of large numbers of genes/proteins. Ligation-independent cloning techniques enable the rapid cloning of multiple genes simultaneously (Hartley, Temple & Brasch, 2000; Berrow, Alderton, & Owens, 2009; Tachibana *et al*., 2009). In addition, miniaturization of cell growth conditions and automation of protein purification protocols allow the rapid production of many hundreds to thousands of target proteins (Chambers, 2002; Dieckman *et al*., 2002; Scheich, Sievert & Büssow, 2003). To improve the yields of protein production, expression conditions (e.g. temperature, time, culture media, host, expression plasmid) need to be optimized to generate soluble proteins in sufficient amounts for the required downstream applications. By using high-throughput tools, a large number of target proteins can be screened simultaneously and a large number of expression and purification conditions can be tested (Bruni & Kloss, 2013).

Selection of the adequate host strain is a key step when designing a bioprocess for recombinant protein production. In the majority of situations, *Escherichia coli* remains the favorite choice for protein production due to its well-characterized genetics, the commercial availability of a range of strains and vectors, low production costs, ease of use, and medium to high yields of protein production (Baneyx, 1999b; Mancia & Love, 2011; Rosano & Ceccarelli, 2014). There are several examples where *E. coli* has been successfully used to produce heterologous proteins at an industrial scale, including the biopharmaceutical and biofuel industries (Huang, Lin & Yang, 2012; Chen *et al*., 2013). However, lack of post-translational machinery and protein accumulation into insoluble and biologically inactive inclusion bodies (IB) constitute a major challenge when overexpressing heterologous proteins, mainly from eukaryotic origin, in *E. coli* (Demain & Vaishnav, 2009; Williams *et al*., 1982). Other problems include poor expression, protein degradation as a result of the bacterial stress response, toxicity and loss of function (Peleg & Unger, 2012). Several strategies have been designed to overcome some of these bottlenecks. To prevent protein aggregation and formation of IB, the strategies available rely on covering their external hydrophobic moieties and include the use of genetic manipulated strains, introduction of chaperone molecules, modification of growth conditions, and fusion protein production (Schein, 1989; Baneyx, 1999b; Sørensen & Mortensen, 2005; Chou, 2007).

The fusion technology by which a recombinant protein is expressed fused to a highly soluble and stable protein partner has recently received considerable attention for solubility enhancement. Fusion tags are commonly proteins or peptides that are expressed at high levels in *E. coli* and are attached to the N- or C-terminus of poorly expressed recombinant proteins conferring mRNA stability while providing improved folding (Arechaga *et al*., 2003; Englander, 2000; Creighton, 1997). Furthermore, although it is still ambiguous how fusion tags act to enhance solubility, it is known that some fusion tags have an intrinsic chaperone-like activity (Kapust & Waugh, 2000; Raran-Kurussi & Waugh, 2012), while others can drive the appended protein into folding pathways mediated by chaperones (Costa *et al*., 2014). The partners used for fusion protein expression to enhance solubility in *E. coli* vary among large molecules, such as maltose-binding protein (MBP) (di Guana *et al*., 1988), glutathione S-transferase (GST) (Smith & Johnson 1988) or N-utilization substance protein A (NusA) (Davis *et al*. 1999), to small peptides, such as thioredoxin (Trx) (LaVallie *et al*. 1993), small ubiquitin-like modifier (SUMO) (Baker, 1996) or Fh8 (Costa *et al*. 2013), and also include disulfide isomerases (e.g. DsbA and DsbC) (Collins-Racie *et al*.,1995; Nozach *et al.* 2013). In general, selection of the most appropriate fusion partner for a poorly expressed or insoluble protein requires the consideration of several factors, such as: the characteristics of the target protein (e.g. stability, hydrophobicity, size), the expression system used and the end use of the recombinant protein (Terpe, 2003). Strategies for the efficient removal of the fusion tag should also be considered for therapeutic proteins, or when structural or biochemical studies are required (Balbás, 2001). It should be taken in consideration that the solubility, as well as the functional activity of the target protein can change after cleavage of the soluble fusion partner (Koehn & Hunt, 2009). The portfolio of fusion tags currently available has been growing, as new fusion systems are constantly emerging.

Here, we have constructed a novel series of prokaryotic expression vectors adapted to a ligation-independent cloning procedure previously described (Fernandes *et. al*., Chapter 3), which allows a large flexibility in the selection of the best fusion tag to increase protein expression and solubility in *E. coli*. The novel vectors were constructed based on the pHTP backbone (Fernandes *et. al*., Chapter 3), by inserting widely tested solubility tags, as well as novel system tags developed in this work. Two recombinant proteins highly expressed in *E. coli* (Rf1 and Rf747) were tested as fusion partners to improve solubility. In addition, a system consisting of a mini-cellulosome and based on the high affinity interactions established between *Clostridium thermocellum* cohesins and dockerins was developed to attempt reduce problems related with protein aggregation and potentiate correct folding. The mini-cellulosome (here named CEL) was also inserted into the pHTP plasmid to serve as fusion partner of recombinant proteins. This study provides a comparison of the novel fusion tags with other established solubility partners to verify their future applicability in the fusion protein technology. Eight target proteins, seven previously described as difficult to express

and one highly expressed in *E. coli* to serve as a control (green fluorescent protein, here named protein G) were cloned into the novel pHTP-derivatives and expressed as fusions with Trx, GST, MBP, NusA, SUMO, DsbA, DsbC, Fh8, CEL, Rf1, and Rf47. The expression vectors were transformed in *E. coli* BL21(DE3) strains and expression and solubility of the fusion proteins produced were evaluated.

4.1.2. Materials and Methods

4.1.2.1. Construction of pHTP-derivative vectors for protein expression in *E. coli*

Confidential

Table 4.1| Primers used for the construction of N-terminal fusion proteins. The *Nco*I restriction site included is underlined.

Vector		Primer sequence (5'→3')	Direction
pHTP-Trx	TRX_F	CACACCATGGGTAGCGATAAAATTATTCACCTG	Forward
	Trx_R	CACACCATGGCAGAACCGGCCAGGTTAGCGTCG	Reverse
pHTP-GST	GST_F	CACACCATGGGTTCCCCTATACTAGGTTATTGG	Forward
	GST_R	CACACCATGGCAGAACCATCCGATTTTGGAGGATGG	Reverse
pHTP-MBP	MBP_F	CACACCATGGGAAAGAAAGGTTTTATGTTGTTTAC	Forward
	MBP_R	CACACCATGGAAGTCTGCGCGTCTTTCAGG	Reverse
pHTP-NusA	NusA_F	CACACCATGGGTAAAGAAATTTTGGCTGTAGTTG	Forward
	NusA_R	CACACCATGGCACTAGTCGCTTCGTCACCGAAC	Reverse
pHTP-SUMO	SUMO_F	CACACCATGGGTGGGTCCCTGCAGGACTCAGAAG	Forward
	SUMO_R	CACACCATGGCACCTCCAATCTGTTCGCGGTG	Reverse
pHTP-DsbA	DsbA_F	CACACCATGGGTAAAAAGATTTGGCTGGCGCTGG	Forward
	DsbA_R	CACACCATGGCTGATCCTTTTTTCTCGCTTAAG	Reverse
pHTP-DsbC	DsbC_F	CACACCATGGGAAAGAAAGGTTTTATGTTG	Forward
	DsbC_R	CACACCATGGATGATCCTTTACCGCTGG	Reverse
pHTP-Fh8	Fh8_F	CACACCATGGGTTCCCCTAGTGTTCAAGAGGTTG	Forward
	Fh8_R	CACACCATGGCTGACAAAATCGAAACGAGTTC	Reverse
pHTP-CEL	CEL_F	CACACCATGGGTGGCAGCAGCCATCACCATCACC	Forward
	CEL_R	CACACCATGGCGCTCGGGATGTCCGTGCCCAC	Reverse
pHTP-Rf1	Rf1_F	CACACCATGGGTGGCGAGTGTCACGGCTATATCG	Forward
	Rf1_R	CACACCATGGCAACGATTGAGTAATCCTTG	Reverse
pHTP-Rf47	Rf47_F	CACACCATGGGTGTCAAAAAGGAAACTCCTAACC	Forward
	Rf47_R	CACACCATGGCAGTTGCTGTATTGAATATCTTTG	Reverse

After gene isolation or synthesis, the nucleic acids encoding the different fusion tags were digested with *Nco*I restriction enzyme and ligated with similarly digested pHTP plasmid with Speedy Ligase (NZYTech, genes & enzymes, Portugal) to generate plasmids pHTP-CEL, pHTP-Trx, pHTP-GST, pHTP-MBP, pHTP-NusA, pHTP-SUMO, pHTP-DsbA, pHTP-DsbC, pHTP-Fh8, pHTP-Rf1 and pHTP-Rf47, respectively. The sequence of the all novel expression vectors was confirmed by DNA sequencing and their properties are summarized in Table 4.2.

Table 4.2| Properties of the novel prokaryotic expression vectors.

Vector	Fusion Protein	Tag size (nt)	MW tag (kDa)	p*I*	GRAVY	Tag position	Resistance
pHTP	His_6-SSGPQQGLR	45	1.75	9.58	-2.060	N-terminal	kan
pHTP-Trx	**Thioredoxin**-MGSS-His_6-SSGPQQGLR	396	14.18	5.92	-0.213	N-terminal	kan
pHTP-GST	**Glutathione S-transferase**-MGSS-His_6-SSGPQQGLR	729	28.69	6.45	-0.466	N-terminal	kan
pHTP-MBP	**Maltose binding protein**-MGSS-His_6-SSGPQQGLR	1221	44.61	5.79	-0.328	N-terminal	kan
pHTP-NusA	**N-utilization substance A**-MGSS-His_6-SSGPQQGLR	1551	57.17	4.67	-0.320	N-terminal	kan
pHTP-SUMO	**Small ubiquitin-like modifier**-MGSS-His_6-SSGPQQGLR	366	13.75	5.87	-0.924	N-terminal	kan
pHTP-DsbA	**Disulfide oxidoreductase DsbA**-MGSS-His_6-SSGPQQGLR	693	25.47	6.75	-0.251	N-terminal	kan
pHTP-DsbC	**Disulfide-bond isomerase DsbC**-MGSS-His_6-SSGPQQGLR	777	28.01	6.86	-0.222	N-terminal	kan
pHTP-Fh8	***Fasciola hepatica* calcium-binding protein (Fh8)**-MGSS-His_6-SSGPQQGLR	270	9.89	6.70	-0.783	N-terminal	kan
pHTP-CEL	**Dockerin**-MGSS-His_6-SSGPQQGLR	288	10.42	7.02	-0.498	N-terminal	kan
pHTP-Rf1	***R. flavefaciens* celullosomal protein** -MGSS-His_6-SSGPQQGLR	870	31.94	4.79	-0.588	N-terminal	kan
pHTP-Rf47	***R. flavefaciens* celullosomal protein**-MGSS-His_6-SSGPQQGLR	815	30.24	5.68	-0.612	N-terminal	kan

nt, nucleotides; MW, molecular weight; kDa, kiloDalton; p*I*, isoelectric point; GRAVY, grand average of hydropathicity; kan, kanamycin

4.1.2.2. Construction of the pHTP28 cloning vector

Confidential

4.1.2.3. Cloning of the target genes into pHTP vectors

The genes encoding seven proteins that were previously shown to form IB when expressed in *E. coli* (here named A, B, C, D, E, F, and H) were cloned into each one of the 11 novel pHTP-derivative vectors and also pHTP by a LIC-based method as described elsewhere (Fernandes *et. al*., Chapter 3). The selected proteins have different origins, including carbohydrate-active enzymes, carbohydrate-binding modules and antigens (Table 4.3). In order to serve as a control of the cloning, expression and purification steps, we included a protein that is produced in the soluble form in *E. coli* under known conditions; the selected protein was the green fluorescent protein, here named G, which encoding gene was synthetically modified for optimized expression in this host. Thus in total, 8 different genes were cloned into 12 different vectors providing a total of 96 cloning reactions. Briefly, the selected coding sequences for the 8 target proteins were amplified by PCR from appropriate templates using primers containing an extension sequence comprising 16 bp homologous to the ends of the pHTP cloning region that allow the cloning by base complementation (Table 4.4). DNA amplification was carried out by NZYProof DNA polymerase (NZYTech, genes & enzymes, Portugal) in 50 µL final volume. After removal of nucleotides and eventual primer dimers using the NZYGelpure kit (NZYTech, genes & enzymes, Portugal) the PCR products were cloned into the different vectors using 240 ng of each circular plasmid, 1 µL of enzyme mix and 2 µL of reaction buffer 10×. Cloning reactions were carried out in 20 µL final volume on a thermal cycler programmed as follows: 37 ºC for 1 hour; 80 ºC for 10 minutes and 30 ºC for 10 minutes. The reaction mixtures were used to transform DH5α *E. coli* competent cells (NZYTech, genes & enzymes, Portugal). Two colonies were picked for each construct and the presence of insert confirmed by PCR using two vector specific primers, T7 and pET24a forward and reverse primers, respectively. NZYLong 2x Green Master Mix (NZYTech, genes & enzymes, Portugal) was used for the verification PCR reaction.

Table 4.3 | Properties of the protein targets used for the expression assay.

Target protein	Organism	Gene	Access Number	Gene Size (nt)	Protein size (aa)	MW (kDa)	p*I*	GRAVY
A	*Clostridium thermocellum*	Putative carbohydrate-active enzyme	CP002416.1	747	249	28.50	4.72	-0.300
B	*Clostridium thermocellum*	Rhamnogalacturonan lyase 11A	CP000568.1	1702	567	61.98	5.18	-0.443
C	*Ruminococcus flavefaciens*	Putative carbohydrate-active enzyme	WP_037281654.1	627	209	23.60	6.17	-0.715
D	*Cellvibrio japonicus*	CBM2A from Xylanase 10A	X15429.1	303	101	10.62	8.05	-0.525
E	*Clostridium cellulolyticum*	CBM3	ACL75844.1	447	149	16.57	4.95	-0.627
F	*Trichomonas vaginalis*	Immuno-dominant variable surface antigen	XP_001330197.1	1782	594	68.35	7.96	-0.476
G	*Aequorea coerulescens*	Green fluorescent protein	AAN41637.1	714	238	26.76	5.72	-0.501
H	*Cellvibrio japonicus*	Chitin-binding protein	WP_012488716.1	1011	337	36.27	6.29	-0.511

CBM, carbohydrate-binding module; nt, nucleotides; aa amino acids; MW, molecular weight; kDa, kiloDalton; p*I*, isoelectric point; GRAVY, grand average of hydropathicity

In this study, a stop codon was included at the end of all genes such that the His_6 tag was engineered at the N-terminus of the proteins derived from pHTP vector or was located internally between the two protein modules in fusion proteins derived from all other vectors. These experiments culminated in the generation of 96 recombinant plasmids resulted from the cloning of the 8 different genes in the 12 vectors.

Table 4.4 | Primers used for cloning of the protein targets.

Target protein	Primer sequence (5'→3')	Direction
A	TCAGCAAGGGCTGAGGGCCTATCTGGATAATGAGCTG	Forward
	TCAGCGGAAGCTGAGGTTATTTTTGAATAACTTCAAACATTGG	Reverse
B	TCAGCAAGGGCTGAGGGCTGGTGCGCGTCAGATGG	Forward
	TCAGCGGAAGCTGAGGTTACGGCACAAGGTAAATATTTGG	Reverse
C	TCAGCAAGGGCTGAGGGATTTCAGCTACTCATCCAATG	Forward
	TCAGCGGAAGCTGAGGTCAGGCCTTGCTGTACTCGAAG	Reverse
D	TCAGCAAGGGCTGAGGGCAACTTGCAGTTATAACATTACC	Forward
	TCAGCGGAAGCTGAGGTTACACAGATCCCGAGCAGATAC	Reverse
E	TCAGCAAGGGCTGAGGTTTAACGCCACCACCAGCGCGACGAC	Forward
	TCAGCGGAAGCTGAGGTTATGGCTCAATACCGCCGATTAAGTTGCC	Reverse
F	TCAGCAAGGGCTGAGGGGCATCAATACAGTTCAAGTAC	Forward
	TCAGCGGAAGCTGAGGTTATTTCTCTCCATTTACTTTATCTTTAAG	Reverse
G	TCAGCAAGGGCTGAGGGTTAGCAAAGGTGAAGAACTG	Forward
	TCAGCGGAAGCTGAGGTTATTTGTACAGTTCATCCATGCC	Reverse
H	TCAGCAAGGGCTGAGGATGAAATACCTGCTGCCGACC	Forward
	TCAGCGGAAGCTGAGGTTAGTGGTGGTGGTGGTGGTG	Reverse

4.1.2.4. Recombinant protein expression

The 96 recombinant plasmids were used to transform BL21(DE3) *E. coli* cells (NZYTech genes & enzymes, Portugal). Recombinant strains were grown in 5 mL of NZY Auto-Induction LB medium (NZYTech, genes & enzymes, Portugal) or Luria-Bertani (LB) broth medium, both supplemented with kanamycin (50 µg/mL). Growth was carried out in 24-deep-well plates sealed with a gas-permeable adhesive in a microplate shaker. Cells were grown at 37 ºC till mid-exponential phase (OD_{600nm} of 0.4-0.6) and gene expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. After induction, cells were further grown at either 16 or 37 ºC for 16 hours. In addition, cells grown in auto-induction media were also cultivated for 20 hours at 37 ºC without IPTG induction. Cells were harvested by centrifugation at 2.500 ×*g* for 10 min (4 ºC). Thus, the 96 recombinant BL21(DE3) *E. coli* strains were grown in five different culture conditions.

4.1.2.5. Expression and solubility analysis

Cells from 5-mL culture pellets were resuspended in 1 mL of NZY Bacterial Cell Lysis Buffer supplemented with 0.1 mg/mL lysozyme and 0.004 mg/mL DNase I (NZYTech, genes & enzymes, Portugal). Cell disruption was conducted at room temperature in a microplate shaker until a clear lysate was observed.

Crude lysates (40 µL) were transferred to 1.5 mL microcentrifuge tubes and the insoluble cell debris were collected by centrifugation at 12,000 x*g* (4 ºC) for 20 min. Approximately 40 µL of the supernatants containing the cell-free extract (Soluble protein fraction, S) were collected in

a new microcentrifuge tube. The cell pellet, containing the IB, was resuspended in 40 µL of water (Insoluble protein fraction, I). Polypeptides contained in both the S and I fractions were separated on 14% (w/v) polyacrylamide gels (SDS-PAGE) and detected with Coomassie brilliant blue. The level of solubility was scored (0, +, ++, and +++) by comparison of levels of soluble protein (S) with insoluble proteins (I) after SDS-PAGE analysis. The score 0 corresponds to no soluble expression; a score of + corresponds to < 50% of protein in the soluble fraction; a score of ++ corresponds to 50% soluble expression; and a score of +++ corresponds to > 50% of protein in the soluble fraction.

4.1.2.6. Small-scale protein purification

The His_6-tagged recombinant proteins were automatically purified from cell-free extracts by immobilized metal-ion affinity chromatography (IMAC) as described elsewhere (Fernandes *et. al*., Chapter 3). Briefly, the crude cell lysates were incubated with sepharose chelating beads (200 µL with bound Ni^{2+}) and then transferred into 96-well filter plates (Macherey-Nagel). The wells were washed with a buffer A (50 mM NaHEPES, 1 M NaCl, 10 mM Imidazole, 5 mM $CaCl_2$ pH 7.5) followed by a second wash with buffer B (50 mM NaHEPES, 1 M NaCl, 35 mM Imidazole, 5 mM $CaCl_2$ pH 7.5) to elute contaminant proteins. The fusion recombinant proteins were eluted from the resin beads with 150 µL of elution buffer (50 mM NaHEPES, 1 M NaCl, 300 mM Imidazole, 5 mM $CaCl_2$, pH 7.5) into 96-deep-well plates. All protein purification steps were automated in a Tecan robot (Switzerland) containing a vacuum manifold. Protein homogeneity was evaluated through SDS-PAGE. Levels of purified protein were also scored (0, +, ++, and +++) by visual inspection of SDS-PAGE gels of purified proteins complemented with A_{280nm} measurements through a NanoVue (GE Healthcare). The score 0 corresponds to no purified protein; a score of + corresponds to > 0.15 and ≤ 0.5 mg of recombinant purified protein (e.g. Figure 4.5 – protein E, lane 1); a score of ++ corresponds to > 0.5 and < 1 mg of recombinant purified protein (e.g. Figure 4.5 – protein E, lane 2); and a score of +++ corresponds to ≥ 1 mg of recombinant purified protein (e.g. Figure 4.5 – protein E, lane 5).

4.1.3. Results and Discussion

4.1.3.1. Proteins selected for these studies

This study aims to develop a novel series of prokaryotic expression vectors that will allow testing the capacity of different fusion tags to promote expression of soluble recombinant proteins at high levels. The proteins were selected on the basis of previous studies that indicated them to be highly prone to form IBs when expressed in *E. coli*. Two of these proteins (A and B) are from the thermophilic bacterium *C. thermocellum* and showed very low soluble expression in *E. coli* in previous attempts to produce them in the soluble form. In the case of protein A, the formerly strategy was based on the fusion of the respective gene

with a dockerin module into pET21 vector (Novagen), and co-expression of the construct in the same cell with a cohesin. This resulted in some soluble production (unpublished data); however the low expression of the cohesin and dockerin peptides into *E. coli* could explain the poor results. Regarding protein B, previous Trx fusion conducted to very low soluble expression (data not shown). Protein C was selected due to the 50:50 expression of insoluble and soluble protein observed in *E. coli* when cloned into pHTP vector (His_6 tag) and expressed at 37 ºC. To serve as control of expression and purification approaches, a protein with high soluble expression in this host when expressed in pHTP at 37 ºC was included in the study (green fluorescent protein, here named protein G). The selected target proteins (see Table 4.3) differ in function, molecular weight (that varies from 10.62 to 68.35 kDa) and biochemical properties. All of them show a hydrophilic nature, with protein A presenting the highest value of GRAVY (-0.300) and protein C the lowest one (-0.715). The p*I* of the protein ranges from around 4 to 8.

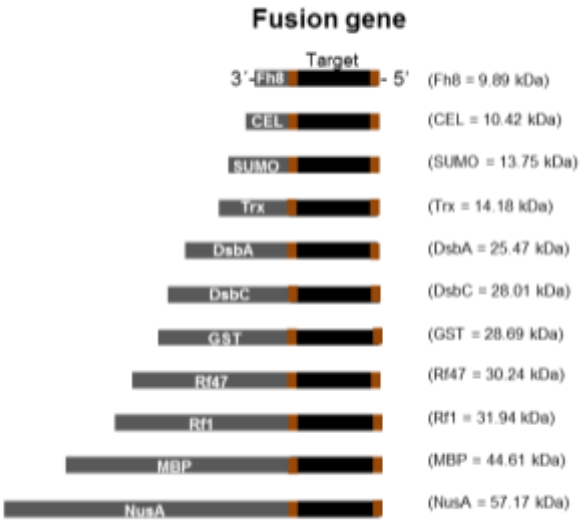
4.1.3.2. Generation of novel prokaryotic expression vectors derivatives of pHTP

Novel vectors for recombinant protein expression in *E. coli* were constructed by inserting different fusion tags into the pHTP backbone (Fernandes *et. al*., Chapter 3) such that the fusion tag would be at the N-terminus of the recombinant protein (see Table 4.2). Eight well-studied fusion partners commonly used to enhance solubility of recombinant proteins in *E. coli* were selected (Trx, GST, MBP, NusA, SUMO, DsbA, DsbC, and Fh8) besides the affinity tag His_6. In addition, three novel system tags were designed (Rf1, Rf47 and CEL) and their capacity to promote recombinant proteins to acquire proper fold and thus circumvent the formation of IB was compared with the traditional tags. The fusion tags Rf1 and Rf47 are two recombinant proteins from the ruminal bacterium *Ruminococcus flavefaciens* previously known to be produced at high yields and in the soluble form in *E. coli* at a range of temperatures and culture media. The CEL fusion tag was designed based on the high-affinity cohesin-dockerin interactions that define the cellulosome machine involved in cellulose degradation. It is well established that in the cellulosome of the anaerobic bacterium *Clostridium thermocellum*, the cohesin domains of the scaffoldin CipA are unable to discriminate between the individual dockerins present in the catalytic subunits (Yaron *et al*., 1995; Lytle *et al*., 1996). The integration of enzymes in the cellulosomal complex through these interactions allows them to display a distended fold which benefits the catalytic mechanism. Based on this premise, we developed a strategy to allow recombinant proteins expressed in *E. coli* to be integrated into a mini-cellulosome in order to acquire the correct spacing and thus preclude protein aggregation. The mini-cellulosome consisting of a cohesin cassette followed by a dockerin module, described in the Materials and Methods section, was cloned into pHTP plasmid so that recombinant proteins will be fused with the dockerin at their N-terminus. The CEL fusion tag has the second lowest molecular weight (10.42 kDa) of

the tags used in this study, excluding the His_6 tag (Fh8 is the smallest one – 9.89 kDa, after His_6). Rf1 (31.94 kDa) and Rf47 (30.24 kDa) have similar molecular weights, which can be comparable to GST (28.07 kDa) and DsbC (28.01 kDa). Regarding to isoelectric points (p*I*), the CEL fusion tag shows the second highest value after His_6. In opposite, the Rf1 tag has a low p*I*, which is similar to NusA. All tags have a hydrophilic tendency predicted from their amino acid sequence, showing negative values for the Grand average of hydropathicity (GRAVY) (Kyte *et al*., 1982). Rf1 and Rf47 have a similar hydrophilic nature, proximal to the CEL, which can be comparable with GST. His_6, SUMO and Fh8 have higher GRAVY values. All vectors have a cloning region containing specific handles that allow cloning and sub-cloning through a base complementation strategy, as reported elsewhere (Fernandes *et. al*., Chapter 3). In Figure 4.1 a schematic representation of all pHTP-derivative vectors is displayed.

Figure 4.1| The pHTP expression vector series.

Parts of this figure are confidential



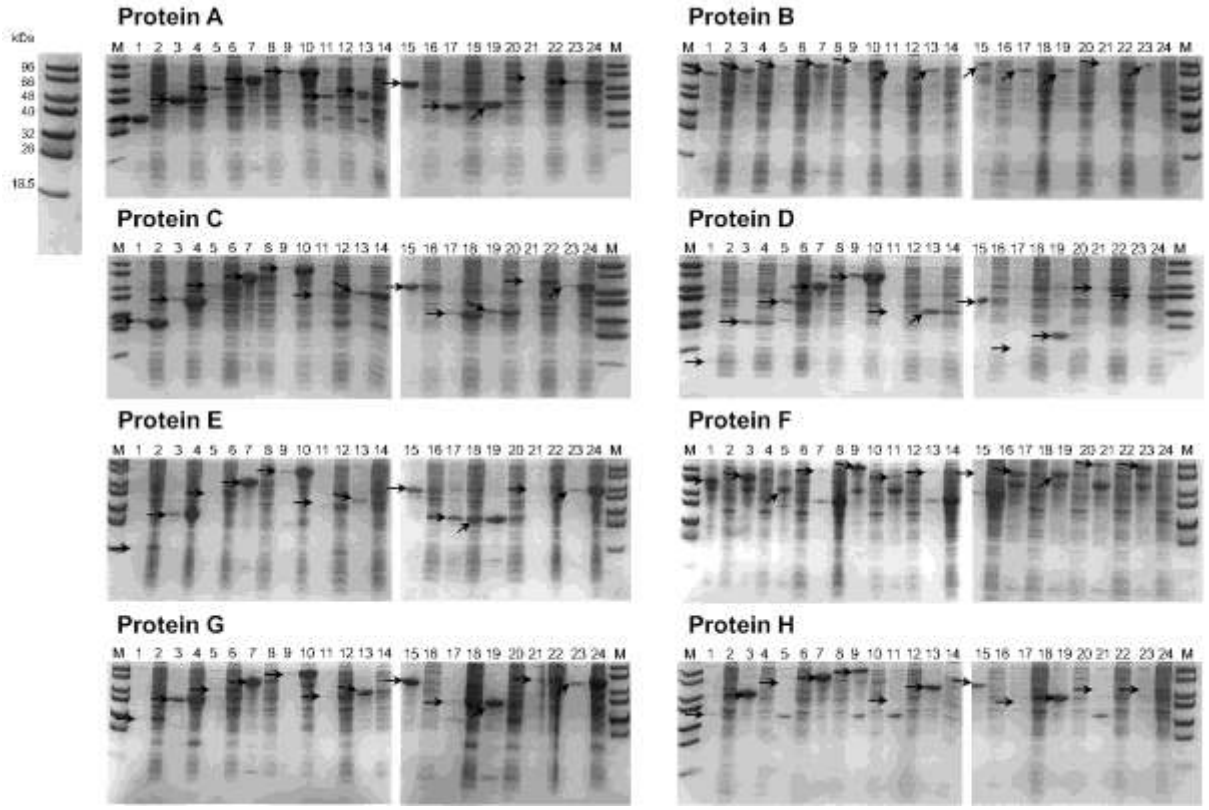
4.1.3.3. Cloning into the pHTP-derivative vectors

The 8 genes were cloned directly into the 12 expression plasmids by a ligation-independent cloning method based on the complementation of single-stranded DNA fragments both generated on the vector and insert. In general, the cloning reactions were highly efficient and only for a low percentage of cases more than two colonies were screened to identify a positive recombinant clone. In general, the sequences of all positive clones accumulated no mutations although for two cases a second gene needed to be sequenced. The experiment generated 96 recombinant plasmids that were used for the expression assays described below.

4.1.3.4. Protein solubility analysis from the small-scale screening

The efficacy of 12 fusion tags (including the His_6 tag) to improve the levels of expression and solubility of 8 different proteins in *E. coli* BL21(DE3) was tested. The 96 strains encoding the different protein versus tag combinations were grown on five different culture conditions: (1) cells grown in auto-induction medium, gene expression induced with IPTG and cells grown at 37 ºC for 16 hours after induction; (2) cells grown in auto-induction medium, gene expression not induced with IPTG and cells grown at 37 ºC for 20 hours; (3) cells grown in auto-induction medium, gene expression induced with IPTG and cells grown at 16 ºC for 16 hours after induction; (4) cells grown in LB medium, gene expression induced with IPTG and cells grown at 37 ºC for 16 hours after induction; and (5) cells grown in LB medium, gene expression induced with IPTG and cells grown at 16 ºC for 16 hours after induction. Cells were harvested and lysed and the presence of soluble and insoluble recombinant proteins was evaluated by SDS-PAGE. Levels of soluble protein were scored (0, +, ++ and +++) according to the scale defined in Materials and Methods section and with example representative gels displayed in Figure 4.2.

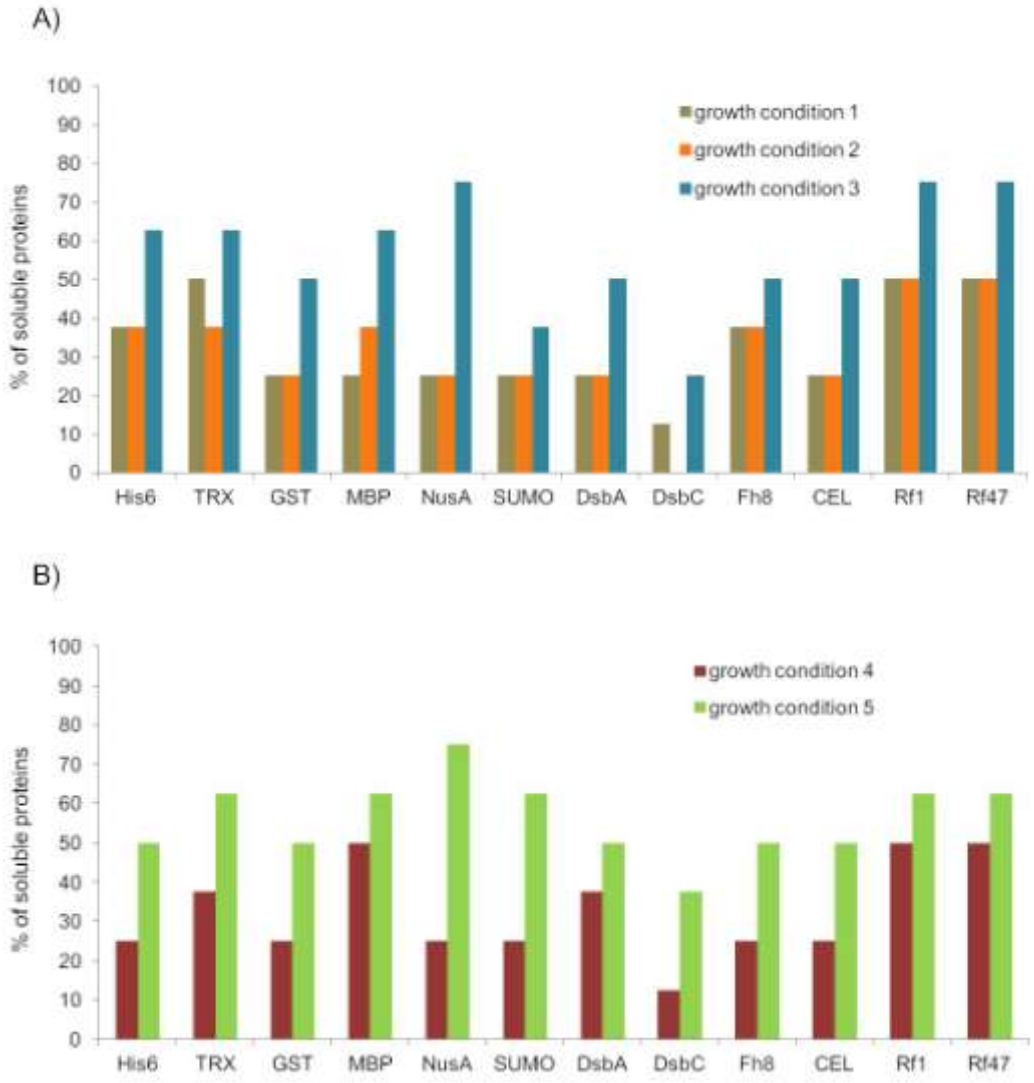
Figure 4.2| SDS-PAGE of His_6-proteins A, B, C, D, E, F, G, and H only fused with the affinity tag (lanes 1 and 2) and fused with Trx, GST, MBP, NusA, SUMO, DsbA, DsbC, Fh8, CEL, Rf1 and Rf47 tags obtained from BL21(DE3) *E. coli* recombinant cultures grown in culture condition 5.



E. coli BL21(DE3) recombinant cultures grown in LB medium at 37 ºC to an OD_{600nm} of 0.4-0.6 and gene expression induced with 1 mM IPTG at 16 ºC overnight (culture condition 5). Both pellet and supernatant fractions from the total cell lysates were separated through SDS-PAGE according to the following order: (1-2) His_6; (3-4) Trx; (5-6) GST; (7-8) MBP; (9-10) NusA; (11-12) SUMO; (13-14) DsbA; (15-16) DsbC; (17-18) Fh8; (19-20) CEL; (21-22) Rf1; (23-24) Rf47. Arrows indicate expected/observed molecular weights for respective fusion proteins according to Table S4.1 (see in Annex). Lane M: low molecular weight protein marker (NZYTech, genes & enzymes, Portugal). In supplementary material (Table S4.2, in Annex), the molecular weight ratios between the target proteins alone and respective fusion proteins are presented.

Initially, we compared the capacity of each tag to drive the production of the 8 recombinant proteins in a soluble form in *E. coli* grown in auto-induction media. The data, presented in Figure 4.3A, suggest that IPTG induction had no effect in the percentage of soluble proteins when all tags were considered: the percentages of soluble proteins obtained when cells were induced with or without IPTG were 32% and 31%, respectively. For Trx and DsbC tags, the addition of IPTG improved the percentage of soluble fusion proteins while for MBP it rather decreased it (by comparing growth conditions 1 and 2). In addition, no DsbC fusion soluble proteins were observed when protein expression was spontaneously induced, suggesting that IPTG induction is required when this tag is used. For all the remaining fusion proteins, addition of IPTG had no effect in protein solubility (Figure 4.3A).

Figure 4.3| Percentage of proteins that were expressed in the soluble form for all the 12 tags tested in *E. coli* cells grown in auto-induction medium (A) and LB-medium (B).



A) The 96 strains encoding the different proteins versus tag combinations were grown in auto-induction media under the different following conditions: (1) gene expression induced with IPTG and cells grown at 37 ºC for 16 hours after induction; (2) gene expression not induced with IPTG and cells grown at 37 ºC for 20 hours; (3) gene expression induced with IPTG and cells grown at 16 ºC for 16 hours after induction. (**B**) The 96 strains encoding the different proteins versus tag combinations were grown in LB media under the different following conditions: (4) gene expression induced with IPTG and cells grown at 37 ºC for 16 hours after induction; and (5) gene expression induced with IPTG and cells grown at 16 ºC for 16 hours after induction.

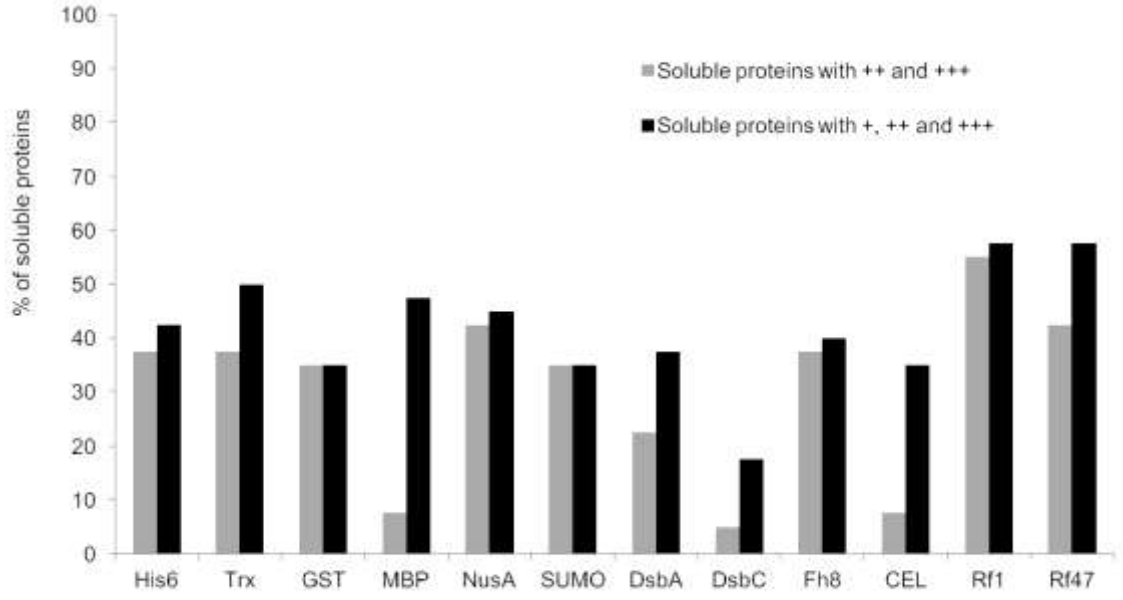
The decrease of the induction temperature from 37 ºC to 16 ºC significantly improved the percentage of soluble proteins produced from *E. coli* grown in auto-induction medium for all tags in test (Figure 4.3A). However, this effect was more pronounced for the NusA tag. The 96 *E. coli* strains were also grown in standard LB media at two different induction temperatures. Again, the data, presented in Figure 4.3B, confirms that a reduction in induction temperature has a dramatic effect in the percentage of soluble proteins, with higher levels of soluble proteins observed at lower temperatures and for all tags analyzed. Taken

together the results considering the growth of 96 *E. coli* strains upon 5 different cultivation conditions suggest that, in both culture media, lower induction temperatures promote higher protein solubility. Indeed, a decrease in the temperature during recombinant gene expression in *E. coli* reduces the rate of protein synthesis, providing additional time for proteins to acquire the correct fold, as well as increasing protein stability while attenuating protein degradation due the lower activity of heat shock proteases (Chesshyre & Hipkiss, 1989; Spiess, Beil, & Ehrmann, 1999; Hunke & Betton, 2003; Sørensen & Mortensen, 2005). In addition, NusA tag was clearly more effective when cells were grown at 16 ºC (growth conditions 3 and 5), suggesting that this tag requires low temperatures to promote protein solubility. When the two culture media are compared, LB versus auto-induction media (Figure 4.3A and 4.3B), the overall percentage of soluble proteins when all tags and proteins are considered was identical both with induction at 37 ºC (32% of soluble proteins, for grown conditions 1 and 4) or at 16 ºC (56% of soluble proteins, for grown conditions 3 and 5). Thus, these data suggest that the type of media has no effect in the capacity of *E. coli* to produce soluble recombinant proteins.

When all five culture conditions were considered (Figure 4.4), the percentage of soluble proteins expressed (scores +, ++ and +++) with Rf1 and Rf47 tags was identical and higher than the percentage obtained with any other tags. However, this derives from the good performance that Rf1 and Rf47 tags presented at 37 ºC (growth conditions 1, 2 and 4) since at 16 ºC (growth conditions 3 and 5) these tags led to similar percentages of soluble protein when compared with the NusA (in auto-induction media) and Trx, MBP and SUMO (in LB media) tags (Figure 4.3). Data presented in Figure 4.4 suggest that levels of soluble proteins generated by CEL, DsbC, DsbA and MBP tags are mostly below 50% of total protein, which contrasts with the other tags, which showed a high percentage of proteins scored with ++ and +++. Thus, these tags have the capacity to generate soluble protein but with lower efficiencies as a significant percentage of IB are still formed. Overall, the less efficient tag was DsbC. The non-eukaryotic origin of the heterologous target proteins used in this study (except protein F) which may not require disulfide isomerization for correct folding could explain the poor solubility observed for DsbC fusion proteins. In addition, a decrease of solubility was observed resulting from the incorporation of some fusion tags. Thus, percentages of soluble proteins were lower for some affinity tags when compared with the His_6-tag, which is solely an affinity purification tag. For example, the levels of soluble expression of His_6 tag protein G (used as control in this study) decreased from score +++ to ++, + or even 0 when using MBP, DsbA, DsbC or CEL as tags (see Figure 4.2 – Protein G). These observations were also previously reported by Bird (2011). Since all target genes are under the control of the same T7 promoter, differences in the soluble expression may be probably due to the properties of the fusion tag (Costa *et al.*, 2013). However, it remains

elusive why the fusion technology may lead to lower levels of soluble protein obtained under some conditions.

Figure 4.4| Percentage of proteins that were expressed in the soluble form (scores +, ++ and +++) or where the level of soluble protein is higher than the level of protein in the form of IB (scores ++ and +++) considering all 5 growing conditions and for all the tags.



The percentage of soluble fusion proteins for each tag was estimated by the number of proteins present in supernatant fractions (scores +, ++ and +++) per total number of proteins tested in the five culture assays. A score of + corresponds to < 50% of protein in the soluble fraction; a score of ++ corresponds to 50% soluble expression; and a score of +++ corresponds to > 50% of protein in the soluble fraction.

Taken together the data presented above suggest that the fusion tags may be ranked in the following order considering their capacity to generate soluble proteins (the sum of +, ++ and +++ scores): Rf1 > Rf47 > Trx > MBP > NusA > His_6 > Fh8 > DsbA > SUMO = GST > CEL > Dsbc. However, as discussed above, some tags, more evident with the MBP tag, although well positioned in the solubility rank generated a lower percentage of soluble protein in relation to the total proteins (Figure 4.4). Indeed, from the 19 soluble MBP fusion protein derivatives, only 3 had the score ++, which corresponds to a 50:50 protein content in the soluble and insoluble fractions (proteins C, D, and E, under the culture condition 4) and no +++ score was attributed. The N-terminal position of this fusion tag could explain this observation since C-terminal MBP fusions were suggested to be more effective (Dyson *et al*., 2004). Interestingly, the affinity His_6 tag showed a good rank position in particular at the lower temperatures of gene induction. The low capacity of GST to enhance protein solubilization described here was also reported in several other comparison studies (Hammarström *et al*., 2002; Dyson *et al*., 2004; Marblestone *et al*., 2006; Bird, 2011; Costa *et*

al., 2013). The solubility of SUMO fusions was occasionally difficult to assess, as observed by the small content of proteins, and thus lower levels of expression, in lanes 11 and 12 of Figure 4.2. As previously reported, proteins fused with SUMO exhibited a molecular weight of about 3 kDa higher than expected after gel migration (Marblestone *et al*., 2006 ; Costa *et al*., 2013). Since high scores for soluble expression in small-scale screenings are consistent predictors of the soluble protein yields in scale-up production (Bird, 2011), Rf1, as well as Rf47 showed good indications for soluble expression enhancement. Nevertheless, no fusion/culture condition combination resulted in the soluble expression of protein F, the largest target protein in study – 68.35 kDa (e.g. for growth condition 5 data is shown in Figures 4.2 and 4.5). For protein B (the second largest protein in study – 61.98 kDa), only NusA (at 16 ºC; e.g. for growth condition 5, data is shown in Figure 4.2, lanes 9 and 10 and Figure 4.5, lane 5) or Rf1 (under culture condition 3) fusions have resulted in soluble proteins. The third protein of higher molecular weight (protein H, 36.27 kDa) was only soluble when produced by fusion with Rf47 or His_6 under the culture condition 3, although its expression was difficult to access in some conditions (data not shown). The correlation between successful soluble expression and a decrease in protein molecular weight had been extensively reported (Dyson *et al*., 2004).

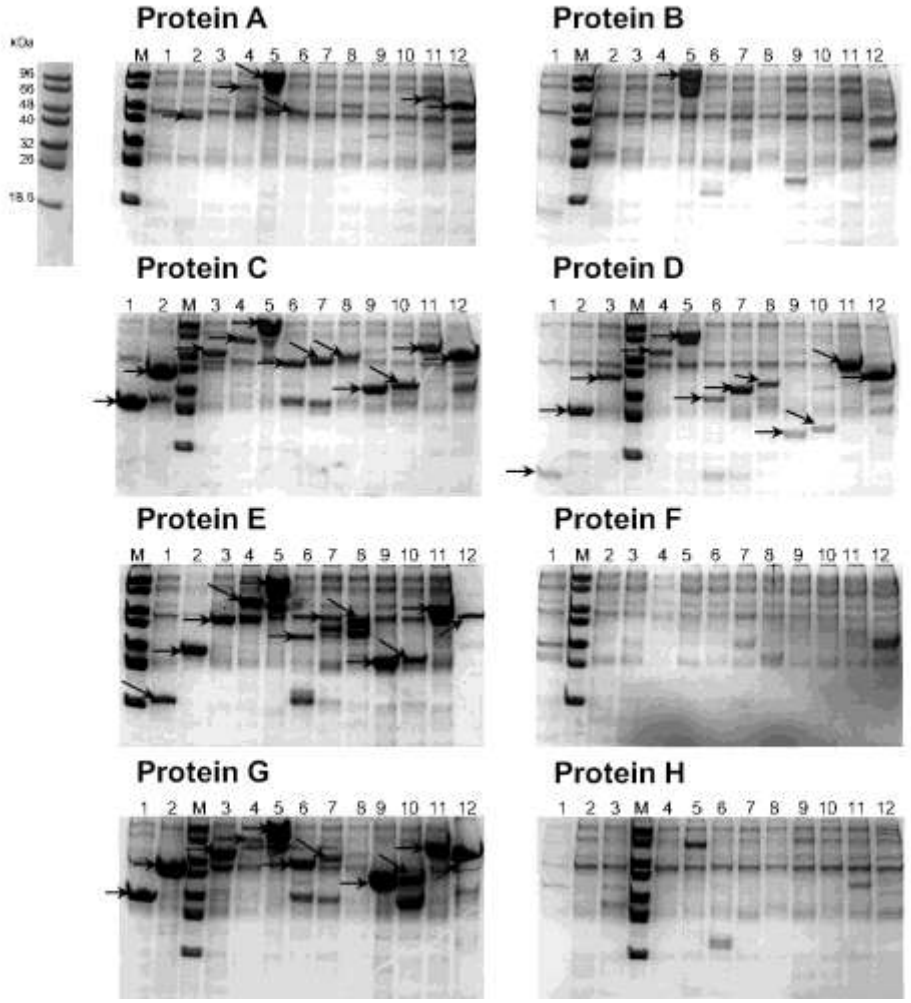
4.1.3.5. Yield and purification efficiency

Although some fusion partners used in this study constitute itself affinity tags (e.g. MBP and GST), the purification approach selected for all recombinant proteins was based on the N-terminal (pHTP) or internal (other vectors) His_6 tag present just before the pHTP cloning site. Thus, all proteins were purified in an automated protocol through IMAC. Levels of purified protein obtained were scored (0, +, ++, and +++) as described in Materials and Methods section and data obtained for all of the five culture conditions are summarized in Figure 4.6 with representative SDS-PAGE gels displayed in Figure 4.5.

In general, the data obtained for the purified recombinant fusion proteins translated the results reported above for solubility, with Rf1 and Rf47 showing the highest protein yields. Only SUMO recombinant fusion proteins were purified in lower levels than those expected by comparing the percentage of soluble proteins scored with +, ++, and +++ (Figure 4.4) with the levels of purified proteins scored with > 0.5 mg (Figure 4.6). The difficulty found to assess the solubility of SUMO fusion proteins, as reported above, could explain this observation. Rf1 presented the highest percentage of recombinant proteins purified in high yields from 0.5 mg to over 1 mg, just followed by NusA and Rf47 tags. Regarding the CEL tag, the percentage of recombinant proteins purified in high yields > 0.5 mg was similar to that for SUMO, DsbA, CEL, or even MBP, suggesting that this tag has a low capacity to enhance protein expression. DsbC fusions showed the lowest yields after purification, as a result of the pronounced tendency for IB formation, as observed above. Attending to the number of fusion

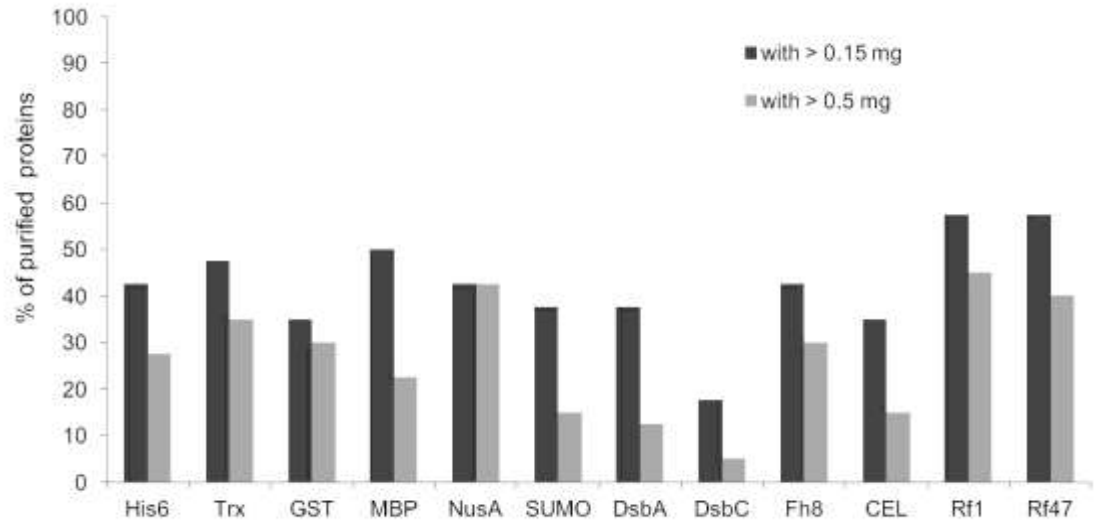
proteins that were purified with the highest yields (> 0.5 mg), the fusion tags were ranked as: Rf1 > NusA > Rf47 > Trx > Fh8 = GST > His_6 > MBP > SUMO = CEL > DsbA > DsbC. Despite GST fusions have shown relatively high yields after purification (30% of proteins purified with > 0.5 mg), the GST tag performed poorly to enhance the number of soluble proteins, as described above. In opposite, MBP highly improved protein solubility, but MBP fusion proteins were poorly produced in high yields (22.5% of proteins purified with > 0.5 mg).

Figure 4.5| SDS-PAGE of His_6-tagged proteins A, B, C, D, E, F, G, and H only fused with the affinity tag (lane 1) and fused with Trx, GST, MBP, NusA, SUMO, DsbA, DsbC, Fh8, CEL, Rf1 and Rf47 system tags following purification through IMAC. Proteins were obtained from *E. coli* BL21(DE3) recombinant cultures grown in the culture condition 5.



E. coli BL21(DE3) recombinant cultures grown in LB medium at 37 ºC to an OD_{600nm} of 0.4-0.6 and protein overexpression induced with 1 mM IPTG at 16 ºC overnight (culture condition 5). Lanes in SDS-PAGE correspond to the following tag order: (1) His_6; (2) Trx; (3) GST; (4) MBP; (5) NusA; (6) SUMO; (7) DsbA; (8) DsbC; (9) Fh8; (10) CEL; (11) Rf1; (12) Rf47. Arrows indicate observed molecular weights for respective fusion proteins according to Table S4.1 (see in Annex). Lane M: low molecular weight protein marker (NZYTech, genes & enzymes, Portugal). In supplementary material (Table S4.2, in Annex), the molecular weight ratios between the target proteins alone and respective fusion proteins are presented.

Figure 4.6| Percentage of purified proteins with a final yield > 0.15 mg and > 0.5 mg. Purified proteins A, B, C, D, E, F, G, and H fused with Trx, GST, MBP, NusA, SUMO, DsbA, DsbC, Fh8, CEL, Rf1 and Rf47 system tags produced in *E. coli* BL21(DE3) under the all five culture conditions.



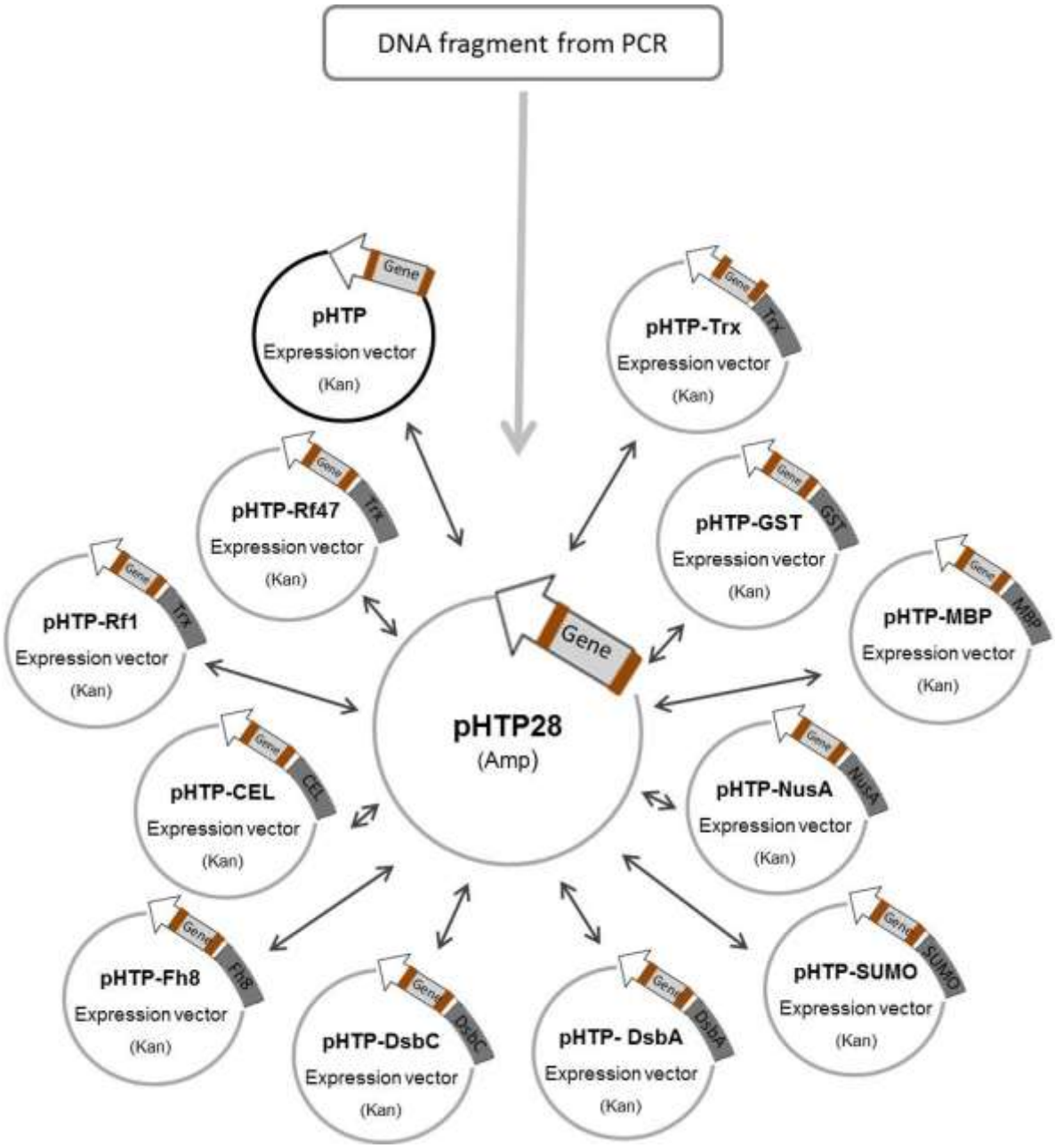
The recombinant fusion proteins were purified through IMAC. The percentage of purified fusion proteins for each tag with a final yield > 0.15 mg (scores +, ++, and +++) or > 0.5 mg (scores ++ and +++) was tested for all five culture conditions.

Interestingly, spontaneous fusion cleavage was observed for some of the largest fusion partners after protein purification, mainly for MBP (44.61 kDa), NusA (57.17 kDa), DsbA (25.47 kDa), Rf1 (31.94 kDa), and Rf47 (30.24 kDa) tags (e.g. Figure 4.5 – proteins C, E, and G, lane 4; protein H, lane 5; proteins C and G, lane 7, or proteins A, B C, and F, lane 12). The cleavage may have occurred in the linker region between the fusion tag and the His_6 sequence (SSGPQQGLR). We observed more fusion cleavages in proteins that were expressed at 37 ºC, at which temperature bacterial proteases are highly produced. In addition, protein conformational rearrangements upon His_6 binding to the Ni^{2+}-column during purification, probably promotes the exposure of the linker regions to protease attack. Besides fusion cleavage, the target proteins were correctly purified when solubly expressed. Regarding the CEL system, it is well known that dockerin-cohesin non-covalent interactions are disrupted by SDS denaturation, so observation of cohesins on the SDS gels was expected. However, cohesin expression was not always verified. For instance, no cohesin expression was observed upon induction at 16 ºC, suggesting poor expression at low temperatures. These observations question the real integration of the insoluble recombinant protein into the mini-cellulosome and suggest that dockerin fusion alone was able to enhance protein solubility. As observed for other fusion partners, dockerin cleavage seems to occur. Research to investigate the most appropriate methods for tag removal is presently ongoing.

4.1.3.6. Construction of simple sub-cloning system applied to the pHTP vector series

Taken together, the results presented here and those reported by other comparative studies (Hammarström *et al*., 2002; Braun *et al*., 2002; Shih *et al*., 2002; Dyson *et al*., 2004; De Marco *et al*., 2004; Marblestone *et al*., 2006; Bird, 2011; Costa *et al*., 2013), suggest that there is no universal tag that appears to work for all different types of proteins but rather the efficiency of the tag depends on the recombinant protein. Thus, small-scale screening assays using a broad repertoire of different tags may constitute the best alternative for selecting the most appropriate solubility tag for each protein type. The LIC-based method in which the pHTP vectors rely allows a simple and rapid sub-cloning system that could be applied for these small-scale screenings. Thus, here we designed a novel method that allows easily to exchange genes from an entry vector to various destination/expression plasmids. Thus, as described in Materials and Methods, we constructed the pHTP28 prokaryotic cloning vector based on the pNZY28 plasmid (NZYTech, genes & enzymes, Portugal). This vector is ampicillin resistant, in contrast with the pHTP series of protein expression vectors that are kanamycin resistant. Thus, the system would work by initially cloning the gene of interest from a PCR product directly into pHTP28 and sequencing the nucleic acid to confirm that no mutations accumulated during the amplification. Subsequently, the gene of interest could be transferred to all the pHTP expression vectors in a single step, generating the expression plasmids. The gene would not need to be sequenced as the transfer will not involve any polymerase chain reaction. Thus, sub-cloning of a target gene can be easily achieved from the pHTP28 entry clone to any of the pHTP derivatives that include different solubility tags. The expression clones generated by this system could be directly used in the solubility screening assays to detect the most efficient tag for the different proteins (Figure 4.7).

Figure 4.7| Sub-cloning system based on the pHTP vector series.



The system allows cloning of a target gene into the pHTP28 vector through a ligation-independent cloning method based on DNA base complementation. Once gene is cloned into the entry cloning vector, the DNA fragment can be transferred into one or more expression vectors simultaneously (pHTP and/or its derivatives that include different solubility tags) using the same cloning approach.

4.1.4. Conclusions

Here we have compared the efficacy of several fusion tags to improve the expression and solubility of 8 different recombinant proteins in *E. coli*. In addition, high-throughput approaches were developed for the rapid generation of expression clones and screening of the optimal conditions for successful protein expression. Overall the data suggest that Rf1

and Rf47 proteins are very promising candidates to increase the number of available solubility tags and thus should contribute to promote an expansion of the protein fusion technology. In contrast, the solubilization strategy based on the cellulosome concept (CEL system tag) failed to enhance protein solubility by providing correct spacing of the recombinant proteins produced in *E. coli*. However, the data suggested that the dockerin module alone effectively contributed for protein solubility. Data presented here confirm that the combination of a fusion partner with the reduction of temperature during protein overexpression potentiate the solubilization of recombinant proteins. Nevertheless, the data suggested that although some tags may be more efficient to enhance protein solubility, there is no universal tag that could work in all situations. Thus, we have developed here a novel technology that will allow the initial cloning of the desired gene into an entry vector (pHTP28) and its subsequent transfer to a series of destination vectors that could be tested for efficacy in generating soluble protein. This approach will very effectively contribute to identify the most favorable solubility tag to use for different protein types.

5. BIOCHEMICAL CHARACTERIZATION OF ONE OF THE MAJOR CELLULOSOMAL ENZYMES OF *Clostridium thermocellum*

5.1. A Mannan binding family 32 carbohydrate binding module influences the activity of the appended mannanase

Vânia O. Fernandes[3,4*], Kimiya Mizutani[1*], Shuichi Karita[2], Ana S. Luís[3], Makiko Sakka[1], Tetsuya Kimura[1], Adam Jackson[5], Xiaoyang Zhang[5], Carlos M.G.A Fontes[3,4], Harry J. Gilbert[5], Kazuo Sakka[1]

[1] Applied Microbiology Laboratory, Graduate School of Bioresources, Mie University, 1577 Kurimamachiya-cho, Tsu 514-8507, Japan; [2] Graduate School of Regional Innovation Studies. Mie University, 1577 Kurimamachiya-cho, Tsu 514-8507, Japan; [3] CIISA-Faculdade de Medicina Veterinária, Universidade de Lisboa, Av. da Universidade Técnica, 1300-477 Lisboa, Portugal; [4] NZYTech Lda, Estrada do Paço do Lumiar, Campus do Lumiar, Edifício E-R/C, 1649-038 Lisboa, Portugal; [5] Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, United Kingdom

*Equal contribution

Adapted from: Fernandes, *et al*. (2012). *Applied and Environmental Microbiology*. 78(14):4781-4787

Abstract

In general, cellulases and hemicellulases are modular enzymes in which the catalytic domain is appended to one or more non-catalytic Carbohydrate-Binding Modules (CBMs). CBMs, by concentrating the parental enzyme at their target polysaccharide, increase the capacity of the catalytic module to bind substrate leading to a potentiation in catalysis. *Clostridium thermocellum* hypothetical protein Cthe_0821, defined here as *Ct*Man5A, is a modular protein comprising an N-terminal signal peptide, a family-5 Glycoside Hydrolase (GH5) catalytic module, a family 32 CBM (CBM32) and a C-terminal type I dockerin module. Recent proteomic studies revealed that Cthe_0821 is one of the major cellulosomal enzymes when *C. thermocellum* is cultured on cellulose. Here we show that the GH5 catalytic module of Cthe_0821 displays endo-mannanase activity. *Ct*Man5A hydrolyses soluble konjac glucomannan, soluble carob galactomannan and insoluble ivory nut mannan, but does not attack the highly galactosylated mannan from guar gum, suggesting that the enzyme prefers unsubstituted β-1,4-mannoside linkages. The CBM32 of *Ct*Man5A displays a preference for the non-reducing end of mannooligosaccharides, although the protein module exhibits measurable affinity for the termini of β-1,4-linked glucooligosaccharides such as cellobiose. CBM32 potentiates the activity of *Ct*Man5A against insoluble mannans but has no significant effect on the capacity of the enzyme to hydrolyze soluble galactomannans and glucomannans. The product profile of *Ct*Man5A is affected by the presence of CBM32.

5.1.1. Introduction

Plant cell wall polysaccharides are the most abundant source of energy and carbon on Earth. Although biofuel production from lignocellulolsic biomass could contribute to a reduction in global warming, the enzymatic degradation of these recalcitrant structures is limited by their chemical and physical complexity (Liu, Saha, & Slininger, 2008). The complexity of microbial plant cell wall degrading systems has become apparent with the sequencing of a variety of genomes of cellulolytic microorganisms. Glycoside hydrolases (GHs) that deconstruct structural polysaccharides are modular enzymes consisting of catalytic module(s) linked, through a diversity of linker regions, to functionally independent non-catalytic Carbohydrate-Binding Modules (CBMs) and/or other ancillary module(s). GHs and CBMs are classified into more than 130 and 70 sequence-based families, respectively (CAZy database, http://www.cazy.org/) (Cantarel *et al*., 2009). Within a GH family there is conservation in the structural fold, catalytic apparatus and mechanism, however, substrate specificity may be highly divergent. For example, GH family 5 (GH5) contains a range of different enzymes exemplified by cellulases, xylanases, chitosanases and mannanases. Therefore, the function of a hypothetical protein within a CAZy family cannot always be predicted from its primary or tertiary structure.

Clostridium thermocellum is one of the most powerful degraders of cellulose. It produces a high molecular mass cellulolytic complex, termed the cellulosome (Bayer *et al*., 2008; Fontes & Gilbert, 2010), and is recognized as an important bacterium in the consolidated bioprocessing of cellulosic biomass to ethanol (Lynd *et al*., 2008). The genome sequence of *C. thermocellum* ATCC27405 was reported (accession number CP000568) recently and proteome analysis was carried out for two different strains, ATCC27405 (Gold & Martin, 2007) and F7 (Zverlov *et al*., 2005). Some common proteins were identified as highly abundant in both strains; e.g., CipA, CelS, CelK, XynC, XynZ, CelA, CelR, and CbhA, although some differences in the expression of cellulosomal proteins were evident in the two bacteria. Thus, CelE and CelJ and the uncharacterized enzyme Cthe_0821 were major components of the cellulosome in strain ATCC27405 but not in strain F7. Cthe_0821 is a modular enzyme comprising a signal peptide, a family 5 GH catalytic module, a family 32 CBM (CBM32) and a C-terminal type I dockerin module (Figure 5.1A). Enzymes containing a similar molecular architecture are found in *Acetivibrio cellulolyticus* (DDBJ/EMBL/GenBank accession No. EFL62306) and in some clostridia (Figure 5.1A), but the biochemical function of these glycoside hydrolases is unknown.

In this study, we showed that Cthe_0821 is an endo-β-1,4- mannanase and was thus defined as *Ct*Man5A. CBM32 binds to the non-reducing end of β-mannans and β-1,4-linked mannooligosaccharides, and plays a critical role in the hydrolysis of insoluble mannans. This paper provides evidence that specificity within CBM32, a family of proteins that generally

binds galactose- or GalNAc-based ligands, can be extended to include β-mannose containing polymers.

5.1.2. Materials and Methods

5.1.2.1. Bacterial strains, plasmids, and bacterial growth conditions

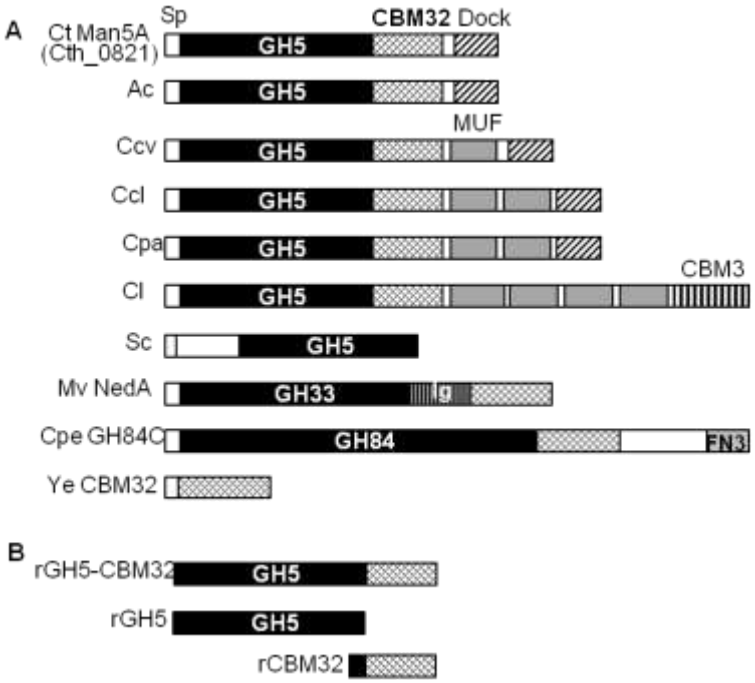
C. thermocellum ATCC27405 was used as a source of genomic DNA. Plasmid vector pET28a(+) (Novagen) was used for gene expression. *E. coli* BL21(DE3)-RIPL (Novagen) carrying a recombinant plasmid was cultivated in Super Broth (3.5% Bacto™ tryptone [BD Diagnostic, Sparks, MD], 2% Bacto™ yeast extract [BD Diagnostic], 0.5% NaCl pH 7.5) supplemented with chloramphenicol (34 μg/mL) and kanamycin (50 μg/mL) at 37°C for protein expression.

5.1.2.2. Expression and purification of rGH5-CBM32, rGH5, and rCBM32

The gene encoding rGH5-CBM32 was amplified by PCR from *C. thermocellum* ATCC27405 genomic DNA with KOD-Plus DNA polymerase (Toyobo, Osaka, Japan) and the PCR primer set, 5'-GGGGCCATGGATGACATTTATCCGGGACTTAGAG-3' and 5'-GGGGGTCGACTTCCGCAATTCCACCCTTTGG-3'; NcoI and SalI sites are underlined. The resulting PCR fragment was digested with *Nco*I or *Sal*I, ligated into similarly restricted pET-28a(+), yielding plasmid pET28a(+)-GH5-CBM32, encoding rGH5-CBM32, residues 27-475 of *Ct*Man5A (Figure 5.1B). Plasmids pET28a(+)-GH5 and pET28a(+)-CBM32 were constructed as described above, except that the following primer sets were used: 5'-GGGGCCATGGATGACATTTATCCGGGACTTAGAG-3' and 5'-GGGGGTCGACCTTGTTCTGCGCAATACTTC-3' for pET28a(+)-GH5, and 5'-GGGGCCATGGATATGACCACCGACGGAAC-3' and 5'-GGGGGTCGACTTCCGCAATTCCACCCTTTGG-3' for pET28a(+)-CBM32. GH5 and CBM32 contained *Ct*Man5A residues 27-349 and 249-475 (Figure 5.1B), respectively. All the recombinant proteins contained a 6×His-tag at C- terminus.

Cultures of recombinant *E. coli* clones (200 mL) were cultivated to mid-log growth phase (absorbance at 600 nm = 0.6) and isopropyl-β-D-thiogalactopyranoside was added to a final concentration of 1 mM. After an additional incubation of 4 h at 37°C, the cells were harvested, washed, and disrupted by sonication. Cell debris was removed by centrifugation. Purification of the recombinant proteins from the cell-free extracts was carried out with the aid of a HisTrap HP column (GE Healthcare Japan, Tokyo) according to the manufacturer's protocol. The purity of each fraction was analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970). Protein concentration was determined with bovine serum albumin (BSA) as the standard, using a Bio-Rad protein assay kit (Bio-Rad Laboratories, K.K., Tokyo, Japan).

Figure 5.1| Molecular architectures of *Clostridium thermocellum* Man5A, related proteins and *Ct*Man5A truncated derivatives.



Schematic of *C. thermocellum* Man5A and some related proteins **(A)** and the truncated derivatives **(B)**. Ct, *C. thermocellum*; Ac, *Acetivibrio cellulolyticus* hypothetical mannanase; Ccv, *C. cellulovorans* hypothetical mannanase; Ccl, *C. cellulolyticum* hypothetical mannanase; Cpa, *C. papyrosolvens* hypothetical mannanase; Cl, *C. lentocellum* hypothetical mannanase; Ss, *Sorangium cellulosum* hypothetical mannanase; Mv, *Micromonospora viridifaciens*; Cpe, *C. perfringens*; Ye, *Yersinia enterolitica*. SP, signal peptide; dock, dockerin module; MUF, module of unknown function; Ig, immunoglobulin-like module.

5.1.2.3. Enzyme assays

Mannanase activity was measured using konjac glucomannan (Megazyme, Wicklow, Ireland), ivory nut mannan (Megazyme), 1,4-β-D-mannan (prepared by controlled hydrolysis of carob galactomannan with β-mannanase and α-galactosidase, Megazyme), carob galactomannan (Megazyme), or guar gum (Sigma-Aldrich Japan, Tokyo) as substrates and incubating at 60°C in 50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer (pH 6.0). Incubation time was 5 min for konjac glucomannan or 5 to 20 min for the other substrates. Britton and Robinson's universal buffer (Britton & Robinson, 1931) was used to determine the optimum pH for enzyme activity. Reducing sugars released from the substrates were determined with the 3,5-dinitrosalicylic acid reagent, as described previously (Miller, 1959).

5.1.2.4. Analysis of hydrolysis products

Mannooligosaccharides (mannose through mannohexaose, 100 μg each; Megazyme) were incubated with 0.18 units (determined with konjac glucomannan) of the purified enzymes, rGH5-CBM32 or rGH5 in 10 μL of 50 mM MES buffer (pH 6.0) at 60°C. Ivory nut mannan (50

µg each; Megazyme) was incubated with 0.54 units (based on activity against konjac glucomannan) of rGH5-CBM32 or rGH5 in 10 µL of 50 mM MES buffer (pH 6.0) at 60°C. Thin layer chromatography (TLC) of the hydrolysis products was performed on a silicagel 60 plate (Merck, Darmstadt, Germany), using a solvent of water:acetic acid:acetone (1:1:2). Hydrolysis products were visualized by spraying the plate with an aniline-diphenylamine reagent (Gasparic & Churacek, 1978).

5.1.2.5. Qualitative polysaccharide binding assays

The binding of rCBM to ivory nut mannan and microcrystalline cellulose (Funacel; Funakoshi, Tokyo, Japan) was determined by mixing 0.1 ml of each protein (30 µg) and insoluble polysaccharides (10 mg of ivory nut mannan or Funacel) in 10 mM 2-[4-(2-hydroxyethyl)-1-piperazinyl] ethanesulfonic acid (HEPES) buffer (pH 7.4) and incubating the mixtures on ice for 3 h with occasional stirring. Following centrifugation, the supernatant fraction was recovered and the precipitate was resuspended in HEPES buffer. The suspension was again centrifuged to separate the wash and precipitate factions. All fractions were analyzed by SDS-PAGE. The affinity of rGH5 and rGH5-CBM32 for soluble mannans, konjac glucomannan, carob galactomannan, and guar gum was examined by native affinity PAGE as described previously (Arai *et al*., 2003). For native affinity PAGE, 1.5 µg of the protein was loaded onto gels. In all experiments BSA was used as the control protein.

5.1.2.6. Isothermal titration calorimetry (ITC)

The thermodynamic parameters of the binding of rCBM32 to mannooligosaccharides were determined by ITC using a VP-ITC calorimeter (MicroCal, Northampton, MA, USA). Briefly, titrations were performed at 25°C by injecting 2 - 10 µL aliquots of 5-20 mM ligand in 50 mM Na-HEPES buffer, pH 7.5, containing 5 mM $CaCl_2$, into the cell containing 100 µM CBM dialyzed into the Na-HEPES buffer, and the release of heat was recorded. The stoichiometry of binding (*n*), the association constant K_a, and the binding enthalpy ΔH were evaluated by using MicroCal Origin 7.0 software. The standard Gibbs energy change ΔG^0 and the standard entropy change ΔS^0 were calculated from $\Delta G^0 = -RT \ln K_a$ and $\Delta G^0 = \Delta H^0 - T\Delta S^0$, where *R* is the gas constant and *T* the absolute temperature.

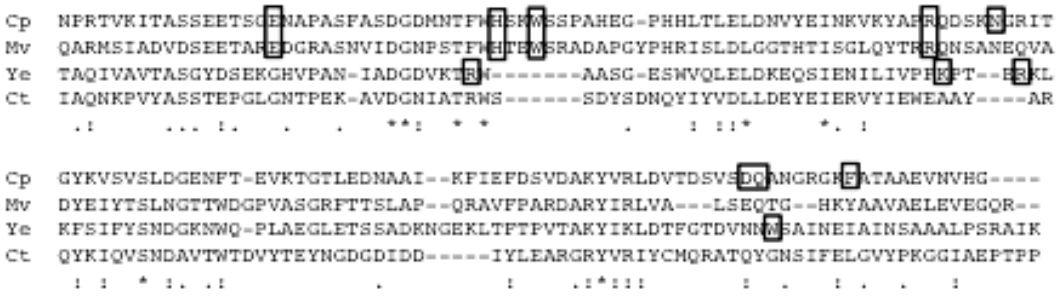
5.1.3. Results

5.1.3.1. Modular structure of *C. thermocellum* Man5A

Mature *Ct*Man5A consists of a GH5 catalytic module, a CBM32, and a type I dockerin module (Figure 5.1A). Homology search using BLAST (www.ncbi.nlm.nih.gov/BLAST) revealed that *Ct*Man5A displayed an identical modular organization with *Acetivibrio cellulolyticus* hypothetical protein EFL62306, which was annotated as a mannanase; sequence identity between the two proteins was 64% and 73% for the catalytic modules and

the CBM32s, respectively. Furthermore, other genes encoding enzymes highly similar to *Ct*Man5A were found in some clostridial genomes such as *C. cellulovorans* (ADL52309), *C. cellulolyticum* (ACL74764), *C. papyrosolvens* (EGD47465) and *C. lentocellum* (ADZ83685). Sequence identities between the catalytic module and CBM32 of *Ct*Man5A and the corresponding regions of the other proteins were 60-66% and 46-50%, respectively. Indeed, the molecular architecture of the proteins encoded by ADL52309, ACL74764, EGD47465 and ADZ83685 were more complex than *Ct*Man5A (Figure 5.1). Although *Sorangium cellulosum* is distantly related to *C. thermocellum*, a catalytic module of its hypothetical enzyme (CAN98252) showed relatively high sequence identity (52%) to that of *Ct*Man5A (Figure 5.1). The CBM-encoding sequence of *Ct*Man5A displays little sequence similarity to the three CBM32 members (out of the 1000 members of CBM32), characterized to date, which target Gal, GalNAc or GalA, Figure 5.2.

Figure 5.2| Alignment of CBM32 of *Ct*Man5A to CBM32 sequences characterized to date.



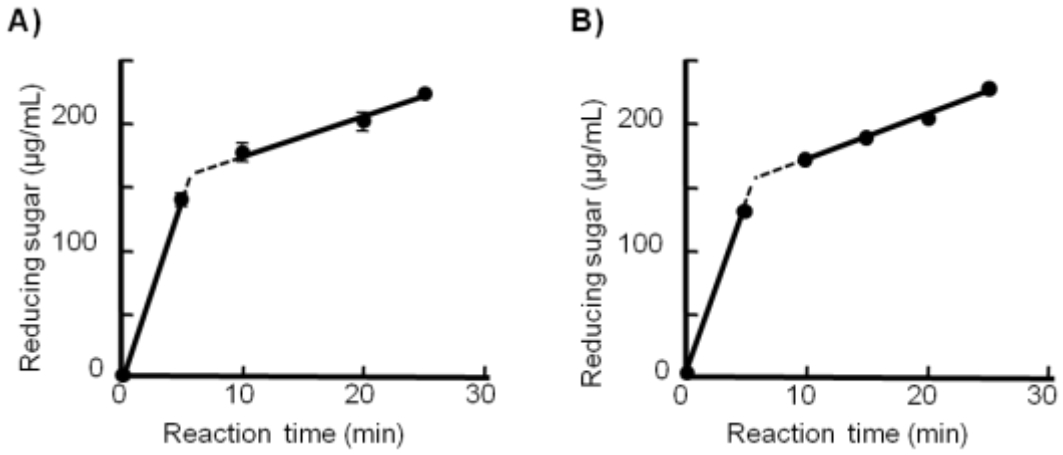
Residues proposed to interact with LacNAc, galactose or polygalacturonic acid are boxed. Cp, CBM32 of *C. perfringens* GH84C; Mv, CBM32 of *Micromonospora viridifaciens* NedA; Ye, CBM32 of *Yersinia enterolitica* polygalacturonic acid-binding protein; Ct, CBM32 of *C. thermocellum* Man5A. Asterisks (*), colons (:) and periods (.) show fully, highly and moderately conserved amino acid residues respectively.

5.1.3.2. Enzymatic properties of rGH5-CBM32 and rGH5

rGH5-CBM32 and rGH5 (Figure 5.1B) were expressed as fusion proteins containing a 6×His tag and purified by IMAC. The purified proteins were present as single bands on SDS-PAGE gels and the observed sizes were in agreement with the expected Mw values (data not shown). When insoluble ivory nut mannan was used as the substrate for rGH5-CBM32 and rGH5, biphasic hydrolysis patterns were observed (Figure 5.3), that is, rapid hydrolysis of the substrate by 5 min (first phase) and slower hydrolysis after 5 min (second phase), under the reaction conditions used. The enzymes may attack easily hydrolysable region (probably amorphous region) of the substrate in the first phase and tougher region (crystalline region) in the second phase. Similar phenomenon was observed for 1,4-β-mannan. Therefore, activity values estimated from both the first and second phases are shown for ivory nut mannan and 1,4-β-mannan in Table 5.1. Both rGH5-CBM32 and rGH5 were highly active toward konjac glucomannan and displayed moderate activity against 1,4-β-D-mannan, ivory

nut mannan and carob galactomannan (Table 5.1). The enzymes exhibited no activity against guar gum, microcrystalline cellulose, carboxymethylcellulose, xylan, barley **β**-glucan, or 4-nitrophenyl β-D-cellobioside. When mannanase activity was determined with konjac glucomannan as the substrate, the enzyme´s optimum temperature was of 60°C and the optimum pH was 6.0, consistent with the observation that the two enzymes retained full activity after incubation for one hour without substrate. The substrate specificity of *Ct*Man5A indicates that the enzyme hydrolyzes THE β-1,4-mannoside linkage and displays a preference for unsubstituted mannans, as guar gum, a highly galactosylated mannan, was not hydrolyzed. The substrate specificity of rGH5 was similar to that of rGH5-CBM32 against soluble substrates (Table 5.1). However, rGH5-CBM32 showed significantly higher activity than rGH5 towards ivory nut mannan and 1,4-β-D-mannan, which are both insoluble polysaccharides.

Figure 5.3| Hydrolysis pattern of ivory nut mannan by rGH5-CBM32 and rGH5.



Biphasic hydrolysis of insoluble ivory nut mannan by rGH5-CBM32 (**A**) and rGH5 (**B**). Ivory nut mannan was incubated with rGH5-CBM32 (A) and rGH5 (B) and the amounts of reducing sugars were determined with the 3,5-dinitrosalicylic acid reagent at the indicated times. Experiments were carried out in triplicate.

Table 5.1| Activities of rGH5-CBM32 and rGH5 toward mannans from different origins.

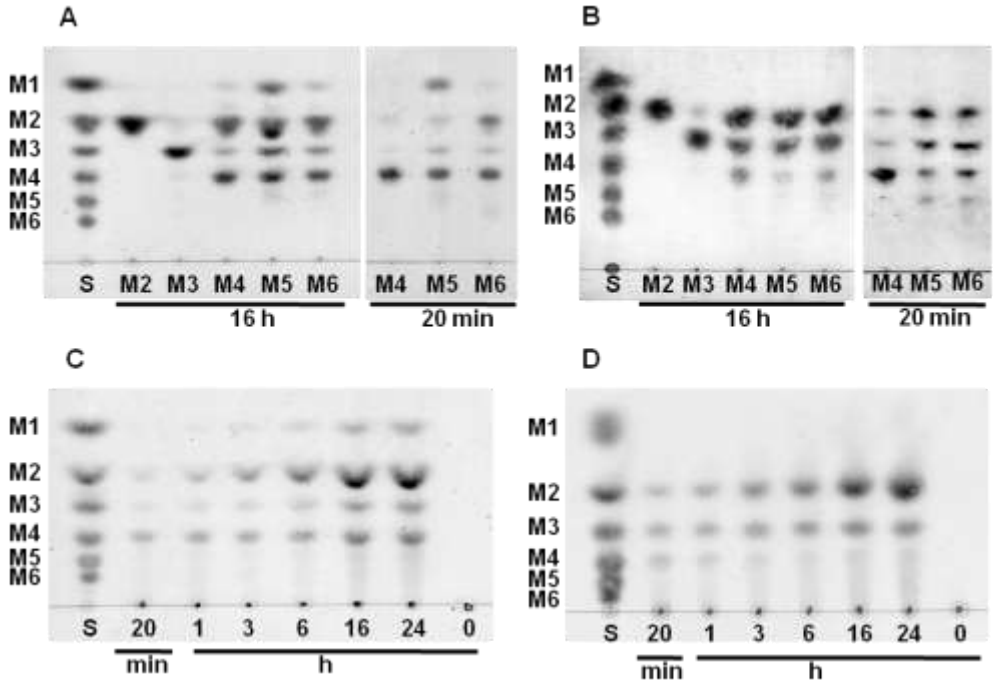
	Activity (µmol/min/µmol)[a] protein toward:				
Enzyme	Konjac glucomannan	Ivory nut mannan	1,4-β-D-Mannan	Carob galactomannan	Guar gum
rGH5-CBM32	23,700 ± 743	10,406 ± 397[b] (1,293 ± 503)[c]	12,679 ± 144 (10,797 ± 145)	11,479 ± 191	ND[d]
rGH5	24,100 ± 1,250	1,119 ± 32.8 (194 ± 16)	4,563 ± 143 (785 ± 26)	6,464 ± 590	ND

[a] Values are means ± standard errors of the means; [b] Activity estimated from the first phase; [c] Activity estimated from the second phase; [d] ND, not detectable

5.1.3.3. Analysis of hydrolysis products by TLC

The action of rGH5-CBM32 and rGH5 against mannooligosaccharides and ivory nut mannan was qualitatively analyzed by TLC (Figure 5.4). When rGH5-CBM32 was incubated with mannooligosaccharides for 16 h, the enzyme displayed negligibly activity against mannobiose and mannotriose, and weak activity against mannotetraose (Fig. 5.4A); the tetrasaccharide was not completely hydrolyzed over 16 h. In contrast, rGH5-CBM32 was highly active on mannopentaose and mannohexaose, with complete conversion of mannopentaose to primarily mannotetraose and mannose, and mannohexaose to mannotetraose and mannobiose within 20 min. rGH5 generated significantly different products to rGH5-CBM32 against mannopentaose and mannohexaose. Thus, rGH5 produced large amounts of mannotriose and mannobiose and a small amount of mannotetraose from both mannopentaose and mannohexaose (Figure 5.4B), while there was a complete absence of mannose (Figure 5.4B). The large amount of mannotetraose and mannotriose generated by rGH5 from mannopentaose and mannotetraose, respectively, which was not mirrored by the appearance of mannose, is strongly indicative of transglycosylation activity. When insoluble ivory nut mannan was treated with rGH5-CBM32 (Figure 5.4C), mannotetraose was detected as the major product in the initial stage of hydrolysis, while mannobiose was the dominant oligosaccharide after 24-h incubation. When rGH5 was incubated with ivory nut mannan mannobiose was the major product; while some mannotetraose was evident no mannose was produced (Figure 5.4D). Similar intensities of hydrolysis products were observed in TLC between rGH5-CBM (Figure 5.4C) and rGH5 (Figure 5.4D), since excessive amounts of the enzymes (0.54 units in 10 μL) were used for digestion of ivory nut mannan.

Figure 5.4| Thin layer chromatography (TLC) showing the enzymatic degradation products of mannooligosaccharides and ivory nut mannan by rGH5-CBM and rGH5.



TLC analysis of hydrolysis products from mannooligosaccharides by rGH5-CBM (**A**) and rGH5 (**B**) and from ivory nut mannan by rGH5-CBM (**C**) and rGH5 (**D**). (A) and (B) Each mannooligosaccharides (100 μg, M2–M6) was incubated with the purified rGH5-CBM or rGH5 (0.18 units each) for 16 h or 20 min and the hydrolysates were analyzed by TLC. (C) and (D) Ivory nut mannan (50 μg each) was treated with of rGH5-CBM or rGH5 (0.54 units each) up to 24 h . Samples were taken at intervals and the hydrolysates were analyzed by TLC.

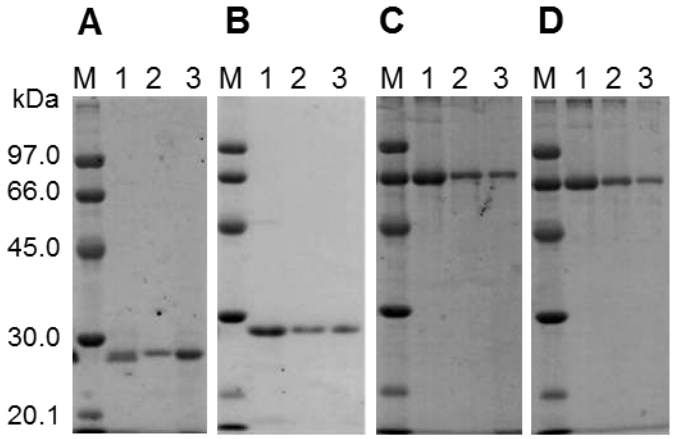
5.1.3.4. Qualitative polysaccharide binding assays

To investigate the function of CBM32, rCBM32 was mixed with insoluble ivory nut mannan and the amount of bound and unbound protein was evaluated by SDS-PAGE. As shown in Figure 5.5A, more rCBM32 was in the precipitated ivory nut mannan than the unbound fraction, while the control protein, BSA, displayed slightly less binding to the insoluble polysaccharide (Figure 5.5C). Both rCBM32 and BSA adsorbed weakly to Funacel, a form of microcrystalline cellulose (Figures 5.5B and 5.5D). These results suggest that rCBM32 may bind to ivory nut mannan, although the number of protein binding sites on the polysaccharide was very low, likely reflecting binding to the non-reducing end of the polysaccharide (see below).

The affinity of rCBM32 for soluble carbohydrates of different origins was qualitatively evaluated by native affinity PAGE (Figure 5.6). The electrophoretic mobility of rCBM32 was slightly retarded by the inclusion of konjac glucomannan and carob galactomannan. In contrast, electrophoresis was not affected by inclusion of guar gum. These results strongly suggest that rCBM32 preferably recognizes less decorated mannans, since guar gum is a highly galactosylated mannan in contrast to carob galactomannan that is less galactosylated

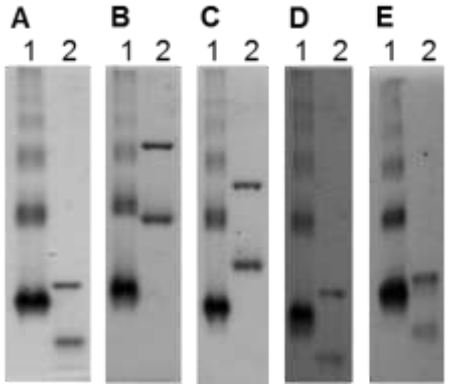
(Daas *et al*., 2002). The specificity of rCBM32 for the termini of mannans is consistent with the limited retardation observed in the mannan-containing gels. In the native affinity PAGE experiments, rCBM32 was detected as two bands, suggesting that rCBM molecules interact weakly to form dimers (Figure 5.6).

Figure 5.5| Adsorption of rCBM to insoluble ivory nut mannan and Funacel.



rCBM was incubated with an insoluble ivory nut mannan (**A**) and Funacel (**B**). After centrifugation, proteins in the supernatant (lane 1), wash (lane 2), and precipitate (lane 3) fractions were analyzed by SDS-PAGE. BSA as a control protein was incubated with ivory nut mannan (**C**) and Funacel (**D**), and each fraction was subjected to SDS-PAGE. Lane M, protein molecular mass standard (molecular masses shown left).

Figure 5.6| Adsorption of rCBM to soluble polysaccharides of different origins.



Affinities of rCBM (lane 2) for konjac glucomannan (**B**), carob galactomannan (**C**), guar gum (**D**), and polygalacturonic acid (**E**) were analyzed by native affinity gel electrophoresis. Lane 1 contains BSA as a control protein. A gel without a polysaccharide served as a reference (**A**).

5.1.3.5. Isothermal titration calorimetry (ITC) analysis

Table 5.2 reports the binding parameters of rCBM32 for *manno*- and *gluco*-configured ligands obtained by ITC analysis. The data revealed broadly similar affinities (~10^4 M^{-1}) for mannobiose, mannotriose and mannotetraose at 25°C. The protein also displayed measurable, but weaker, binding to cellobiose and 4-nitrophenyl β-D-cellobioside. Binding to mannose, N-acetyl glucosamine, and glucose was also observed, but the affinity was too low

to determine accurately (<10^3 M^{-1}). rCBM32 displayed no binding to galactose, β-1,4-mannobiose and 0.1 % high viscosity carob galactomannan. Although binding to 1 % low viscosity carob galactomannan was observed it was too weak to quantify. However, when the polysaccharide was partially digested with *Ct*Man5A binding was evident (K_a ~4 x 10^3) with a low estimated coverage at saturation of one rCBM32 molecule for every 24 mannose residues. These data suggest that rCBM32 contains two sugar binding sites that display a preference for mannose, although they will also recognize glucose-configured ligands. The substantial increase in binding to galactomannans digested with *Ct*Man5A (compared to undigested high or medium viscosity galactomannan) indicates that rCBM32 recognizes the termini of polysaccharides. The observation that the CBM binds to 4-nitrophenyl β-D-cellobioside suggests that the protein recognizes the non-reducing termini of *gluco*- and by inference *manno*-configured β-glycans. The stoichiometry of binding was ~1 indicating that rCBM32 contains one ligand binding site. Typical of CBMs that bind soluble ligands, the interaction of rCBM32 with its carbohydrate targets was driven by a negative enthalpy with the change in entropy having a detrimental effect on affinity.

Table 5.2| Affinity and thermodynamics of the binding of CBM32 to gluco- and manno configured molecules[a].

Ligand	*N*[b]	K_a (M^{-1})	Δ*H* (cal mol^{-1})	TΔ*S* (cal mol^{-1})	Δ*G* (cal mol^{-1})
β-1-4-Mannobiose	1.0 ± 0.02	9.3 x 10^3 ± 338	-10.4 x 10^3 ± 510	-5.0 x 10^3 ± 546	-5.4 x 10^3 ± 36
β-1-4-Mannotriose	1.0 ± 0.05	12.4 x 10^3 ± 2066	-10.2 x 10^3 ± 1267	-4.6 x 10^3 ± 1369	-5.6 x 10^3 ± 106
β-1-4-Mannotetraose	1.2 ± 0.05	11.1 x 10^3 ± 721	-10.7 x 10^3 ± 1698	-5.2 x 10^3 ± 1744	-5.5 x 10^3 ± 50.0
Cellobiose	1.1 ± 0.9	1.6 x 10^3 ± 170	-7.2 x 10^3 ± 6860	-2.9 x 10^3	-4.4 x 10^3
4-Nitrophenyl β-D-cellobioside	1.2 ± 0.3	2.8 x 10^3 ± 200	-5.5 x 10^3 ± 1780	-0.75 x 10^3	-4.7 x 10^3

[a] The binding of CBM32 to the various ligands was determined by ITC. Standard errors of the means for the mannooligosaccharides were determined from triplicate titrations while errors for the fitted line are displayed for cellobiose and 4-nitrophenyl β-D-cellobioside.

[b] *N*, stoichiometry of binding.

5.1.4. Discussion

Clostridium thermocellum Man5A preferred unsubstituted substrates such as konjac glucomannan to highly galactosylated galactomannan such as guar gum (Table 5.1). Thus, galactose residues in galactomannans likely interfere with the binding of *Ct*Man5A to its target substrates. The low activity of *Ct*Man5A against ivory nut mannan can be ascribed to

the crystalline structure of the substrate, which greatly reduces its access to enzyme attack. Indeed, GH5 mannanases generally display low or undetectable activity against crystalline mannan. The activity of GH5 mannanases for soluble substrates varies. For example, *B. circulans* K-1 ManG displayed highest activity toward konjac glucomannan and moderate activity toward locust bean galactomannan (Yoshida, Sako, & Uchida, 1998). In contrast, *C. japonicus* Man5A and *C. josui* Man5A showed the highest activity toward carob or locust bean galactomannan (Gold & Martin, 2007; Sakka *et al*., 2010). These differences in substrate specificities are partially explained by the preference of the subsites for glucose or mannose. By definition mannanases bind mannose at the -1 subsite (active site), as they cleave mannosidic linkages. However, specificity at the distal subsites for glucose or mannose can be highly variable explaining why different mannanases display a preference for galactomannanas and/or glucomannans (Tailford *et al*., 2009).

The ligand preference of CBM32 for β-1,4-*manno*-configured polymers reflects the substrate specificity of the GH5 catalytic module of *Ct*Man5A. CBM32 is a large family with, potentially, a diversity of ligand specificities. Out of the more than one thousand family 32 CBMs, the ligand specificities of only three proteins have been characterized. CBM32 from *Clostridium perfringens N*-Acetyl-β-hexosaminidase GH84C was shown to bind preferentially to β-D-galactosyl-1,4-β-D-*N*-acetylglucosamine (LacNAc) and lactose (Ficko-Blean & Boraston, 2006). In contrast, CBM32 from *Micromonospora viridifaciens* sialidase NedA recognized galactose and lactose (Newstead *et al.*, 2005) while a periplasmic single module polypeptide consisting of CBM32 from *Yersinia enterolitica* selectively bound to highly polymerized galacturonic acid (Abbott, Hrynuik, & Boraston, 2007). The crystal structures of these CBM32s reveal a β-sandwich fold. While the ligand binding site is conserved in these proteins, there are significant differences in the residues that bind to Gal/GalNAc in the CBM32s from *C. perfringens* GH84C and *M. viridifaciens* NedA, compared to the CBM32 *Y. enterolitica* that targets GalA (Abbott *et al*., 2007). Since the CBM32 of *Ct*Man5A has a binding specificity that is different from those well-characterized CBM32s, it is not surprising that the Gal/GalNAc/GalA binding residues are not conserved in the mannan binding CBM32 (Figure 5.2).

Comparison of the hydrolytic activities of rGH5-CBM32 and rGH5 toward different mannans indicated that the CBM32 in *Ct*Man5A plays an important role in the degradation of insoluble mannans, ivory nut mannan and 1,4-β-D-mannan (Table 5.1). Distinguished biphasic actions of both rGH5-CBM32 and rGH5 were observed toward insoluble ivory nut mannan (Figure 5.3), suggesting that these enzymes attack amorphous region of the substrate first and then crystalline region. Since only a weak biphasic action of rGH5-CBM32 was observed toward insoluble 1,4-β-D-mannan and the first phase activity of rGH5-CBM32 toward ivory nut mannan was comparable to the second phase activity toward 1,4-β-D-mannan (Table 5.1),

1,4-β-D-mannan seems to consist of solely amorphous structure. Nevertheless, rGH5 showed a biphasic action toward 1,4-β-D-mannan, suggesting that this substrate consists of two distinct regions, more sensitive or more resistant to rGH5 attack. It is plausible that the presence of CBM32 enhanced catalytic activity of the appended catalytic module by increasing enzyme concentration in the vicinity of the substrate. In addition, the presence of CBM32 in rGH5-CBM32 may decrease resistance of the substrate to rGH5 attack. Activity-stimulating effects of CBMs toward insoluble substrates have been observed in cellulases and other glycoside hydrolases (Maglione *et al*., 1992; Mangalaa *et al*., 2003; Sakka *et al*., 2011). For example, the family-3 CBM of *Paenibacillus curdlanolyticus* B-6 Xyn10D has been recently shown to be important for hydrolysis of insoluble arabinoxylan and natural biomass by comparison of the parental and truncated enzymes (Sakka *et al*., 2011), while the addition of family-6 CBMs of *Clostridium stercorarium* XynA to *Bacillus holodurans* xylanase XylA increased hydrolytic activity of the chimeric enzyme toward insoluble oat spelt xylan but not soluble birchwood xylan (Mangalaa *et al*., 2003). Uniquely the CBM32 in rGH5-CBM32 appeared to also influence the mode of action of the catalytic module. Thus, rGH5-CBM32 produced equal amounts of mannose and mannotetraose from mannopentaose, and mannose and mannotriose from mannotetraose. In contrast rGH5 generated mannobiose and mannotriose from mannotetraose, while the production of the tetrasaccharide from mannopentaose was not associated with the appearance of mannose (Figure 5.4). This unbalance between complementary oligosaccharide release (e.g. if mannotetraose is hydrolyzed to generate mannotriose, then an equal amount of mannose should be generated) is typical of transglycosylation reactions. GH5 enzymes, including *Ct*Man5A, have a retaining reaction mechanism that proceeds by a double displacement reaction in which a covalent glycosyl-enzyme intermediate is firstly formed (glycosylation step) and then hydrolyzed (deglycosylation step) through general acid/base-catalysis. Transglycosylation occurs when the glycone of the glycosyl-enzyme intermediate is transferred to an oligosaccharide acceptor rather than water. In the case of rGH5-CBM32, it is possible that the presence of CBM32 sterically hinders the access of large acceptor molecules, such as mannooligosaccharides, to the +1 subsite of the covalent glycosyl- enzyme intermediate, reducing transglycosylation.

5.1.5. Conclusions

In conclusion, CBM32 from *Ct*Man5A displays a preference for the non-reducing end of β-*manno*-configured oligosaccharides, a specificity not previously observed in family 32 CBMs. The module appears to play an important role in the hydrolysis of insoluble mannans. Furthermore, the presence of CBM32 affects the mode of action of the catalytic module of *C. thermocellum* Man5A.

6. BIOCHEMICAL CHARACTERIZATION OF PECTATE LYASES FROM *Clostridium thermocellum*

6.1. Role of pectinolytic enzymes identified in *Clostridium thermocellum* cellulosome

Vania O. Fernandes[2,3*], Soumyadeep Chakraborty[1*], Fernando M.V. Dias[2], Jose A.M. Prates[2], Luis M.A. Ferreira[2,3], Carlos M.G.A. Fontes[2,3], Arun Goyal[1], Maria S.J. Centeno[2]

[1] Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India; [2] CIISA-Faculdade de Medicina Veterinária, Universidade de Lisboa, Av. da Universidade Técnica, 1300-477 Lisboa, Portugal; [3] NZYTech Lda, Estrada do Paço do Lumiar, Campus do Lumiar, Edifício E-R/C, 1649-038 Lisboa, Portugal

*Equal contribution

Adapted from: Fernandes, *et al*. (2015*). PLOS one*. 10(2):e0116787

Abstract

The cloning, expression and characterization of three cellulosomal pectinolytic enzymes viz., two variants of PL1 (PL1A and PL1B) and PL9 from *Clostridium thermocellum* was carried out. The comparison of the primary sequences of PL1A, PL1B and PL9 revealed that these proteins displayed considerable sequence similarities with family 1 and 9 polysaccharide lyases, respectively. PL1A is a putative catalytic domain of protein sequence ABN54148.1, while PL1B and PL9 are two putative catalytic domains of ABN53381.1. These two protein sequences also contain putative carbohydrate binding module (CBM) and type-I dockerin. The associated putative CBM of PL1A showed strong homology with family 6 CBMs while those of PL1B and PL9 showed homology with family 35 CBMs. Recombinant derivatives of these three enzymes showed molecular masses of approximately 34 kDa, 40 kDa and 32 kDa for PL1A, PL1B and PL9, respectively. PL1A, PL1B and PL9 displayed high activity toward polygalacturonic acid and pectin (up to 55% methyl-esterified) from citrus fruits. However, PL1B showed relatively higher activity towards 55% and 85% methyl-esterified pectin (citrus). PL1A and PL9 showed higher activity on rhamnogalacturonan than PL1B. Both PL1A and PL9 displayed maximum activity at pH 8.5 with optimum temperature of 50 ºC and 60 ºC, respectively. PL1B achieved highest activity at pH 9.8, under an optimum temperature of 50°C. PL1A, PL1B and PL9 produced two or more unsaturated galacturonates from pectic substrates as displayed by TLC analysis confirming that they are endo-pectate lyase belonging to family 1 and 9, respectively. This report reveals that pectinolytic activity displayed by *Clostridium thermocellum* cellulosome is coordinated by a sub-set of at least three multi-modular enzymes.

6.1.1. Introduction

Plant cell wall degradation carried out by saprophytic and phytopathogenic microbes is essential for the recycling of carbon stored in plant biomass and of intrinsic biotechnological importance. Plant cell walls are composed of a complex network of polysaccharides, primarily cellulose, hemicelluloses and pectic substances (Carpita & Gibeaut, 1993). Pectins are a highly heterogeneous group of polymers, containing a high quantity of galacturonic acid, which contribute to the firmness and structure of plant tissues, and is predominantly found in the primary cell wall and middle lamella. This recalcitrant carbohydrate is more soluble in water than cellulose and hemicelluloses, suggesting that it constitutes the initial target for plant associated microbes attack (Ochiai *et al*., 2007). Pectins are divided into three polysaccharides i.e., homogalacturonan (HG), rhamnogalacturonan type-I (RG-I), and rhamnogalacturonan type-II (RG-II). HG is present as a linear backbone, while RG-I and RG-II are branched carbohydrates (Darvill, McNeil, & Albersheim, 1978). The backbone of RG-I pectin is composed of alternating rhamnose and galacturonic acid residues with a disaccharide repeated unit consisting of [(1,2)-α-L-rhamnose-(1,4)-α-D-galacturonic acid] where galacturonic acid residues may be acetylated at the O2 or O3 positions (McNeil, Darvill, & Albersheim, 1980; McDonough *et al*., 2004). It is often referred as “hairy” because of the presence of multiple side chains of neutral polymers like arabinans, galactans and arabinogalactans which are attached to C4 of the rhamnose residues (Ridley, O Neil, & Mohnen, 2001; De Vries & Visser, 2001). RG-II consists of a polygalacturonan backbone with side chains complexes of about 30 monosaccharides including rare molecules such as apiose and aceric acid (O Neill *et al*., 1996).

Due to its structural complexity, pectin degradation requires the concerted action of several enzymes. The enzymatic degradation of polygalacturonan involves two well-known enzymatic mechanisms: i) hydrolysis by glycoside hydrolases (GH) that cleave glycoside bonds in the polysaccharide and ii) β-elimination reactions carried out by polysaccharide lyases resulting in oligomers with Δ4,5 unsaturated residues at the non-reducing end (Linhardt, Galliher, & Cooney, 1986; Davies & Henrissat, 1995). Polysaccharide lyases (EC 4.2.2.-) belong to a large group of enzymes defined as carbohydrate-active-enzymes and have been classified into 23 families (February 2015), according to CAZy database (Cantarel *et al*., 2009). Pectate lyases of families 1, 2, 3, 9, and 10 catalyse the β-eliminative cleavage of α-1,4-glycosidic bond between D-galactopyranosyluronic acid (Gal*p*A) residue in pectate (a low methylesterified form of pectin), and generate Δ4,5 unsaturated Gal*p*A as the product (Pagès *et al*., 2003), which exhibits a maximum absorbance at around 235 nm (Jurnak *et al*., 1996).

Pectate lyases are widely distributed among microbial plant pathogens like *Erwinia* (Hugouvieux-Cotte-Pattat *et al*., 1996; Pissavin, Robert-Baudouy, & Hugouvieux-Cotte-Pattat, 1996; Shevchik, Robert-Baudouy, & Hugouvieux-Cotte-Pattat, 1997) although they

have also been found in saprophytic bacteria including the genus *Bacillus* (Soriano, Diaz, & Pastor, 2006; Ochiai *et al*., 2007; Sukhumsiirchart *et al*., 2009) and *Clostridium* (Pagès *et al*., 2003). *Clostridium thermocellum* is an anaerobic, saccharolytic and thermophilic bacterium that organizes a consortium of plant cell wall degrading enzymes in a large multienzymatic complex termed the cellulosome (Bayer, Kening, & Lamed, 1983; (Lamed, Setter, & Bayer, 1983). The cellulosome is assembled via the interaction of individual type-I dockerins located at the C-terminus of enzymes into one of the nine cohesins of the scaffoldin subunit, CipA. CipA also bears a family 3 carbohydrate-binding module (CBM) which accounts for its cellulose-targeting function and a dockerin type-II that mediates the attachment of the entire complex into the bacterial cell surface (Bayer, Kening, & Lamed, 1983; Bayer *et al*., 2008). Despite its specialization in the hydrolysis of crystalline cellulose, the cellulosome contains in addition to several cellulases, an extensive group of hemicellulases, which have been extensively characterized (Zverlov *et al*., 1994; Halstead *et al*., 1999; Fernandes *et al*., 1999; Blum *et al*., 2000; Fontes & Gilbert, 2010), and are believed to increase the accessibility of the bacterium into its primary substrate.

The majority of glycoside hydrolases that attack cellulose and hemicelluloses are modular enzymes consisting of catalytic modules appended to non-catalytic carbohydrate-binding modules (CBMs) (Davies & Henrissat, 1995). Pectinases on the contrary generally have a relatively simple structure lacking CBMs, which is possibly explained by the accessibility of pectins to soluble biocatalysts (McKie *et al*., 2001). CBMs are described to date into 71 families (February 2015), according to CAZy database (http://www.cazy.org) and continue to expand. The present study provides data indicating that *C. thermocellum* cellulosome secretes modular polysaccharides lyases belonging to PL families 1 and 9. The role of this subset of enzymes in the anaerobic conversion of biomass by cellulosomes was investigated.

6.1.2. Materials and Methods

6.1.2.1. Bacterial strains, plasmids and culture conditions

The *Escherichia coli* strains used in this study were NZYStar (NZYTech, genes & enzymes, Portugal.), BL21 (DE3) and BL21(DE3) pLysS (Novagen). The plasmid vectors used were pNZY28 (NZYTech, genes & enzymes), pGEM-T Easy vector (Promega), pET21a and pET28a (Novagen). *E. coli* strains containing recombinant plasmids were cultured in LB broth medium supplemented with 100 µg/mL ampicillin or 50 µg/mL kanamycin. To generate the recombinant proteins encoded by pET21a or pET28a expression vectors, *E. coli* BL21(DE3) were cultured at 37 ºC to mid-exponential phase (OD_{550nm} = 0.6) and at this point isopropyl-β-D-thiogalactoside (IPTG) was added to a final concentration of 1 mM. Incubation conditions after induction with IPTG were 16h at 19°C for PL1A and PL9, 12h at 24°C for PL1B.

6.1.2.2. Substrates used in enzyme assays

Polygalacturonic acid (PGA) from citrus fruits, rhamnogalacturonan from soybean (RGAS) and potato (RGAP), pectic galactans from potato (PGP) and lupin (PGL), were purchased from Megazyme. Pectins from citrus fruits (with varying degrees of methyl-esterification, PC) and apple (PA) were purchased from Sigma Chemical Co., USA.

6.1.2.3. General recombinant DNA procedures

Bacterial transformation, agarose gel electrophoresis, plasmid DNA preparation, restriction endonuclease digestion and ligation of DNA sequences were followed as described elsewhere (Sambrook, Fritsch, & Maniatis, 1989).

6.1.2.4. Construction of recombinant plasmids

Genes encoding PL1A, PL1B and PL9 were amplified from *C. thermocellum* genomic DNA, using one IU of thermostable DNA polymerase NZYSpeedy Proof (NZYTech, genes & enzymes, Portugal) and primer pairs, described in Table 6.1. *Nhe*I/*Xho*I restriction sites were used for directional cloning of the respective amplified DNA sequences into the expression vectors pET21a and pET28a. The reactions, in a final volume of 50 µL, were subjected to 30 cycles at the following temperatures: 95 ºC for 1 min, 55 ºC for 1 min and 72 ºC for 2.5 min. The amplified genes after running on agarose gel were purified by gel extraction kit (Qiagen), and were cloned into pNZY28 (NZYTech, genes & enzymes, Portugal) or pGEMT-Easy (Promega) and sequenced to ensure that no mutation occurred during PCR. Recombinant pNZY28 and pGEMT-Easy derivatives were digested with *Nhe*I/*Xho*I restriction enzymes (NZYTech, genes & enzymes or Promega). *pL*1a and *pL*9 genes were cloned into similarly digested expression vector pET21a(+) whereas *pL*1b gene was cloned into pET28a(+) vector. Recombinant PL1A and PL9 proteins contained a C-terminal His_6-tag, whereas PL1B contained an N-terminal His_6-tag.

Table 6.1| **Primers used in PCR of *pl1A*, *pl1B* and *pl9* genes.** The nucleotides shown in bold are the restriction enzyme sites, which were used to clone amplified genes into the expression vectors pET21a(+) and pET28a(+).

Protein	Vector	Primers
PL1A	pET21a(+)	5´- ctc**gctagc**accacttatgcccagacc -3´
		5´- cac**ctcgag**ggagaatgttcccgggac -3´
PL1B	pET28a(+)	5´- ctct**gctagc**gcgccaagctttgaactg -3´
		5´- cgcg**ctcgag**ctgctgagtatttttcgg -3’
PL9	pET21a(+)	5´- ctc**gctagc**gcaatccctgtggaaggagac -3´
		5´- cac**ctcgag**tttaaatattgcattgtcaag -3´

6.1.2.5. Expression and purification of PL1A, PL1B and PL9

PL encoding genes were expressed using *E. coli* BL21(DE3) where the cells were induced with 1mM IPTG, only when the culture OD A_{600} reaches 0.4 – 0.6. Induced cells were grown overnight at 24 °C and 180 x*g* for protein production. Cells were harvested by centrifugation at 12000 x*g*, 4 ºC for 20 min and the bacterial pellets were resuspended in 50 mM sodium HEPES buffer (pH 7.5) containing 1 M NaCl, 5 mM $CaCl_2$ and 10 mM imidazole for PL1A and PL9. PL1B containing cells were resuspended in 50 mM Tris-HCl buffer (pH 8.6) containing only 100 mM NaCl. These three recombinant proteins containing His-tags were purified by immobilized nickel ion affinity chromatography as described previously (Carvalho *et al*., 2004). For PL1A and PL9 the buffer was exchanged to 50 mM Tris-HCl, pH 8.5, containing 5 mM $CaCl_2$ and 100 mM NaCl and for PL1B the buffer was exchanged to 50 mM Tris-HCl (pH 8.6) containing 100 mM NaCl. The purity and size of recombinant enzymes were evaluated by SDS-PAGE (Laemmli, 1970).

6.1.2.6. Enzyme assays

The enzyme activity of PL1A, PL1B and PL9 was determined against different pectic substrates. 30 µg of PL1A or PL9 was incubated with 0.5% (w/v) of substrate dissolved in 50 mM Tris-HCl buffer pH 8.5 containing 5 mM $CaCl_2$ and 100 mM NaCl at 60 ºC for 20 min. The assay of PL1B was carried out by incubating 7 µg of enzyme with 0.1% (w/v) of substrate in 50 mM Glycine-NaOH buffer (pH 9.8) containing 0.6 mM $CaCl_2$ for 15 min at 50 ºC. The reactions were stopped by incubation on ice for 10 min and centrifuged at 13,000 x*g* for 5 min. The supernatant containing the released unsaturated products was measured by spectrophotometer (Ultrospec III Pharmacia and Cary 100 Bio Varian). The molar extinction coefficient used for the unsaturated product released at A_{232} nm, was 5,200 $M^{-1}cm^{-1}$ (Collmer, Riad, & Mount, 1988) and at A_{235} nm, was 4,600 $M^{-1}cm^{-1}$ (Hasegawa & Nagel, 1966). 1 Unit of enzyme was defined as the amount of enzyme that forms 1 µmol of 4,5-unsaturated product per minute, under the described assay conditions.

To determine the maximum activity of PL1A and PL9 at different pH values, all enzymes were incubated with appropriate substrates at 50°C in the following buffers: 50 mM MES (pH 6.5); 50 mM Tris-HCl (pH 7.0 to 8.5) and 50 mM $NaHCO_3$ (pH 9.0 to 12.0), and the activity was determined at A_{232} nm as described above. Activity of PL1B at different pH values was determined by incubating with PGA at 50°C using following buffers: 50 mM Tris-HCl (pH 7.6-8.8), 50 mM Glycine-NaOH (pH 9.0-10.6) and 50 mM Na_2HPO_4-NaOH (10.8-12), and the activity was determined at A_{235} nm as mentioned earlier. The optimal activities of PL1A and PL9 at a range of temperature from 10 to 100 ºC, were determined spectrophotometrically at A_{232} nm, by incubating the enzymes in 50 mM Tris-buffer pH 8.5, for 20 min. The optimal activity of PL1B in the temperatures range from 10 to 100 ºC was spectrophotometrically determined at A_{235} nm in 50 mM Glycine-NaOH buffer pH 9.8, after 15 min of incubation. The

thermostability of PL1A and PL9 was evaluated, by incubating the enzyme at different temperatures (30 to 100 °C) in 50 mM Tris-HCl pH 8.5, and for PL1B the enzyme was incubated at same temperature range in 50 mM Tris-HCl (pH 8.6) for 30 min (Fontes *et al*., 1995b), and then the residual activity was measured by assay methods described earlier.

Kinetic parameters of these three Clostridial enzymes were measured against PGA (citrus) having an average molecular weight approximately, 25000 g/mol (White, Katona, & Zodda, 1999). 20 µL (1.5 mg/mL) of enzyme (PL1A or PL9) was used in 1 mL reaction mixture containing 50 mM Tris-HCl (pH 8.5), 5 mM $CaCl_2$ with varying concentrations (0.01 to 0.5% w/v) of PGA were incubated at 50 ºC. The unsaturated product formation was monitored spectrophotometrically at A_{232} nm. Similarly, 20 µL (1.4 mg/mL) of PL1B was used in 1 mL of reaction volume containing 50 mM Glycine-NaOH (pH 9.8), 0.6 mM $CaCl_2$ and varying concentration of PGA (0.01 to 0.5% w/v). The reaction mixture was incubated at 50 ºC and released unsaturated product was monitored spectophotometrically at A_{235} nm. K_{cat} and K_m were determined using the Michaelis-Menten equation. All the reactions were carried out in triplicate and results were reported as mean±SD.

6.1.2.7. Analysis of enzyme degradation products

PL1A (6 µg), PL1B (7 µg) and PL9 (6 µg) were separately incubated in 1 mL reaction volume containing 0.1% (w/v) PGA or citrus pectin (25% methyl-esterified). The reaction was carried out under optimized conditions of pH and temperature for different time intervals from 0 to 60 min. After the reaction enzyme was deactivated by keeping on ice for 5 min and the sample was treated with equal volumes of ethanol to precipitate un-hydrolyzed polysaccharides and protein. Ethanol was removed and samples were concentrated to 500 µL by heating at 50 °C. 1 µL of sample was then loaded on the TLC plate (readymade silica coated aluminum TLC plates obtained from Merck, Germany) for running the degradation products under a solvent system containing butan-1-ol/water/acetic acid in the ratio of 5:3:2 (Lojkwoska *et al*., 1995). The spots on TLC plates were visualized by a solution containing 0.5% (w/v) α-naphthol and 5% (v/v) sulphuric acid in ethanol (Cote & Leathers, 2005), after heating at 95°C for 10 min in hot air oven. Standard oligogalacturonides like D-galacturonic acid (S1), di-galacturonic acid (S2) and tri-galacturonic acid (S3) (procured from Sigma Chem. Co., USA) were used to analyze the degradation product formed from different substrates upon enzymatic treatment.

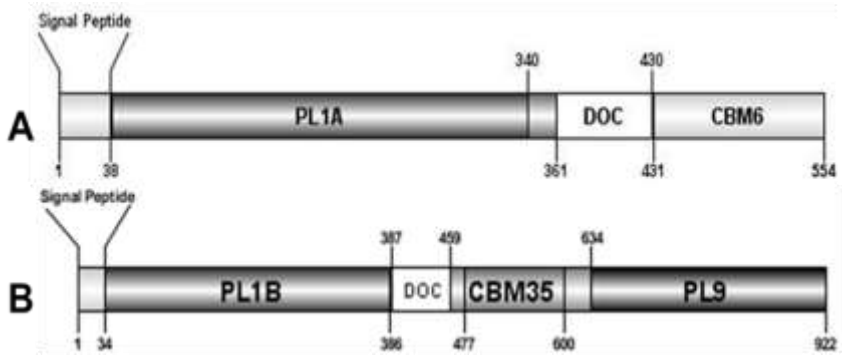
6.1.3. Results

6.1.3.1. Molecular architecture of three pectin degrading enzymes from *Clostridium thermocellum*

Inspection of two protein sequences from *C. thermocellum* revealed that they contain one gene (*pL*1A) in the sequence ABN54148.1 (Figure 6.1A) and two genes (*pL*1B and *pL*9) in the sequence ABN53381.1 (Figure 6.1B) all putatively expressing lyase activities. Both the

protein sequences are associated with type I dockerin, which is the signature module of the cellulosomal proteins. Analysis of the deduced amino acid sequence of the three enzymes revealed characteristic N-terminal signal sequences with putative cleavage sites located between Ala-37/Thr-38 (ABN54148.1) and between Ala-33/Ala-34 (ABN53381.1) suggesting that the proteins are exported into the extracellular space. Homology searches using Blast (www.ncbi.nlm.nih.gov/BLAST), revealed downstream the signal peptide of ABN54148.1 a putative 302-aa family 1 PL (PL1A) followed by a 70-aa dockerin domain and a 124-aa C-terminal family 6 CBM (CBM6) (Figure 6.1A). ABN53381.1 contains a 353-aa N-terminal family 1 PL (PL1B) and a 289-aa C-terminal family 9 PL (PL9). Sandwiched between these two catalytic domains is a 73-aa dockerin domain and 124-aa family 35 CBM (CBM35) (Figure 6.1B).

Figure 6.1| Molecular architecture of modular protein sequences with accession numbers ABN54148.1 and ABN53381.1.



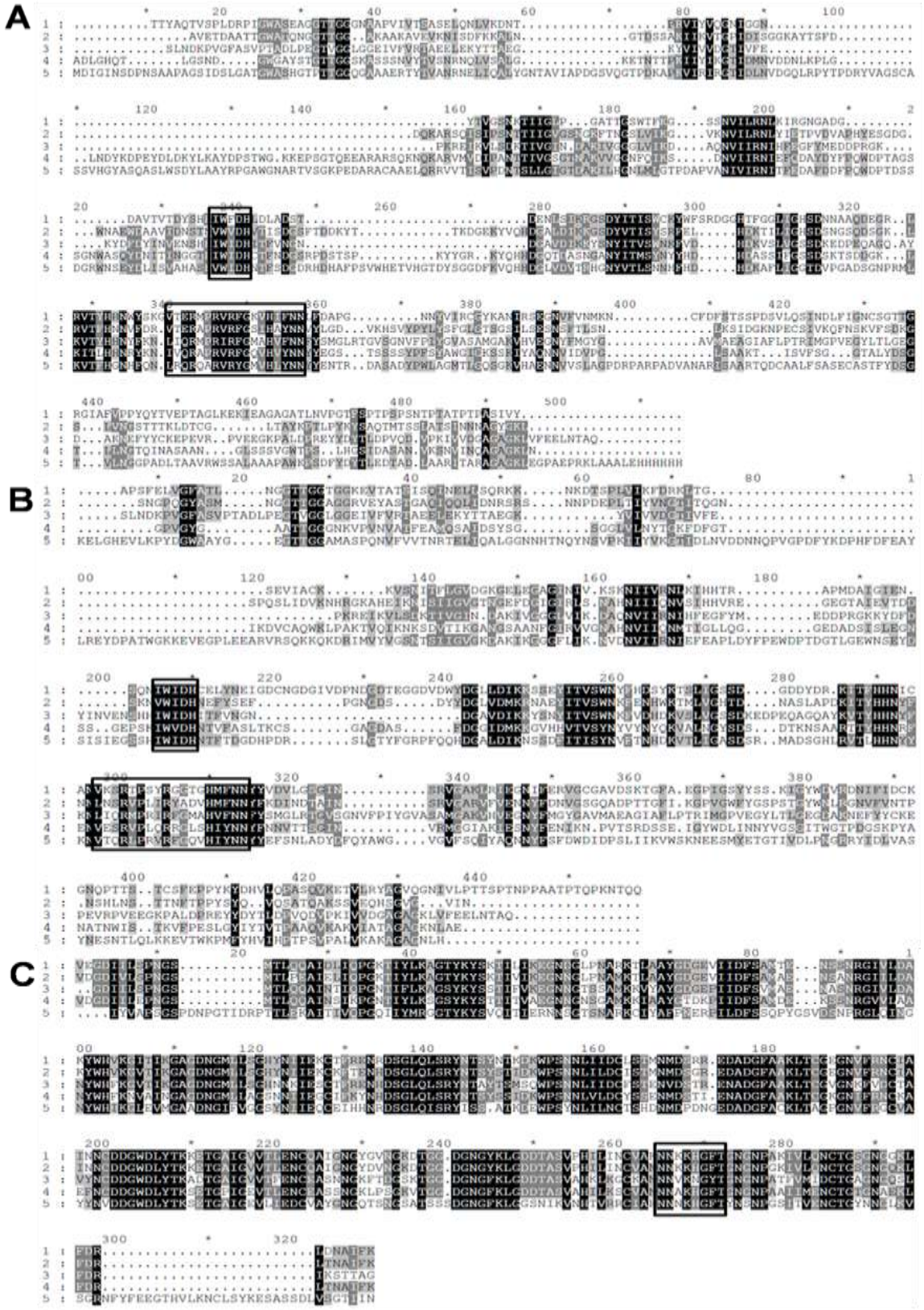
(A) Protein sequence with accession no. ABN54148.1 that comprises of N-terminal PL1A catalytic domain, followed by DOC, type-I dockerin and C terminal CBM6 binding domain. **(B)** Protein sequence with ac no. ABN53381.1 that comprises of N terminal PL1B catalytic domain, followed by DOC, type-I dockerin and CBM35 binding domain, with a C-terminal PL9 catalytic domain.

Alignment of PL1A (Figure 6.2A) and PL1B (Figure 6.2B) domain with other PL1 homologues revealed two consensus sequence patterns in the enzymes catalytic domains, “VWIDH” and “VxxRxPxxRxGxxHxxxN”, which are signature regions of pectate lyases (Pel) (Hinton *et al*., 1989; Barras, Van Gigsegem, & Chatterjee, 1994; Henrissat *et al*., 1996). The conserved arginine residue observed in the second region, identified as R-218 in the superfamily pectate lyase C (PelC from *Erwinia chysanthemi*) is the catalytic residue involved in proton abstraction (Yoder, Keen, & Jurnak, 1993; Scavetta *et al*., 1999). Cleavage of glycosidic bonds in which the aglycone sugar is galacturonic acid can be acid-base-assisted catalysis, mediated by glycoside hydrolases (Koshland, 1953) or via β-elimination reaction, which is initiated by proton abstraction from C-5 of the galacturonosyl residue on the reducing end of the glycosidic bond (Moran, Nasuno, & Starr, 1968). As R-218 belongs to a potential group

or groups involved in the proton abstraction in PelC and as it is highly conserved in catalytic domains of PL1A (position 190 in Figure 6.2A) and PL1B (position 209 in Figure 6.2B) it suggests that these three enzymes cleave α-1,4-linked galacturonic acid units of the pectate component of the plant cell wall by a β-elimination mechanism as expected for pectate lyases. The “VWIDH” region is highly conserved in PL1A and PL1B (Figure 6.2 A and B) and it is involved in the membrane transport and in the protein fold (Bruhlmann & Keen, 1997). Structural motifs (parallel β-helix) are also identified in pectate lyases from *Erwinia chysanthemi* and *Bacillus subtilis* (Yoder *et al*., 1993; Lietzke *et al*., 1994; Pickersgill *et al*., 1994).

PL9 catalytic domain bears profound similarity with Pel9A from *Erwinia chryanthemi*. Pel9A showed an endolytic cleavage pattern where it cleaves the polysaccharide by anti-β-elimination mechanism, where a base catalyzed abstraction of proton is carried out from the C5 carbon (Anderson, 1998). In case of Pel9A the putative base is Lysine rather than Arginine found in other endo-pectate lyases (Jenkins *et al*., 2004). Such a consensus sequence was found in the catalytic domain of PL9, whose translated amino acid sequence is highly conserved in all the aligned sequences marked within a box in Figure 6.2C. The lysine residue involved in proton abstraction during β–elimination is located in position 269 of the PL9 sequence (Figure 6.2C).

Figure 6.2| Multiple Sequence Alignment of family 1 and 9 polysaccharide lyases performed by CLUSTALW program and viewed in GeneDoc ver2.7.



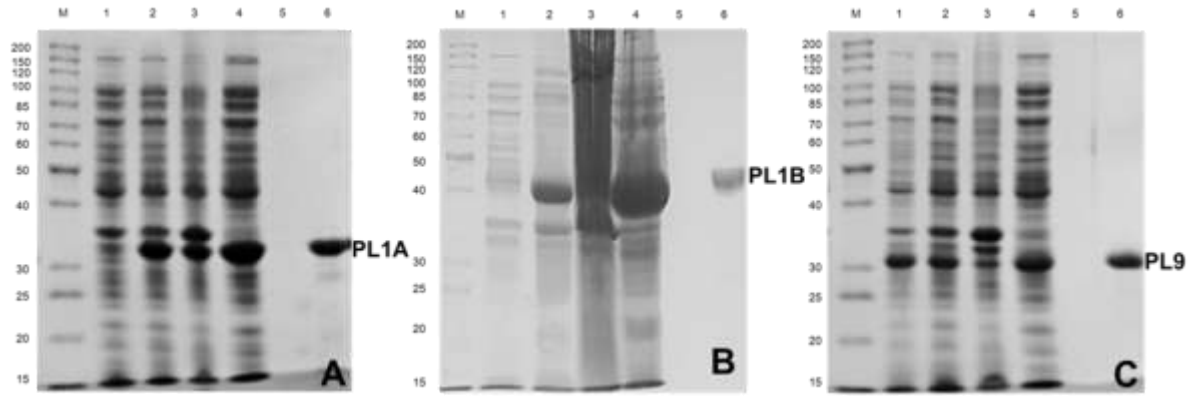
Identical amino acid residues common in all sequences are represented by black shade, and similar amino acids which are common in atleast 3 out off 5 sequences are represented by grey shades. Consensus sequences that characterize pectate lyases are inside of black squares. (**A**) PL1A alignment was with following proteins: 1 (PL1A,

C. thermocellum ATCC 27405); 2 (*Erwina chrysanthemi*, PDB: 1PCL); 3 (*Thermotoga maritima*, PDB: 3ZSC); 4 (*Bacillus subtilis*, PDB: 3KRG) and 5 (Acidovorax Avenae Subsp Citrulli, PDB: 4HWV). (**B**) PL1B was aligned with following proteins: 1 (PL1B, *Clostridium thermocellum* ATCC 27405); 2 (Bacillus Sp. N16-5, PDB: 3VMV); 3 (*Thermotoga Maritima*, PDB: 3ZSC); 4 (*Xanthomonas Campestris* ATCC 33913, PDB: 2QX3); 5 (*Bacillus Sp*. TS-47, PDB: 1VBL). (**C**) PL9 was aligned with following proteins: 1 (PL9, *Clostridium thermocellum* ATCC 27405); 2 (*Clostridium straminisolvens* JCM 21531, GAE89695.1); 3 (Clostridium cellulovorans 743B, YP_003842407.1); 4 (*Acetivibrio cellulolyticus*, WP_010245176.1); 5 (*Caldicellulosiruptor kristjanssonii* I77R1B, YP_004026944.1).

6.1.3.2. Cloning, expression and purification of recombinant PL1A, PL1B and PL9

DNA sequences of 906, 1059 and 867 bp, encoding PL1A, PL1B and PL9 respectively, were amplified by PCR and cloned into pET21a and pET28a expression vector as described in the method section. The recombinant proteins containing the His_6-tags were purified by immobilized metal ion affinity chomatography. The expression and purification of PL1A, PL1B and PL9 proteins was analyzed by SDS-PAGE as shown in Figure 6.3 A, B and C displaying molecular size of approximately 34, 40 and 32 kDa, respectively.

Figure 6.3| Hyper-expression and purification of PL1A, PL1B and PL9 using *E. coli* BL21(DE3) cells.



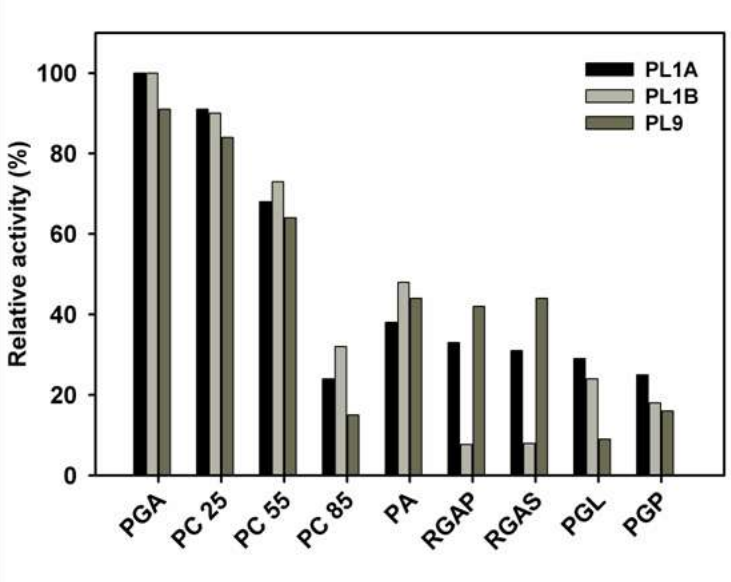
The purity of the proteins was analysed by SDS-PAGE using 10% (w/v) gel showing **(A)** PL1A (34 kDa); **(B)** PL1B (40 kDa); (C) PL9 (32 kDa**);** Lane M: Fermentas high range protein molecular weight marker; Lane 1: Uninduced BL21 cells; Lane 2: IPTG induced BL21 cells; Lane 3: Cell pellet after sonication; Lane 4: Cell free extract, Lane 5: Last wash from column and Lane 6: Purified recombinat enzyme.

6.1.3.3. Biochemical properties of PL1A, PL1B and PL9

The biochemical role of *C. thermocellum* cellulosomal PL1A, PL1B and PL9 enzymes was investigated by analyzing their activity against different substrates. All the three enzymes PL1A, PL1B and PL9 were predominantly active towards polygalacturonic acid (PGA) and pectin both from citrus (Figure 6.4). PL1B displayed relatively higher activity with 55% and 85% methyl-esterified pectins from citrus than PL1A and PL9. On the other hand PL1A and

PL9 showed 30-40% relative activity with rhamnogalacturonan from potato (RGAP) and soybean (RGAS), whereas PL1B showed only 8% relative activity.

Figure 6.4| Substrate specificity of PL1A, PL1B and PL9 towards pectic polysaccharides.



PGA: Polygalacturonic acid, PC 25: Pectin (Citrus) (25% methyl-esterified), PC 55: Pectin (Citrus) (55% methyl-esterified), PC 85: Pectin (Citrus) (85% methyl-esterified), PA: Pectin (Apple), RGAP: Rahmnogalacturonan (Potato), RGAS: Rahmnogalacturonan (Soybean), PGL: Pectic galactan (Lupin), PGP: Pectic galactan (Potato).

The effect of pH and the temperature on the activity of the recombinant PL1A, PL1B and PL9 enzymes against PGA was determined. The results showed that PL1A, PL1B and PL9 were active under alkaline conditions. PL1A and PL9 were active within pH range (6.5-9.5) showing highest activity at pH 8.5 (Figure 6.5A and C). PL1B was active within pH range (8-10) displaying highest activity at pH 9.8 (Figure 6.5B).

The optimum temperature was found at 50 ºC for both PL1A and PL1B and 60 ºC for PL9 (Figure 6.6 A, B and C) which were expected because these enzymes originated from a thermophilic bacterium. Nevertheless all the three recombinant enzymes, PL1A, PL1B and PL9 displayed thermostability within the temperature range of 30 ºC to 70 ºC for 30 min (Figure 6.7 A, B and C).

Figure 6.5| Effect of pH on the activity of (A) PL1A; (B) PL1B; (C) PL9 towards PGA as substrate.

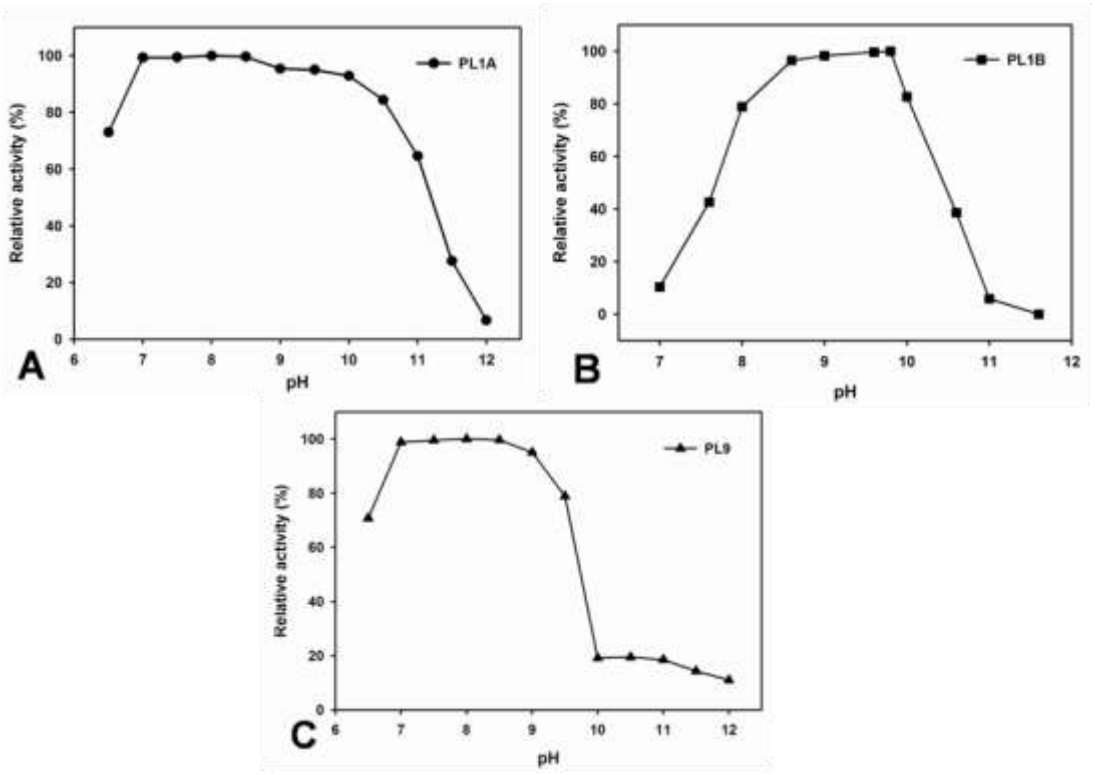


Figure 6.6| Effect of temperature on the activity of (A) PL1A; (B) PL1B; (C) PL9 towards PGA as substrate.

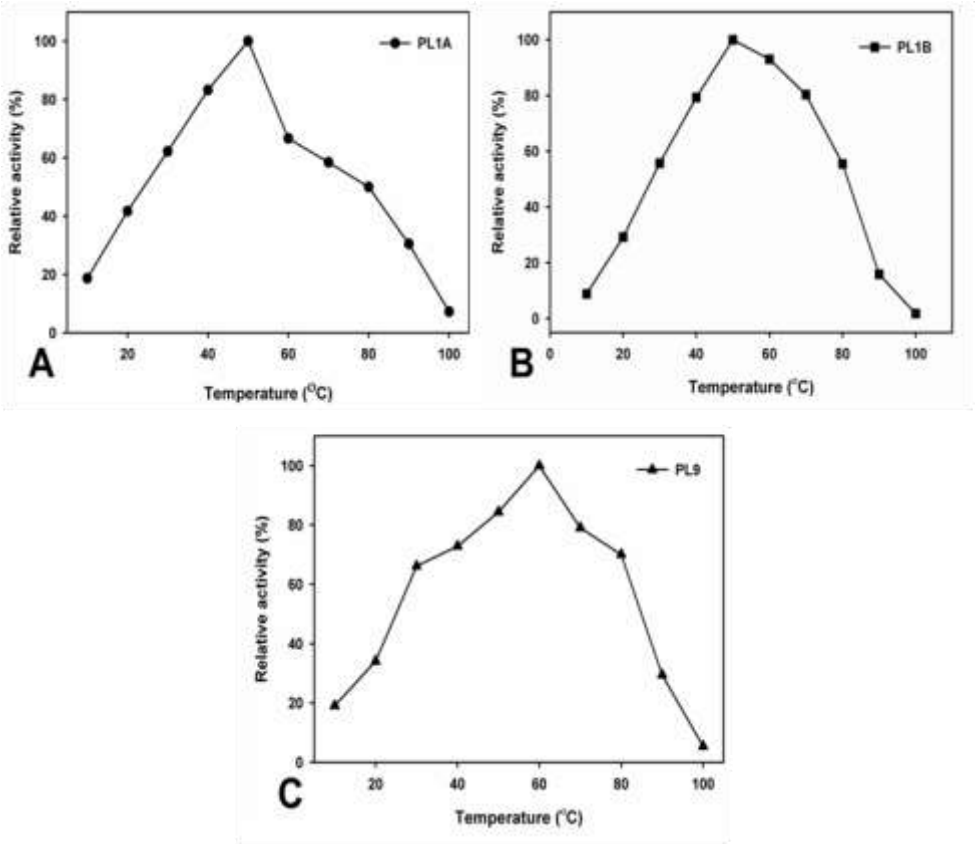
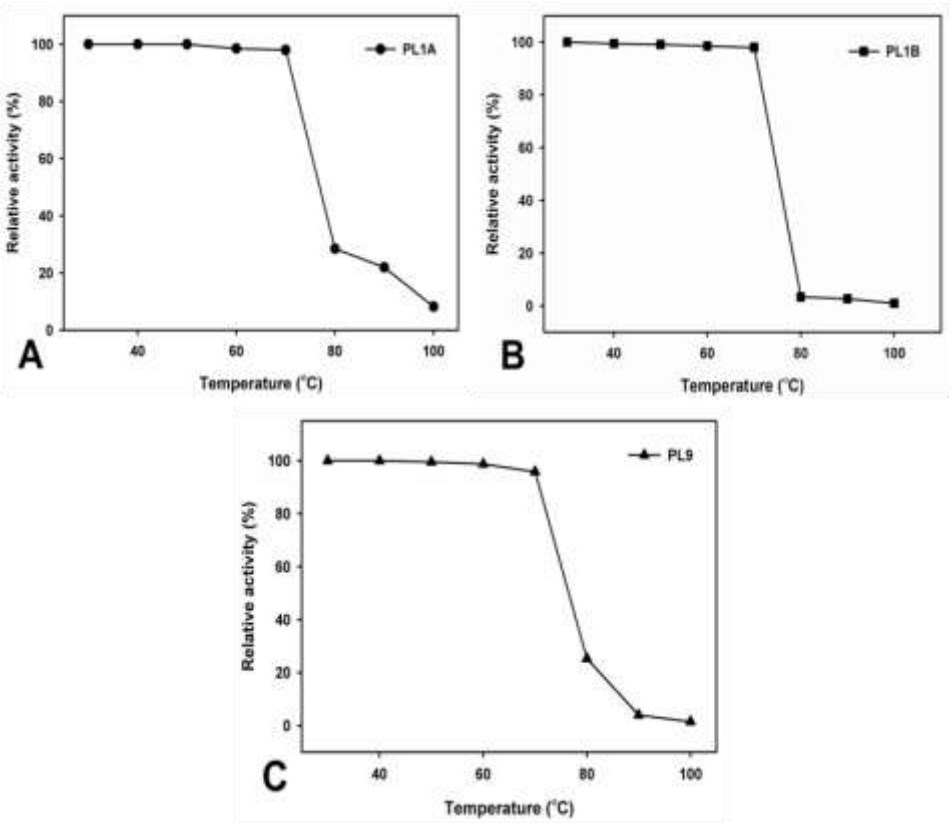


Figure 6.7| Thermostability of (A) PL1A; (B) PL1B; (C) PL9 towards PGA, after 30 min of incubation of the enzyme at different temperatures.



All the three enzymes showed an exclusive requirement of Ca^{2+} ions to achieve their maximum activity. PL1A and PL1B showed only 20% of their maximum activity in the absence of 5 mM and 0.6 mM Ca^{2+} ions, respectively, whereas PL9 showed 50% of its maximum activity in absence of 5 mM Ca^{2+} ions (Figure 6.8A, B and C). The optimum Ca^{2+} ion concentration required to achieve 100% pectinolytic relative activity were 5 mM for both PL1A and PL9, whereas 0.6 mM for PL1B.

Kinetic parameters of these three enzymes were determined against PGA and are presented in Table 6.2. The data revealed that PL1A, PL1B and PL9, showed turnover number values (K_{cat}) of 1.3, 1.76, and 1.32 min^{-1} respectively (Table 6.2). The catalytic efficiency (K_{cat}/K_m) values exhibited by PL1A, PL1B and PL9 were 41, 62 and 35 $mM^{-1}min^{-1}$ respectively, revealing that PL1B exhibit higher catalytic efficiency on PGA, than PL1A and PL9.

Figure 6.8| Effect of concentration of Ca^{2+} ions on the activity of (A) PL1A; (B) PL1B; (C) PL9 against PGA as substrate.

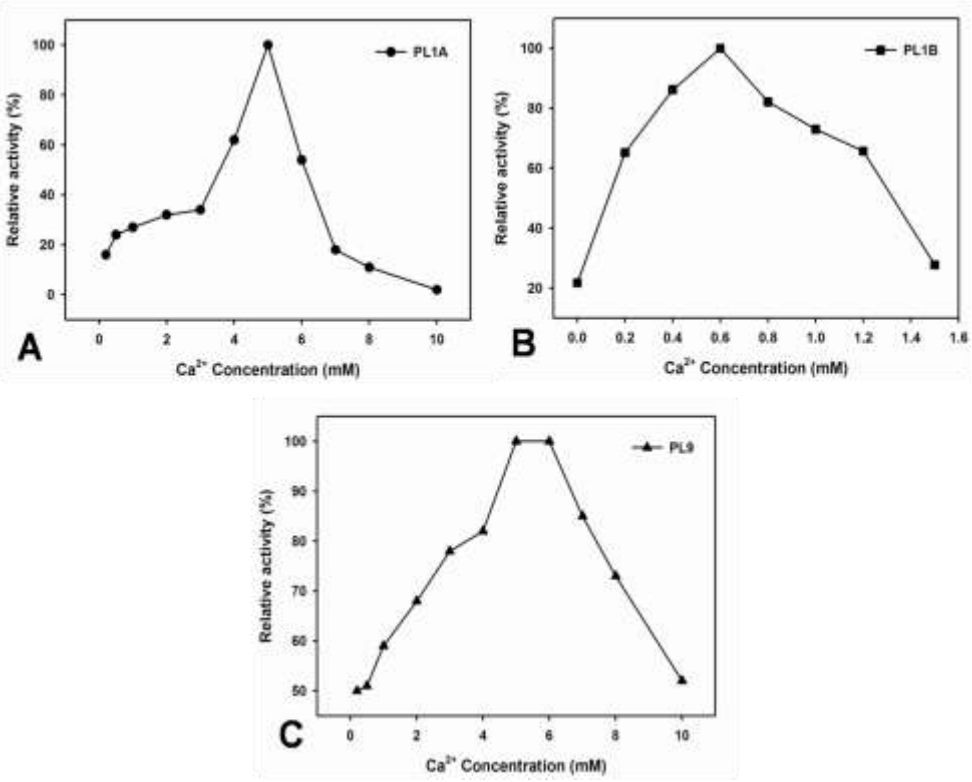


Table 6.2| Kinetic parameters of PL1A, PL1B and PL9 with Polygalacturonic Acid (PGA) from citrus. One unit of enzymatic activity (U) was defined as the amount of enzyme in mg that produces 1 mmol/L of unsaturated product per minute.

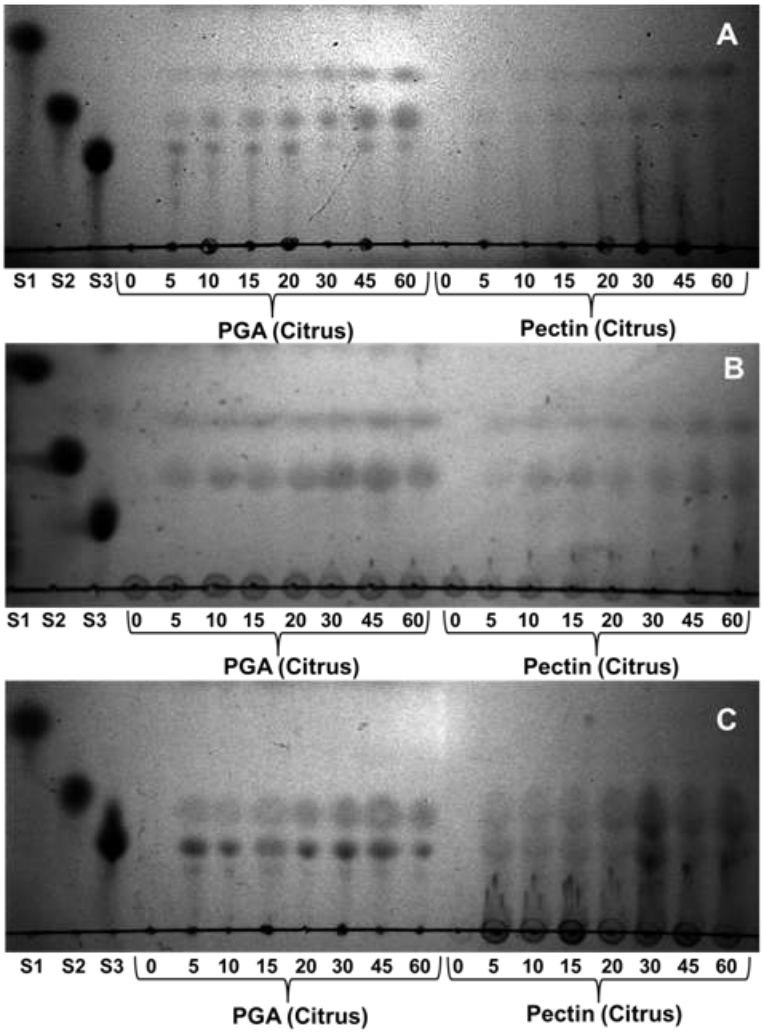
Enzyme	Substrate	Kcat (min^{-1})	Km (mM)	K_{cat}/K_m (mM^{-1} min^{-1})
PL1A	PGA (citrus)	1.3±0.03	0.0313±0.0005	41±0.23
PL1B	PGA (citrus)	1.76±0.05	0.0286±0.0002	62±0.43
PL9	PGA (citrus)	1.32±0.05	0.0378±0.0004	35±0.4

6.1.3.4. Functional properties of recombinant PL1A, PL1B and PL9

Products released by the enzymatic cleavage of PL1A, PL1B and PL9 of PGA and pectin (citrus) were determined. The reactions were carried out under optimum conditions of pH and temperature for each individual enzyme as mentioned in Methods section. The samples from enzymatic reaction were collected at different time intervals of 0, 5, 10, 15, 20, 30, 45 and 60 min and separated through TLC. PL1A produced unsaturated di- and tri-galacturonates along with other oligosaccharides of higher size. The accumulation of unsaturated tri-galacturonates and higher size oligosaccharides was predominant after 5 min of the start of reaction (Figure 6.9A). It was evident from the TLC results that an increase in the amount of unsaturated di- and tri-galacturonates was found with increase in time and found to be

highest at 60 min of reaction time (Figure 6.9A). PL1B from the beginning of the reaction started producing unsaturated di and tri-galacturonates, and no higher size oligosaccharides were observed. Accumulation of unsaturated di- and tri-galacturonates increased with time and found to be highest at 60 min of the reaction (Figure 6.9B). PL9 produced unsaturated tri-galacturonates and oligosaccharides of higher size, and the accumulation of this two products were found to be constant from 5 to 60 min of the reaction time (Figure 6.9C). The cleavage pattern of these three enzymes suggests that they cleave within the poly-galacturonan main chain of PGA and pectin (citrus) thus following an endo cleaving pattern.

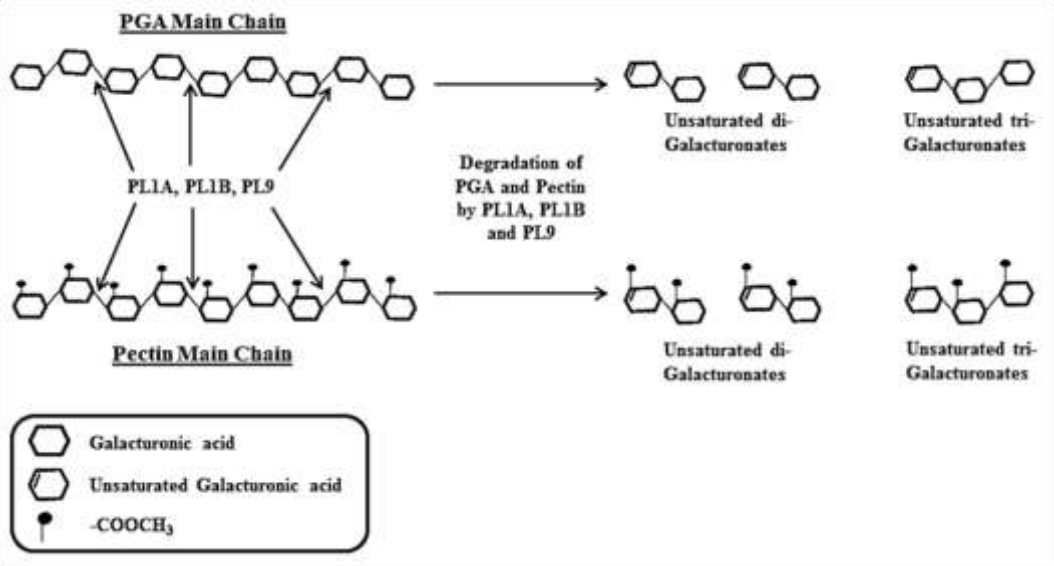
Figure 6.9| Thin layer chromatography (TLC) showing the enzymatic degradation products of PGA (citrus) and pectin (citrus) (25% methyl-esterified).



Chromatogram displaying hydrolysis by (**A**) PL1A (**B**) PL1B and (**C**) PL9 at 0, 5, 10, 15, 20, 30, 45 and 60 min. Standard oligosaccharides used were S1: D-galacturonic acid; S2: Di-galacturonic acid; S3: Tri-galacturonic acid.

The mechanism by which these three enzymes from *Clostridium thermocellum* cleaves the α-1,4 linkages in pectic polysaccharides thus resulting in enzymatic degradation is explained schematically in Figure 6.10.

Figure 6.10| Schematic presentation of mode of action of PL1A, PL1B and PL9 against PGA and pectin (citrus)



Hydrolysis of PGA and pectin (citrus) by PL1A, PL1B and PL9 leading to production of corresponding unsaturated oligo-galacturonates.

6.1.4. Discussion

Clostridium thermocellum is known to hydrolyze not only cellulose but also hemicelluloses (Zverlov *et al*., 1994; Halstead *et al*., 1999; Fernandes *et al*., 1999; Blum *et al*., 2000; Fontes & Gilbert, 2010; Zverlov, Fuchs, & Schwarz, 2002). It was also shown that *C. thermocellum* could utilize polygalacturonic acid and pectins as carbon sources (Spinnler, Lavigne, & Blachere, 1986). The results described in this paper demonstrate that *C. thermocellum* cellulosome is composed of enzymes that are able to attack pectin and can degrade these complex polysaccharides. We have identified and characterized for the first time three cellulosomal pectinolytic enzymes PL1A, PL1B and PL9 from this microorganism. The data revealed that PL1A, PL1B and PL9 have catalytic activity on polygalacturonic acid (PGA) and pectin (citrus). Sequence similarity studies with proteins in biological databanks placed PL1A, PL1B and PL9 in families 1 and 9 of pectate lyases (PL), respectively. Up to the present date there are 23 families of polysaccharide lyases, of which PL families 1, 2, 3, 9 and 10 contain pectate lyases (http://www.cazy.org).

Analysis of primary sequences of the cellulosomal enzymes under analysis here revealed a modular organization with the presence of a CBM, which is rare in pectinases (Pagès *et al*., 2003). It is generally believed that pectins are more accessible to enzyme attack than cellulose and hemicelluloses, and as such it has been assumed that there has been less evolutionary pressure for pectinases to contain CBMs. However, Rgl11A and Pel10A from *P. cellulosa* and Pel4A from *Clostridium cellulovorans* are examples of prokaryotic pectinases that contain a cellulose-binding domain (Pagès *et al*., 2003; McKie *et al*., 2001; Brown *et al*.,

2001). This report showed that CBMs are prevalent within cellulosomal pectinases and might be involved in potentiating the degradation of less recalcitrant substrates. Previous studies showed that functionally active family 6 and 35 CBM's bind strongly to cellulose (Henshaw *et al*., 2004; Bolam *et al*., 2004). Family 6 CBMs display considerable promiscuity in ligand binding with different modules showing affinity for amorphous cellulose, xylans and β-glucans (Czjzek *et al*., 2001). In addition, CBM family 35 also reveals considerable plasticity in ligand recognition which is not surprising considering that this family shares sequence similarities with CBM 6. Both these families, CBM 6 and CBM 35, are structurally related to the β-jelly-roll CBM superfamily (Boraston *et al*., 2004) and can be viewed as a subfamily of the large β-jelly-roll CBM superfamily (Tunnicliffe *et al*., 2005). CBMs are prevalent in plant cell wall degrading enzymes and as a general function promote the interaction of the enzyme with their target substrate (Boraston *et al*., 2004). PL1A contains a CBM 6, while PL1B and PL9 exhibit a CBM 35. The presence of CBMs in the structure of PL1A, PL1B and PL9 suggests that they are important in increasing their catalytic efficiency by bringing the enzymes into close proximity to their target substrates. However, in a recent work Montanier and colleagues (Montanier *et al*., 2009) while analysing the biological role of 4 members of family CBM35, revealed that the biological role of CBM35s is not dictated solely by the substrate specificity of their appended catalytic domains as members of these CBM family may recognize the products of pectin hydrolysis. Structurally, PL1A, PL1B and PL9 consist of an individual dockerin-containing enzyme integrated into the *C. thermocellum* cellulosome by CipA cohesin-dockerin interaction with a non-catalytic module CBM-like.

PL1A, PL1B and PL9 are characteristic pectate lyases and preferentially degrade polygalacturonic acid, though they also act on pectins. Moreover, all the three enzymes displayed significantly higher activity with 55% and 85% methyl-esterified pectin (citrus). Similar high activity of pectate lyase on pectins with high degree of methyl-esterification has been previously reported only from *Bacillus subtillis* (Soriano *et al*., 2006). PL1A and PL9 showed significant activity with rhamnogalacturonan from potato (RGAP) and soybean (RGAS), as compared with PL1B. The enzymes that degrade the backbones of pectic substances utilize two distinct cleavage mechanisms, the hydrolysis or the β-elimination. The method used to evaluate the PL1A, PL1B and PL9 activities provide evidence that these enzymes are lyases, as they catalyse the β-eliminative cleavage of glycosidic bonds with the production of Δ4,5 unsaturated galacturonates, which can be followed spectrophotometrically at absorbances of 232 to 235 nm. PL1A and PL9 was active within the pH range of 6.5-9.5 with highest activity at pH 8.5, which is similar to those of Rgl11Y from *C. cellulolyticum* cellulosome (pH 8.5) and from *C. cellulovorans* cellulosome Pel4A (pH 8.0) (Pagès *et al*., 2003; Tamaru & Doi, 2001), but PL1B was active within the pH range of 8-10 and showed highest activity at pH 9.8 similar to PelA from *Bacillus sp*. which showed pH optima of 10 (Soriano *et al*. 2000). All enzymes have a requirement of Ca^{2+} ions to achieve their maximum

activity. It has been suggested that the pH value of some plant tissues changes during microbial attack which possibly indicate that the degradation of plant cell wall polysaccharides occurs sequentially according to the pH of plant tissues (Pagès *et al*., 2003). It is also known that pectate lyases require Ca^{2+} for *in vitro* activity and presumably utilize the abundant Ca^{2+} in the plant cell wall for *in vivo* activity (Barras *et al*., 1994; Herron *et al*., 2003).

Analysis of the degradation products of cellulosomal PLs by TLC conclusively inferred that PL1A, PL1B and PL9 followed an endo cleavage pattern on PGA and pectin (citrus), cleaving these substrates endolytically as was previously reported for PelA from *Clostridium cellulovorans* (Tamaru & Doi, 2001). These enzymes produced unsaturated di, tri and higher oligogalacturonates from PGA and pectin (citrus). PelC from *B. subtilis* (Soriano *et al*., 2006), also an endo pectate lyase, showed a similar cleavage pattern producing mixtures of different degradation products, whereas PelX from *Erwinia chysanthemi* an exo-pectate lyase always produced a single degradation product either unsaturated di or tri-galacturonates (Shevchik *et al*., 1999). Hence, PL1A, PL1B and PL9 under investigation are conclusively endo pectate lyases. Pectic substrates are highly heterogeneous which may require many enzymes with different specificities and catalytic mechanisms for their complete breakdown. Therefore, the ability of these cellulosomal enzymes to degrade pectic substances suggests that cellulosomes are designed for the degradation of an entire set of carbohydrates within plant cell walls, and not only cellulose and hemicellulose. It is clear that within cellulosomes other enzymes presently of unknown function may target the degradation of pectic polysaccharides.

6.1.5. Conclusion

Thermostable enzymes are important resources in various industrial processes that occur at higher temperatures. Hence enzymes described in this study will be competent enough for industrial processes like fruit juice extraction, vegetable and fruit maceration or bioscouring of cotton fabric to increase the efficiency of dying at improved temperatures. These enzymes can be used as a cocktail for further efficient and complete degradation of pectic polysaccharides.

7. β-GLUCANASES FOR ANIMAL FEED SUPPLEMENTATION

7.1. β-1,3-1,4-glucanases and not β-1,4-glucanases improve the nutritive value of barley-based diets for broilers

Vania O. Fernandes[1,2*], Mónica Costa[1*], Teresa Ribeiro[1], Luís Serrano[1], Vânia Cardoso[1,2], Helena Santos[1], Madalena Lordelo[3], Luís M.A. Ferreira[1,2] and Carlos M.G.A. Fontes[1,2]

[1] CIISA-Faculdade de Medicina Veterinária, Universidade de Lisboa, Avenida da Universidade Técnica, 1300-477, Lisboa, Portugal; [2] NZYTech genes & enzymes, Campus do Lumiar, Estrada do paço do Lumiar, 1649-038 Lisboa, Portugal; [3] Instituto Superior de Agronomia, Tapada da Ajuda, 1349-017 Lisboa, Portugal.
*Equal contribution

Adapted from a submitted manuscript

Abstract

Barley-based diets contain a significant proportion of highly soluble β-1,3-1,4-glucans that are highly anti-nutritive for monogastric animals, in particular for poultry. Cleavage of mixed linked glucans by the addition of exogenous enzymes leads to a significant reduction in the degree of polymerization of the polysaccharide chains and a consequent reduction of digesta viscosity. Hydrolysis of β-1,3-1,4-glucans can result from the action of strictly specific β-1,3-1,4-glucanases (EC 3.2.1.6) or β-1,4-glucanases (EC 3.2.1.4), generally termed endo-cellulases that cleave the β-1,4-linkages of the β-1,3-1,4-glucan chains. Here we evaluated the capacity of two *Clostridium thermocellum* enzymes, β-1,3-1,4-glucanase 16A, termed *Ct*Glc16A, and β-1,4-glucanase 8A, termed *Ct*Cel8A, to improve the nutritive value of barley based diets for broilers. The data revealed that *Ct*Glc16A improves the performance of broilers fed a highly viscous barley-based diet. In contrast, although remaining active and retaining its molecular integrity during passage through the GI tract, *Ct*Cel8A was unable to affect the nutritive value of the cereal based diet. *In vitro* studies revealed that both *Ct*Glc16A and *Ct*Cel8A are equally effective in reducing the viscosity of a pure β-1,3-1,4-glucan solution. However, the data demonstrate that the capacity of *Ct*Cel8A to cleave β-1,3-1,4-glucans is significantly affected by the presence of cellulosic substrates. Taken together the results suggest that while active against β-1,3-1,4-glucans, *in vivo* β-1,4-glucanases tend to act preferentially on cellulosic substrates and not on mixed linked glucans. Thus, a significant optimization of current enzyme mixtures used in poultry nutrition could derive by estimating the contribution of β-1,4-glucanases to the overall β-1,3-1,4-glucanase activity expressed by the commercial enzyme mixtures.

7.1.1. Introduction

Industrial poultry diets are based in cereals. However, these diets may contain high levels of soluble non-starch polysaccharides (NSPs) that significantly affect the efficiency of the digestive process, impairing animal performance (Smits & Annison, 1996; Choct, 1997). Barley incorporation in poultry diets is limited by its high content in NSP, mainly soluble β-1,3-1,4-glucans, which upon solubilization lead to an increase in digesta viscosity. An increase in digesta viscosity slows digesta passage rate and affects the interaction of the endogenous digestive enzymes with their target substrates (Smits & Annison, 1996; Pettersson & Aman, 1989; Jozefiak *et al*., 2006). In addition, prolonged digesta passage rates promotes a modification in gut physiology such as the enlargement of gastrointestinal organs, which are detrimental to final carcass yields (Smits & Annison, 1996). High digesta viscosity also favors the proliferation of anaerobic microbes in the upper parts of the gastrointestinal (GI) tract, which affects animal's health (Jozefiak *et al*., 2006). All these negative effects may lead to a decrease in feed intake as well as a reduction in nutrient digestibility, which consequently promotes a decrease in birds productivity (Pettersson & Aman, 1989). To counter the negative effects related with the solubilization of barley β-1,3-1,4-glucans, commercial enzyme mixtures expressing cellulase and hemicellulase activities are widely used to supplement broiler diets (Bedford & Morgan, 1996). Enzyme supplementation of barley based diets for poultry results in the hydrolysis of β-1,3-1,4-glucans allowing for a reduction in the degree of polymerization of the polysaccharide leading to a reduction in digesta viscosity, an improvement of nutrients digestibility and feed intake (Bedford & Morgan, 1996; Choct, 2006). In addition, it is now well established that broilers fed barley-based diets display a maximal response to enzyme supplementation at the early stages of the production cycle (Newman & Newman, 1988; Rotter *et al*., 1988; Nahas & Lefrancois, 2001) when the young chick has a poorly developed digestive system. Hence, the production of endogenous digestive enzymes at an early stage of growth is scarce and may hinder feed digestion (Nitsan *et al*., 1991; Dunnington & Siegel, 1995; Kirjavainen & Gibson, 1995). Therefore, by effectively contributing to reduce digesta viscosity, exogenous plant cell wall hydrolases contribute to improve the effectiveness of endogenous enzymes enhancing the animal's digestive capacity, particularly when the raw-materials used in animals' diets are highly prone to display higher viscosities, such as barley.

Exogenous polysaccharidases used to supplement poultry diets are usually composed of enzyme mixtures displaying a large range of polysaccharide specificities. However, enzymes required to improve the nutritive value of barley based diets for poultry need to display β-1,3-1,4-glucanase activity in order to depolymerize the anti-nutritive mixed linked glucans. β-1,3-1,4-glucanases are Glycoside Hydrolases (GH; EC. 3.2.1). GH are classified in families based in primary sequence homology (Henrissat, 1991). Presently there are 133 families of GHs in the constantly updated Carbohydrate Active enZyme (CAZy) database

(www.cazy.org) (data collected on February 2015). Enzymes expressing strict specificity for β-1,3-1,4-glucan linkages (EC 3.2.1.6), also generically termed as β-glucanases, are currently grouped in GHs families 9 and 16. In contrast, enzymes that participate in the cleavage of β-1,4-glucan linkages (EC 3.2.1.4), which are generally referred as cellulases, belong to GHs families 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 74 and 124. Barley β –glucan is the major cell wall polysaccharide in endosperm cells of barley and is constituted by mixed linkages of β-1,3- and β-1,4 glucosidic bonds, generating a linear polysaccharide more soluble than cellulose (Xue *et al*., 2003). Barley β-glucan is composed mainly of β-1,4 linkages, with a ratio β-1,3 to β-1,4 linkages of 1:2.5 approximately (Xue *et al*., 2003; Jamar, Jardin, & Fauconnier, 2011). Thus mixed linked glucans may be cleaved by β-1,4-glucanases, which are enzymes that display a broader substrate specificity as they can degrade both β-1,4-glucans and β-1,3-1,4-glucans, or by β-1,3-1,4-glucanases that strictly cleave β-1,3-1,4-glucans and have no activity on β-1,4-glucans. Thus, β-1,3-1,4-glucanases do not hydrolyze cellulose, while β-1,4-glucanases (cellulases) are able to degrade cellulosic substrates but also the more soluble β-1,3-1,4-glucans by acting at the more abundant β-1,4-glucan regions (Xue *et al*., 2003). Enzyme mixtures currently used to supplement barley-based diets are evaluated by their β-1,3-1,4-glucanase activity. However, it is usually unknown which GHs contribute to the total β-1,3-1,4-glucanase activity; if the highly specific β-1,3-1,4-glucanases or the general β-1,4-glucan cutters generally referred as cellulases. Thus, it remains to be established which enzymes contribute most to improve the nutritive value of barley-based diets for poultry, if β-1,3-1,4-glucanases or β-1,4-glucanases.

Here, we have compared the capacity of a highly specific β-1,3-1,4-glucanase and a typical endo-acting β-1,4-glucanase to improve the nutritive value of a barley-based diet for broilers. The enzymes were selected from the anaerobic thermophilic bacterium *Clostridium thermocellum* and consisted in the family 16 β-1,3-1,4-glucanase A (Ribeiro *et al*., 2012), *Ct*Glc16A, and the family 8 β-1,4-glucanase A, *Ct*Cel8A, also known as cellulase 8A (Cornet *et al*., 1983). The two enzymes were recombinantly expressed in *Escherichia coli*, purified to become free of other side activities, and used to supplement a barley based diet for broilers.

7.1.2. Materials and Methods

7.1.2.1. Gene isolation and cloning into prokaryotic expression vector

The thermostable β-1,3-1,4-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 encoding mature *Ct*Glc16A (residues 31-251) was amplified from *C. thermocellum* genomic DNA through PCR using the NZYProof DNA polymerase (NZYTech, genes & enzymes, Portugal) and the following primers: 5´-CTC**GCTAGC**ACTGTGGTAAATACGCC -3´ and 5´-CAC**CTCGAG**ATTATCTTGCGGAACAC -3´ (NZYTech Ltd., Portugal). Primers included

engineered *Nhe*I and *Xho*I restriction sites (in bold) that allowed the subsequent cloning of the resulting nucleic acid into similarly restricted pET21a, generating the plasmid pGH16. The thermostable β-1,4-glucanase of *C. thermocellum*, termed *Ct*Cel8A, is also a modular enzyme containing an N-terminal glycoside hydrolase family 8 catalytic domain and a C-terminal dockerin (Cornet *et al*., 1983). The gene encoding *Ct*Cel8A (residues 396-1734) was amplified from *C. thermocellum* genomic DNA through PCR using the NZYProof DNA polymerase (NZYTech, genes & enzymes, Portugal) and the following primers: 5´-CTC**CATATG**GCAGGTGTGCCTTTTAAC -3´ and 5´-CAC**GGATCC**CTAATAAGGTAGGTGGGG -3´. Primers included *Nde*I and *BamH*I engineered and restriction sites (in bold) that allowed the subsequent cloning of the resulting nucleic acid into similarly restricted pET21a, generating the plasmid pGH8. Recombinant *Ct*Glc16A and *Ct*Cel8A contained an engineered C-terminal His6 tag allowing the direct purification of the two polypeptides by immobilized metal affinity chromatography (IMAC).

7.1.2.2. Expression and purification of *Ct*Glc16A and *Ct*Cel8A

To express the proteins, *E. coli* BL21(DE3) cells harboring the appropriate recombinant plasmid, pGH16 or pGH8, were cultured in LB medium containing ampicillin (100 μg ml^{-1}) at 37 ºC to mid-exponential phase (A_{600} 0.4). At this point, isopropyl 1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 1 mM and the cultures were further incubated at 19 ºC during 16 hours. Recombinant cells were harvested at 5,000 ×g and the resulting pellet was resuspended in 10 mM Imidazol buffer (10 mM Imidazol, 50 mM NaHepes, 1 M NaCl, 5 mM $CaCl_2$, at pH=7.5), submitted to ultrasonication and centrifuged at 4 ºC, 17000 ×g during 30 min. The obtained His_6- 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 analyzed by SDS-PAGE (Laemmli, 1970).

7.1.2.3. Biochemical properties of *Ct*Glc16A and *Ct*Cel8A

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. Reactions were stopped by adding a DNSA based solution (1% DNSA, 1% NaOH and 0.2% phenol) following the method described by Miller (1959). To explore the pH profile of *Ct*Glc16A and *Ct*Cel8A 50 mM MES (2-N-morpholino- ethanesulfonic acid) (pH 4.5-7), 50 mM Tris-HCL (pH 7-9.5), and 50 mM $NaHCO_3$ (pH 9.5-11) buffers were used in enzyme assays employing 0.25% (w/v) barley β-glucan as the substrate. For thermostability experiments, the two proteins were incubated at temperatures ranging from 50 to 90 ºC. After

20 min at the required temperature, samples were withdrawn and residual activity was determined at 50 ºC and 45 ºC, for *Ct*Glc16A and *Ct*Cel8A, respectively, by measuring the amount of reducing sugar released from barley β-glucan as described above. Determination of temperature of maximal enzyme activity was performed by incubating the enzyme at temperatures ranging from 50 to 90 ºC and measuring reducing sugar release as previously described. One unit of catalytic activity is defined as the amount of enzyme required to release 1 µmol of product per min. Resistance to proteolysis was tested essentially as described previously (Fontes *et al*., 1995a) by incubating the proteins with porcine pancreatine (Sigma P-1500) at 37 ºC and measuring residual enzyme activity as described for thermostability experiments. In addition the molecular integrity of the enzymes was accessed by SDS-PAGE analysis with 14% polyacrylamide separation gels. The capacity of the two recombinant enzymes to decrease the viscosity of barley β-glucan was evaluated *in vitro*. A solution of 1.5% (w/v) of barley β-glucan (Megazyme, Ireland) was prepared in PC buffer. Precisely 111 µL of each enzymatic extract, which supplied 1400 U/ kg of enzyme activity was added to the barley β-glucan preparation. Viscosity was measured at 6 x*g* using a viscometer (Model LVDVCP-II, Brookfield Engineering Laboratories, Middleboro, MA) with a cup maintained at 37 ºC, at time 0 and then at each minute after enzyme addition until levels of viscosity were stabilized. After stabilization, viscosity was measured until min 90 every 15 minutes. At the same time, as a control reaction, the viscosity of barley β-glucan without addition of enzymes was measured over time as described above. The capacity of *Ct*Glc16A and *Ct*Cel8A to depolymerise barley β-glucan in the presence of feed was investigated by using the commercial Azo-barley glucan method (Megayme, Ireland) with some adaptations. The both enzymes were incubated individually with 0.15 g of a barley based diet feed sample (reaction) or with 0.15 mL of PC buffer (control) during 10 minutes. A volume of 0.25 mL of azo-barley glucan was then added to the previous preparations and incubated at 37 ºC during 20 minutes. Reactions were stopped and β-1,3-1,4 glucanase activity was measured following manufacturer procedures.

7.1.2.4. Feed analysis for β-glucanase activity

In a previous study it was shown that levels of barley's endogenous β-glucanase activity affect the efficacy of exogenous enzymes used to improve the nutritive value of barley-based diets for poultry (Ribeiro *et al*., 2011). To select a barley batch expressing lower levels of endogenous β-glucanases, five different barley lots available commercially were analyzed. The barley lots were milled at 0.5 mm and enzyme extraction was performed by adding 1 mL of PC buffer into 0.75 g of barley. The supernatant was recovered and used to quantify the β-1,3-1,4-glucanase activity using a β-glucanase commercial assay kit (Megazyme, Ireland), following the manufacturer protocol. The incubation period was extended to 3 hours. The

barley lot expressing the lowest levels of endogenous β-1,3-1,4-glucanase activity (52.5 U/ kg of barley) was used as the main component of the basal diet described below.

7.1.2.5. Enzyme feed incorporation

In order to standardize the number of β-1,3-1,4-glucanase units used to supplement the barley-based diets, the catalytic activity of the recombinant enzymes against barley β-glucan, including the commercial enzyme mixture (Rovabio™ Excel AP ®), were determined in parallel at 40 ºC by measuring the release of reducing sugars as described above (Fontes *et al*., 2000). The extract containing the commercial enzyme was prepared by ressuspending 250 mg of the enzyme in 10 ml of PC buffer, followed by an overnight incubation at room temperature with gentle agitation and centrifugation at 16000 ×g for 10 min. The two recombinant enzymes and the commercial enzyme mixture were incorporated in the feed at a concentration of 1400 U/ kg feed.

7.1.2.6. Animals and diets

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 86/609/EEC). One hundred and twenty 1-d-old Ross 308 male broiler birds were assigned to 40 pens of 3 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 β-1,3-1,4-specific glucanase *Ct*Glc16A (treatment *Ct*Glc16A) or the broadly specific β-1,4-glucanase *Ct*Cel8A (treatment *Ct*Cel8A). The trial duration was 35 days. As stated above, 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 *Penicillium funiculosum* β-1,4-xylanase (EC 3.2.1.8) and β-1,3-1,4- glucanase (EC 3.2.1.6). The basal diet, which composition was displayed in Table 7.1, contained 618 g/kg of barley and was formulated to ensure a nutrient availability as defined by the NRC (N.R.C, 1994). 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 viscosity as described above.

Table 7.1| Ingredient composition and calculated nutrient content.

Ingredients	g/ kg
Barley	615
Soybean meal 47%	294
Soybean oil	59
Salt	2.5
Calcium carbonate	8.4
Dicalcium phosphate 18%	17.4
DL-Methionine	1.7
Mineral and vitamin premix[1]	2
Nutrient content	
Energy (MJ ME/kg DM)	12.1
Crude protein	208
Ether extract	75
Crude cellulose	51
Methionine	4.69
Methionine + Cysteine	8.20
Lysine	10.9
Threonine	7.58
Calcium	9.20
Available phosphorus	4.10

[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

7.1.2.7. Analytical procedures performed in digesta samples

To evaluate the levels of β-1,3-1,4-glucanase activity present in the four animal diets and in digesta samples collected in different parts of the broilers´ GI tract, samples were initially centrifuged at 16000 ×g for 5 min and the supernatants were recovered for analysis. The barley and feed samples (0.75 g) were previously mixed in 1 mL of PC buffer. The mixture samples were subjected to vigorous stirring during 30 min, centrifuged and the supernatant was analyzed. Qualitative analysis of β-1,3-1,4-glucanase activity 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). Catalytic activity was detected based in the method described by

Ponte *et al*. (2004) with some modifications. Briefly, a Pasteur pipette was used to create wells in the agar plates, which were filled with 20 µL of enzyme extract from digesta samples. Subsequently, the plates were incubated for 16 h at 37 °C and dyed with 1% Congo Red (E.Merck AG®) in 10 mM Tris-HCl (pH=8) for 30 min. To observe regions where the polysaccharides were degraded, the plates were subjected to three 15 min washes with 1 M NaCl in 10 mM Tris-HCl (pH=8). For measuring the viscosity of small intestine contents, samples collected from the duodenum plus jejunum and ileum were centrifuged for 10 min at 10867 ×*g* and the viscosity of sample supernatant was measured at 6 ×*g* using a viscometer (Model LVDVCP-II, Brookfield Engineering Laboratories, Middleboro, MA) with a cup maintained at 24 ºC. Zymogram analysis was performed as described by Fontes *et al*. (2004) and Ribeiro *et al*. (2011). Summarily, digesta proteins were separated through SDS-PAGE electrophoresis in 14% acrylamide gels containing 0.1% of barley β-glucan (Megazyme, Ireland), according to Laemmli (1970). After electrophoresis, polypeptides were renatured by subjecting the gel to five 30 min washes in 100 mM sodium succinate, pH 6.3, 10 mM $CaCl_2$ and 1 mM DTT. The gels were incubated for 16 h at 37 ºC in the same buffer and proteins were stained in a solution comprising 40% (v/v) methanol, 10% (v/v) glacial acetic acid and 0.4% (w/v) Coomassie Brilliant Blue R. After destaining, gels were washed in 0.1M Tris-HCl (pH=8) for 20 min. β-glucanase activity was detected using a 0.1% (w/v) Congo Red solution for 15 min and washing with 1M NaCl in 10mM Tris-HCl (pH=8) until excess dye was removed. After Congo Red staining the gels were counter stained with 1N HCl as described elsewhere (Ribeiro *et al*., 2011). Areas of catalytic activity appeared as colorless zones in a grey background.

7.1.2.8. Statistical Analyses

Data related to bird performance were subjected to ANOVA according to the general linear models procedure of SAS (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 3 animals (n=10). Unless otherwise stated, differences were considered significant when $P < 0.05$.

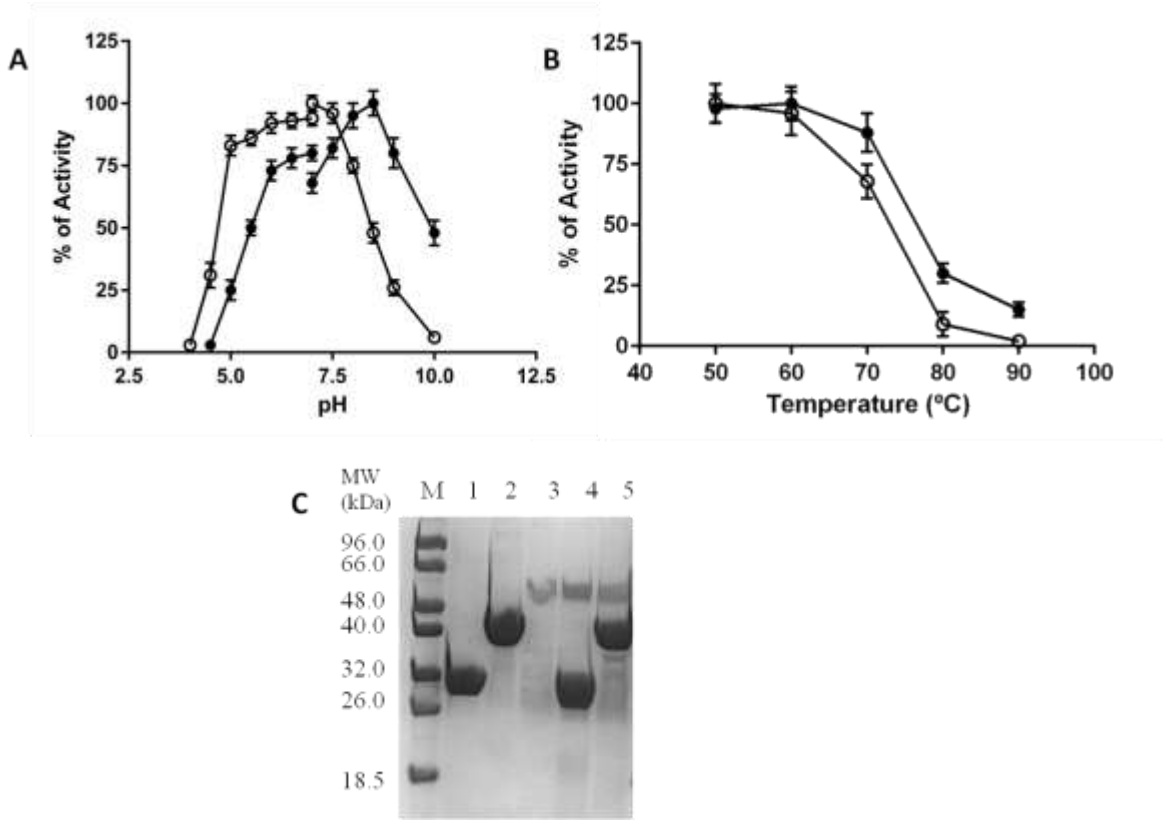
7.1.3. Results and Discussion

7.1.3.1. Biochemical properties of the *Ct*Glc16A and *Ct*Cel8A

Two recombinant enzymes from the anaerobic thermophilic bacterium *C. thermocellum* were selected for this study. *Ct*Glc16A is a family 16 GH with a restricted activity against β-1,3-1,4-glucans (Ribeiro *et al*., 2012). In contrast, *Ct*Cel8A is a typical endo-β-1,4-glucanase being capable of hydrolyzing not only cellulosic polysaccharides but also β-1,3-1,4-glucans through the cleavage of β-1,4-linkages (Cornet *et al*., 1983). To evaluate the capacity of *Ct*Glc16A

and *Ct*Cel8A to remain stable and active throughout passage in the broilers GI tract the biochemical properties of the two biocatalysts were compared. The data, presented in Figure 7.1, reveals that both enzymes display a similar pH and temperature profiles, which may reflect their common origin in *C. thermocellum*. The two enzymes display broad pH optima with *Ct*Glc16A more active on the neutral to alkaline range and *Ct*Cel8A on the neutral acidic range (Figure 7.1A). In addition, the two recombinant enzymes were shown to be completely resistance to proteolytic inactivation by pancreatic proteases. Thus, enzyme activity was not affected by incubation with a pancreatic extract and enzyme molecular integrity was retained after incubation with proteolytic enzymes (Figure 7.1C). Taken together these data confirms that both *Ct*Glc16A and *Ct*Cel8A express the required biochemical properties to actively contribute to the depolymerization of β-1,3-1,4-glucans in the conditions prevalent in the poultry GI tract.

Figure 7.1| pH (Panel A) and temperature (Panel B) profiles of *Ct*Glc16A (●) and *Ct*Cel8A (○) and resistance of the two enzymes against proteolytic inactivation.



In **Panel A**, these enzymes were incubated at standard conditions in MES, Tris or $NaHCO_3$ buffers expressing different pHs, and β-glucanase activity determined. In **Panel B**, *Ct*Glc16A and *Ct*Cel8A activities were determined with barley β-glucan at different temperatures. For thermostability, the enzymes were incubated for 30 min at different temperatures, and residual activity determined at 50 ºC or 45 ºC, for *Ct*Glc16A and *Ct*Cel8A, respectively. In **Panel C**, *Ct*Glc16A (lane 1) and *Ct*Cel8A (lane 2) were incubated for 30 min with a pancreatic protein extract and protein integrity evaluated by SDS-PAGE (lanes 4 and 5, for *Ct*Glc16A and *Ct*Cel8A, respectively). In lane 3 the pancreatic mixture of enzymes was analyzed. M represents the low molecular protein marker.

7.1.3.2. *Ct*Glc16A but not *Ct*Cel8A improve the nutritive value of barley-based diets for broilers

In this study we compared the capacity of a highly specific β-1,3-1,4-glucanase, *Ct*Glc16A (Laemmli, 1970), and a typical family 8 β-1,4-glucanase, *Ct*Cel8A (Cornet *et al*., 1983), to improve the nutritive value of barley-based diets for broilers. The both enzymes were recombinantly produced in *E. coli* and the enzymatic units of β-1,3-1,4-glucanase activity used to supplement the diets were measured such that they were identical and similar to the positive control enzyme mixture. Results concerning the productive parameters of broilers are presented in Table 7.2. Enzyme supplementation affected body weight of the animals since day 14 of the trial. At days 14 and 21 of the experiment, body weight of birds supplemented with the commercial enzyme mixture and recombinant *Ct*Glc16A were significantly higher than those of non-supplemented and *Ct*Cel8A supplemented birds, although animals of the *Ct*Glc16A and *Ct*Cel8A groups had similar body weights. At days 28 and 35 of the trial, birds supplemented with the commercial enzyme mixture had similar body weights to birds supplemented with *Ct*Glc16A and animals from these two groups had significantly higher body weights than non-supplemented or *Ct*Cel8A supplemented birds. Differences in body weight result from similar differences in weight gain for the period of 7-14 and 14-21 days of the trial. Feed intake was not affected by the addition of exogenous enzymes. In addition, feed conversion ratio (FCR) displayed a tendency to be globally different (0-35 days) among treatments, showing lower values ($P<0.1$) for animals supplemented with the commercial mixture and the *Ct*Glc16A. Taken together, the data suggest that the strict β-1,3-1,4-glucanase *Ct*Glc16A contributes to improve the nutritive value of the barley based diet for broilers and the recombinant β-1,4-glucanase *Ct*Cel8A although able to degrade β-1,3-1,4-glucans, is unable to significantly improve broiler performance.

The better body weight displayed by animals from PC and *Ct*Glc16A groups at early stages of growth confirms the importance of enzyme supplementation of barley-based diets for young broilers. Several reports suggested the importance of supplementing both the poor production of endogenous enzymes and the insufficient microbial flora in young birds (Newman & Newman, 1988; Nahas & Lefrancois, 2001; Dunnington & Siegel, 1995), emphasizing the importance of exogenous supplementation at this initial period of the productive cycle. A previous study have shown the importance of enzyme supplementation of barley-based diets mainly in the first 11 days of the broilers growth, suggesting that enzyme supplementation in the initial period of broilers production is crucial to improve digestion efficiency and that enzyme supplementation in later stages might not be important since the capacity of birds in producing endogenous enzymes are already established and can counter the antinutritive effects of the β-glucans until the end of the production cycle

(Cardoso *et al*., 2014). Although the feed intake was not considerable different among the four groups of animals, the data suggest that the animals supplemented with the commercial enzymatic mixture and those supplemented with *Ct*Glc16A revealed improved nutrients digestibility, which was reflected by the better body weight and FCR in these animals, mainly in the earlier periods of the production cycle.

Table 7.2| Growth performance parameters of broilers fed on a barley-based diet non-supplemented (NC) or supplemented with different exogenous β-glucanases. PC, a commercial enzyme mixture; *Ct*Glc16A, the recombinant β-1,3-1,4-glucanase; *Ct*Cel8A, the recombinant β-1,4-glucanase.

	NC	PC	*Ct*Glc16A	*Ct*Cel8A	SEM	p(*F*)
Body Weight (g)						
0d	41.2	41.1	40.6	40.7	0.381	0.6270
7d	121	123	115	120	3.19	0.4518
14d	280 ^{b}	325 ^{a}	313 ^{ac}	291 ^{bc}	10.8	0.0200
21d	631 ^{b}	732 ^{a}	693 ^{ac}	661 ^{bc}	19.2	0.0052
28d	1137 ^{b}	1252 ^{a}	1225 ^{a}	1161 ^{b}	26.3	0.0189
35d	1794 ^{b}	1877 ^{a}	1901 ^{a}	1753 ^{b}	41.0	0.0490
Weight Gain (g)						
0-7d	80	82	74	79	3.038	0.3498
7-14d	159 ^{b}	203 ^{a}	197 ^{a}	171 ^{b}	8.67	0.0023
14-21d	352 ^{b}	412 ^{a}	390 ^{ac}	370 ^{bc}	11.2	0.0035
21-28d	505	530	501	500	15.2	0.4700
28-35d	657	626	676	592	24.7	0.3606
0-35d	1753	1836	1860	1712	41.5	0.3310
Feed Intake (g)						
0-7d	98	100	93	97	4.45	0.6791
7-14d	233	252	267	243	15.9	0.4878
14-21d	523	573	552	525	16.2	0.1120
21-28d	815	816	803	798	16.8	0.8282
28-35d	1105	1013	1037	1020	32.7	0.1912
0-35d	2774	2755	2751	2682	57.7	0.6972
Feed Conversion Ratio						
0-7d	1.24	1.24	1.26	1.23	0.050	0.9900
7-14d	1.47	1.28	1.36	1.43	0.069	0.2100
14-21d	1.49 ^{a}	1.39 ^{b}	1.42 ^{b}	1.42 ^{b}	0.028	0.1000
21-28d	1.62	1.55	1.62	1.60	0.034	0.4081
28-35d	1.73	1.77	1.72	1.76	0.052	0.9263
0-35d	1.58 ^{a}	1.52 ^{b}	1.55 ^{ab}	1.57 ^{a}	0.019	0.0814

7.1.3.3. Exogenous enzymes affect the dimensions of broilers GI tract

The effects of exogenous enzyme supplementation in the length and weight of birds' GI compartments were measured and the data is presented in Table 7.3. Regarding to relative weight of the GI compartments, significant differences were found only at ileum weights. Ileum was lighter in animals supplemented with the commercial mixture or *Ct*Glc16A, and heavier in non-supplemented or *Ct*Cel8A supplemented animals. Differences were more pronounced in the relative length of GI compartments. Thus, duodenum was found to be shorter in the animals supplemented with the commercial enzyme mixture while was longer in the birds supplemented with the *Ct*Cel8A. Non-supplemented and *Ct*Glc16A supplemented animals displayed similar duodenum lengths. In contrast, the length of jejunum and ileum was significantly smaller for animals supplemented with *Ct*Glc16A and the commercial enzyme mixture and longer for non-supplemented or *Ct*Cel8A supplemented animals.

Table 7.3| Relative length and weight of GI tract compartments and viscosity of digesta samples of broilers fed on a barley-based diet non-supplemented (NC) or supplemented with different exogenous β-glucanases. PC, a commercial enzyme mixture; *Ct*Glc16A, the recombinant β-1,3-1,4-glucanase; *Ct*Cel8A, the recombinant β-1,4-glucanase.

	NC	PC	*Ct*Glc16A	*Ct*Cel8A	SEM	p(F)
Relative Weight (g/kg BW)						
Crop	3.26	2.67	2.78	3.16	0.312	0.4762
Gizzard	9.3	10.2	11.2	10.4	0.612	0.2079
Liver	22.6	23.4	22.9	23.7	0.791	0.7558
Duodenum	6.68	6.38	6.51	7.34	0.307	0.1306
Jejunum	14.9	12.6	13.6	13.6	0.674	0.1424
Ileum	11.4 ^{a}	9.56 ^{b}	9.40 ^{b}	11.5 ^{a}	0.581	0.0162
Caecum	2.44	2.66	3.04	2.43	0.168	0.0517
Relative Length (cm/kg BW)						
Duodenum	16.3 ^{ab}	15.2 ^{b}	16.3 ^{ab}	17.1 ^{a}	0.456	0.0500
Jejunum	42.1 ^{a}	37.6 ^{b}	36.9 ^{b}	40.1 ^{a}	0.898	0.0008
Ileum	43.3 ^{a}	38.7 ^{b}	38.3 ^{b}	42.3 ^{a}	1.261	0.0169
Caecum	9.8	9.7	10.2	10.0	0.394	0.8227
Content Viscosity (cP)						
Duodenum + Jejunum	8.79 ^{a}	5.84 ^{b}	5.57 ^{b}	6.14 ^{b}	0.410	0.0001
Ileum	12.9 ^{a}	6.85 ^{b}	7.57 ^{b}	7.26 ^{b}	0.959	0.0001

Several studies reported a decrease in weight and length of GI compartments of birds supplemented with exogenous cellulases and hemicellulases. This effect is believed to be related with the decrease in digesta viscosity, which will increase feed passage rate and decrease the physical aggression of the digestive lumen derived from a higher viscous

digesta. Interestingly the size of the caecum tended to be higher (P<0.1) in animals supplemented with the *Ct*Glc16A recombinant enzyme. Mathlouthi *et al*. (2002) also found improved gut morphology when supplementing a rye-based diet with a xylanase. In addition, Adeola and Cowieson (2011) also referred a possible increment in caecum dimensions as a consequence of enzyme supplementation. Enlargement of the last compartment of the birds GI tract was suggested to be related with an increase in the production of short chain fatty acids (SCFA) usually associated with diets supplemented with prebiotics (Van der Wielen *et al*., 2000; Campbell, Fahey, & Wolf, 1997) or diets supplemented with enzymes that promote the release of fermentable oligosaccharides in the small intestine (Adeola & Cowieson, 2011). This increase in the SCFA in the caecum is related with the increase of the total anaerobic microorganisms and promotes the decrease in caecum pH. In addition, exogenous enzymes have been reported to modulate the gut microflora (Bedford & Cowieson, 2009) and to increase the proportion of lactic acid produced. Moreover, Jamroz *et al*. (1996) referred an increase in caecum fermentation when high incorporation (63%) of barley in diet was performed. Further studies aiming at profiling gut microflora in the latter portions of birds GI tract are required to evaluate if exogenous β-1,3-1,4-glucanase might mediate their effects through the microflora route as well as the decrease in viscosity.

The effect of enzyme supplementation in digesta viscosity in a pool of duodenum and jejunum samples was measured and data is reported in Table 7.3. Digesta viscosity was affected by diet supplementation with exogenous enzymes. Digesta viscosity was significantly reduced (P<0.05) when animals were supplemented with the commercial enzyme mixture or the two recombinant enzymes. Since intestinal viscosity was decreased efficiently with the addition of both *Ct*Glc16A and *Ct*Cel8A in the diet, the results suggest that poor performance found in animals supplemented with the *Ct*Cel8A was not a consequence of the inability of the β-1,4-glucanase to decrease digesta viscosity (see below). However, these results should be viewed with some caution as viscosity was solely measured at the end of the experiment and not in the first weeks of the animal growth when it is supposed to have a higher impact in animal nutrition. It is well known that a lower digesta viscosity in the small intestine in the first period of broiler growth is critical for broiler performance (Jozefiak *et al*., 2006; Fontes *et al*., 2004; Cardoso *et al*., 2014; Ponte *et al*., 2008).

7.1.3.4. *Ct*Glc16A and *Ct*Cel8A remain stable during passage through broilers GI tract

In order to evaluate if the above reported differences in the capacity of the two recombinant enzymes to improve the nutritive value of a barley-based diet result from differences in polypeptide stability, samples from the various GI were collected and β-1,3-1,4-glucanase activity measured. The data, presented in Table 7.4, revealed that although β-1,3-1,4-

glucanase activity was detected in the crop and caecum of animals from all groups only birds receiving diets supplemented with *Ct*Glc16A and *Ct*Cel8A display significant levels of enzyme activity in the other regions of the GI tract. Intriguingly, non-supplemented animals and animals supplemented with the commercial mixture display similar levels of enzyme activity although the intensity of the halos collected from birds of the PC treatment were higher than those from non-supplemented animals (data not shown).

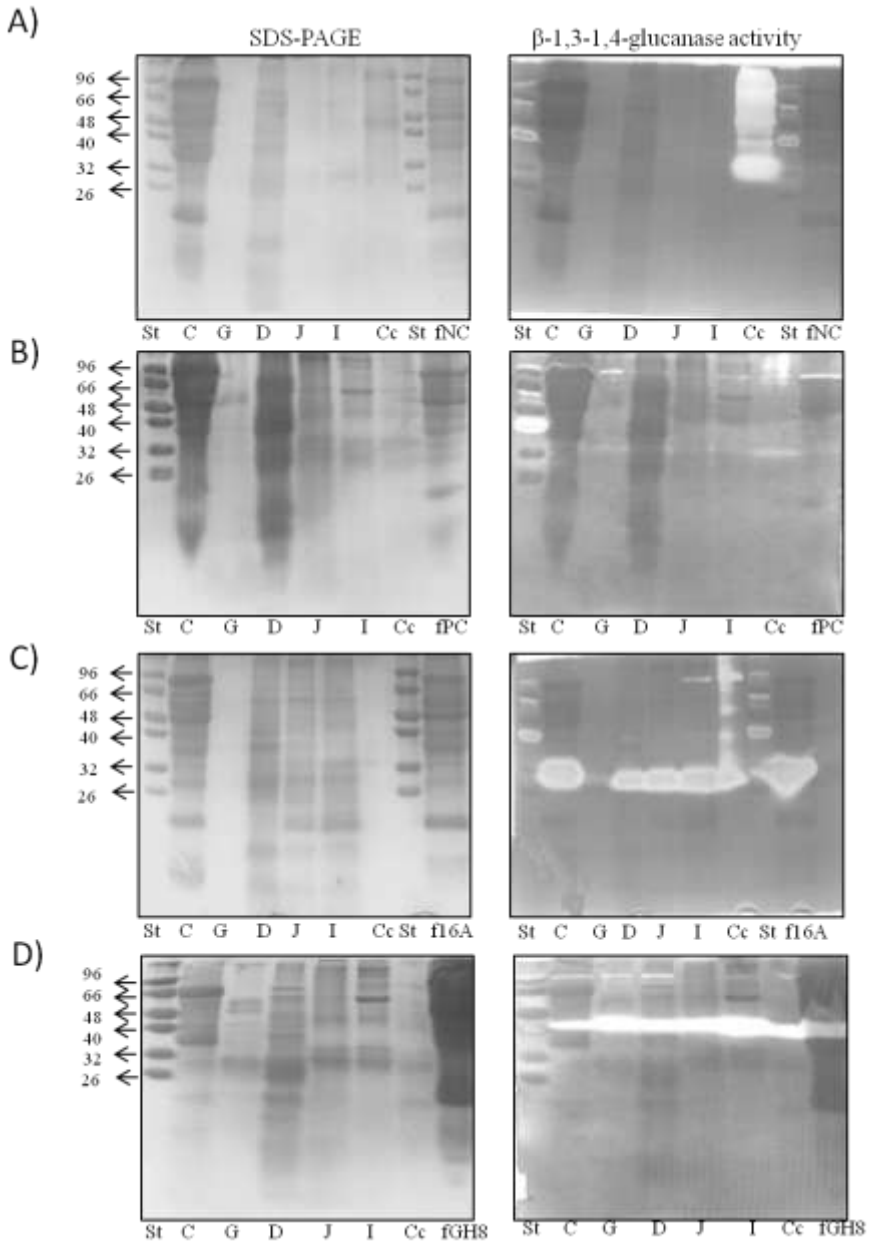
Table 7.4| Number of birds, out of 10 animals analyzed, fed on a barley-based diet non-supplemented (NC) or supplemented with different exogenous β-glucanases, presenting β-1,3-1,4-glucanase activity in digesta samples collected in different GI compartments. PC, a commercial enzyme mixture; *Ct*Glc16A, the recombinant β-1,3-1,4-glucanase; *Ct*Cel8A, the recombinant β-1,4-glucanase.

	NC	PC	*Ct*Glc16A	*Ct*Cel8A	Chi-Square	P- value
Crop	8	10	6	9	-	-
Gizzard	2	1	9	6	18.4	0.0004
Duodenum	2	2	10	8	16.6	0.0008
Jejunum	1	2	10	10	29.8	0.0001
Ileum	2	4	9	8	18.9	0.0003
Caecum	8	10	9	10	-	-

Zymogram analysis was used to evaluate the effect of exogenous enzyme passage through the birds GI tract in recombinant protein integrity. The data, presented in Figure 7.2, suggest that both *Ct*Glc16A and *Ct*Cel8A retain their molecular integrity throughout the entire GI tract. Thus, polypeptide bands of approximately at 32 (*Ct*Glc16A) and 48 (*Ct*Cel8A) kDa are detected with unchanged variations in size in all GI samples collected in animals receiving the recombinant *Ct*Glc16A and *Ct*Cel8A, respectively. The data also suggest that *Ct*Glc16A displays a low catalytic activity in the gizzard of the animals supplemented with this recombinant enzyme. The commercial β-1,3-1,4-glucanase mixture used in this study is mainly represented by a 32 kDa enzyme and it presented at low but detectable levels of catalytic activity throughout the GI tract. In addition, an enzyme with approximately 96 kDa was detected in samples collected from birds of the three groups receiving exogenous enzymes, suggesting the presence of an endogenous enzyme in the GI tract of these animals. In a previous study (Cardoso *et al*., 2014), the presence of an enzyme of high molecular weight was also reported and it was suggested to be from microbial origin since it was not present in the animal feed. However, more studies should be performed to evaluate the secretion of endogenous enzymes presenting β-glucanase activity and the presence of

endogenous enzymes at different levels of the GI tract. Significantly, caecum samples of all animals including those from the non-supplemented group revealed an heterogeneous and wide range of β-1,3-1,4-glucanases with different molecular weights.

Figure 7.2| Zymogram analysis for detection of catalytic activity against barley β-1,3-1,4-glucan of digesta samples collected from various regions of the GI tract of birds fed on a barley-based diet non-supplemented or supplemented with different exogenous β-glucanases.

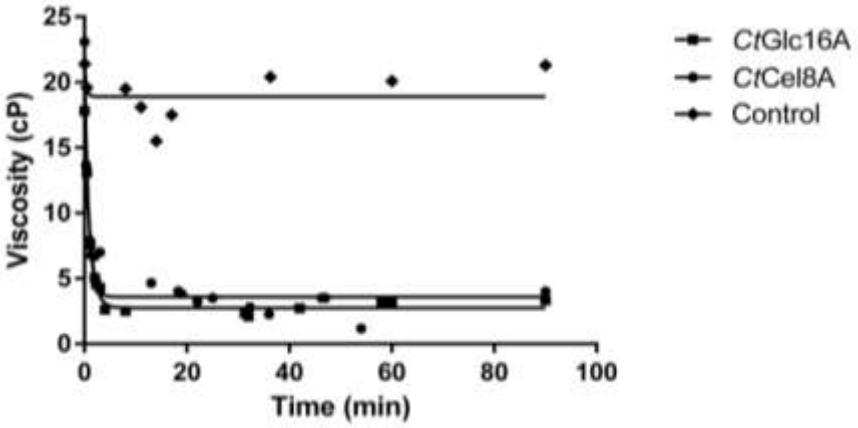


A) samples from birds fed on a diet non-supplemented; **B)** samples from birds fed on a diet supplemented with the commercial enzyme mixture; **C)** samples from birds fed on a diet supplemented with the recombinant β-glucanase *Ct*Glc16A; **D)** samples from birds fed on a diet supplemented with the recombinant cellulase *Ct*Cel8A. Proteins were fractionated through SDS-PAGE and stained for β-glucanase activity after enzyme renaturation. Abbreviations: St, low molecular weight protein standard; C, crop; G, gizzard; D, duodenum; J, jejunum; I, ileum; Cc, caecum; fNC, feed non-supplemented with exogenous enzymes; fPC, feed supplemented with the commercial enzyme mixture; f16A, feed supplemented with *Ct*Glc16A; f8A, feed supplemented with *Ct*Cel8A.

7.1.3.5. Strict β-1,3-1,4-glucanase specificity is required for exogenous enzymes used to improve the nutritive value of barley-based diets for broilers

Data presented above suggest that differences in the capacity of recombinant exogenous enzymes to improve the nutritive value of barley-based diets cannot be merely explained by differences in enzyme stability and integrity. Thus, to investigate if enzymes displaying strict β-1,3-1,4-glucanase specificity (*Ct*Glc16A) or broad β-1,3-1,4-glucanase and β-1,4-glucanase specificities (*Ct*Cel8A) have different capacities to reduce the viscosity of β-1,3-1,4-glucans, the two enzymes were incubated with a barley β-glucan suspension and levels of viscosity measured over a 90 min period. The data, presented in Figure 7.3, suggest that both enzymes are effective in reducing the viscosity of barley β-1,3-1,4-glucans. This is not completely surprising as a reduction in viscosity results from the reduction in the degree of polymerization of barley β-1,3-1,4-glucans and by cleaving the β-1,4-linkages of the mixed linked glucan, *Ct*Cel8A effectively contributes to the production of smaller glucan chains.

Figure 7.3| Effect of *Ct*Glc16A and *Ct*Cel8A in the viscosity of a barley β-glucan preparation.

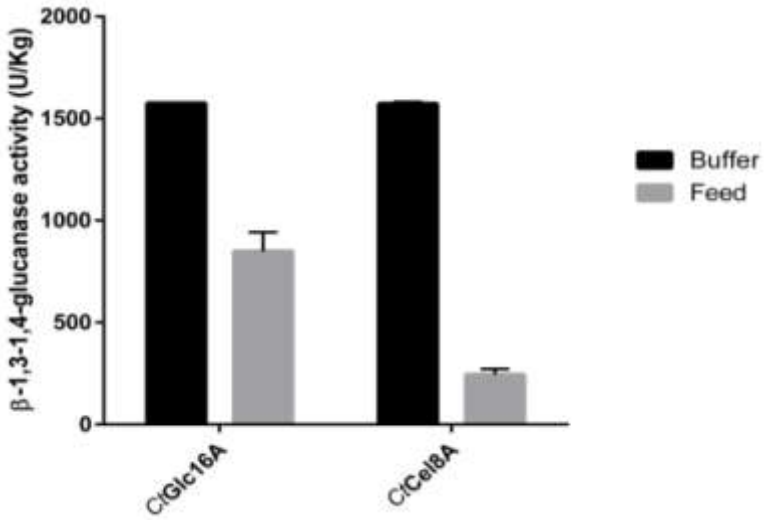


The two recombinant enzymes were incubated with a barley β-glucan solution (1.5%) and viscosity measured up to 90 min. Viscosity of the same solution was also measured when the polysaccharide was not exposed to the enzymes (Control).

Although β-1,4-glucanases can effectively reduce the degree of polymerization of barley β-1,3-1,4-glucans, as revealed by the *in vitro* experiment described above, it is possible that presence of β-1,4-glucans may reduce the efficacy of *Ct*Cel8A to cleave β-1,4-linkages within β-1,3-1,4-glucans. To evaluate this possibility, the capacity of *Ct*Glc16A and *Ct*Cel8A to hydrolyse β-1,3-1,4-glucans in the presence of the barley-based feed was investigated. The data, presented in Figure 7.4, revealed that there is a significant reduction of the activity of both enzymes in the presence of a barley-based feed extract. However, the magnitude of this reduction is much higher for *Ct*Cel8A, as only approximately 15 % of its β-1,3-1,4-glucanase activity was retained in the presence of the animal feed (Figure 7.4). In contrast, *Ct*Glc16A activity against the artificial substrate was reduced to only 54% in the presence of the barley-

based feed. Overall these data suggest that *Ct*Cel8A is preferentially targeting the β-1,4-linkages of β-1,4-glucans that are highly abundant in cereal-based diets (more than 25% of structural polysaccharides found in barley is cellulose). In contrast, the strict specificity revealed by *Ct*Glc16A limits the involvement of the enzyme in non-productive interactions with β-1,4-glucans and thus improves its efficacy against mixed linked glucans. Overall, this observation may explain why, *in vivo*, *Ct*Glc16A is more effective in improving the nutritive value of barley-based diets for broilers than *Ct*Cel8A.

Figure 7.4| Effect of feed in the activity of recombinant enzymes *Ct*Glc16A and *Ct*Cel8A.



The activity of the two recombinant enzymes was measured in the absence (Buffer) or the presence (Feed) of the barley based feed.

7.1.4. Conclusions

Here we have investigated if the improvement of the nutritive value of barley based diets for broilers results from the action of highly specific β-1,3-1,4-glucanases or β-1,4-glucanases displaying a broad substrate specificity that includes cleavage of β-1,3-1,4-glucans. The data firmly confirms that although retaining its molecular integrity and catalytic activity during its passage through birds GI tract, β-1,4-glucanase *Ct*Cel8A is unable to affect the nutritive value of barley based diets. *In vitro* it was observed that *Ct*Cel8A can effectively contribute to reduce the viscosity derived from the solubilization of pure mixed linked glucans. However, data suggests that presence of cellulosic substrates at high levels in barley-based diets may lead *Ct*Cel8A to establish non-productive interactions with β-1,4-glucans, thus limiting its capacity to hydrolyze the anti-nutritive β-1,3-1,4-glucans. These observations suggest that there is a considerable scope to optimize enzyme mixtures employed to supplement barley-based diets by increasing the predominance of β-1,3-1,4-glucanases in relation to β-1,4-glucanases. In addition, it might be highly relevant to know for each microbial enzyme mixture used in poultry nutrition what is the contribution of β-1,4-glucanases to the overall β-1,3-1,4-glucanase activity expressed by the enzyme supplements.

7.2. Construction of GH16 β-glucanase mini-cellulosomes to improve the nutritive value of barley-based diets for broilers

Vania O. Fernandes[1,3*], Mónica Costa[1*], Teresa Ribeiro[1], Luís Serrano[1], Vânia Cardoso[2], Helena Santos[1], Madalena Lordelo[2], Luís M.A. Ferreira[1,3] and Carlos M.G.A. Fontes[1,3]

[1] CIISA-Faculdade de Medicina Veterinária, Universidade de Lisboa, Avenida da Universidade Técnica, 1300-477, Lisboa, Portugal; [2] Instituto Superior de Agronomia, Tapada da Ajuda, 1349-017 Lisboa, Portugal; [3] NZYTech genes & enzymes, Campus do Lumiar, Estrada do paço do Lumiar, 1649-038 Lisboa, Portugal.
*Equal contribution

Adapted from: Fernandes *et al*. (2014). *Journal of Agricultural and Food Chemistry*. 62(30):7496-7506

Abstract

Anaerobic cellulolytic bacteria organize a comprehensive range of cellulases and hemicellulases in high molecular weight multi enzyme complexes termed cellulosomes. Integration of cellulosomal components occurs via highly ordered protein:protein interactions between cohesins and dockerins, whose specificity allows the incorporation of cellulases and hemicellulases onto a molecular scaffold. Here we report the production of two mini-cellulosomes containing one (GH16-1C) or three (GH16-3C) copies of *Clostridium thermocellum* Glucanase 16A (*Ct*Glc16A). Assembling of *Ct*Glc16A in the two protein complexes had no effect in the pH and thermal properties of the cellulosomal enzyme. Due to the presence of high levels of β-1,3-1,4-glucans, barley-based diets express considerable anti-nutritive value for monogastric animals, in particular for poultry. Thus, GH16-1C and GH16-3C were used to supplement barley based diets for broilers. The data revealed that the two mini-cellulosomes very effectively improved the nutritive value of barley-based diets for broilers, although the efficacy of GH16-3C complex seemed to be lower than GH16-1C. Analysis of the molecular integrity of the two mini-cellulosomes suggested that although cohesins and the *Ct*Glc16A catalytic domain are highly resistant to proteolytic inactivation, linker sequences separating protein domains in scaffoldins and cellulosomal catalytic units are highly susceptible to proteolytic attack. Thus, overall the data suggest that efficacy of mini-cellulosomes to improve broiler performance results from the action of *Ct*Glc16A per se rather from the association of the enzyme in a multi-enzyme complex.

7.2.1. Introduction

Plant cell wall degrading anaerobic bacteria organize a large repertoire of glycoside hydrolases, carbohydrate esterases and pectate lyases into a high molecular weight multi-enzyme complex termed the cellulosome, one of nature´s most elegant and effective nanomachine characterized so far (Fontes & Gilbert, 2010). The major player in the organization of cellulosomes is a non-catalytic modular protein termed scaffoldin, which contains a variable number of cohesin domains. Cellulosomal enzymes contain a C-terminal dockerin that tenaciously bind scaffoldin cohesins and this constitutes the primary mechanism of cellulosome assembly (Carvalho *et al*., 2003). Cohesin-dockerin interactions involved in cellulosome assembly were termed of type I. In addition, several scaffoldins were found to contain a C-terminal type II dockerin that does not interact with their internal type I cohesins but rather binds type II cohesins located in anchoring scaffoldins located at the bacterium cell surface (Adams *et al*., 2005). It is now well established that in contrast to type II dockerins, type I dockerins contain two identical cohesin binding interfaces (Carvalho *et al*., 2007). This dual binding mode is most probably responsible to the introduction of an intrinsic flexibility into the quaternary structure of highly diverse and populated enzyme machinery (Fontes & Gilbert, 2010). Organization of cellulases and hemicellulases in cellulosomes improves enzyme stability and activity and provides a rational for the increased efficiency displayed by anaerobic organisms upon the degradation of highly recalcitrant polysaccharides such as cellulose and hemicellulose.

The high-affinity protein:protein structure established between cohesins and dockerins ($>10^9$ M^{-1}) forms a blue print for the production of tailored multicomponent catalytic nanomachines for a range of biological processes that might benefit from enzyme proximity. This will require engineering scaffoldin surfaces to interact with defined dockerins to allow construction of macromolecular assemblies with specific functions, generically termed mini-cellulosomes. Recently, a variety of examples have been explored where mini-cellulosomes were used to optimize different biotechnological applications, mainly in what concerns to the production of second-generation biofuels originated from cellulosic biomass. The importance of these biofuels, predominantly the cellulosic bioethanol, relies on the origin of renewable sources of energy which can be responsible for a reduction in greenhouse gas levels leading to a decreased environmental impact and a diminished cost investment compared to first generation energies (Malça & Freire, 2006; Nordon, Craig, & Foong, 2009; Tamaru *et al*., 2010; Sheridan, 2009). Recent studies described the possible involvement of engineered cellulosomal protein complexes in a new method called consolidated bioprocessing (BPC), which further combines enzyme production with cellulose saccharification and fermentation (SSCF) into a single process to produce ethanol (Lynd *et al*., 2002b; Lynd *et al*., 2008). Thus, Tsai, Goyal, & Chen (2010) reported an almost two-fold increase in both cellulose hydrolysis and ethanol production (0.475 g of ethanol/ g of sugar consumed) when using an optimized

consortium composed by four different engineered yeast strains capable of either displaying a trifuncional mini-scaffoldin carrying three divergent cohesin domains from *Clostridium thermocellum*, *C. cellulolyticum* and *Ruminococcus flavefaciens* or secreting one of the three required dockerin-tagged enzymes (endoglucanase, exoglucanase or β-glucosidase). More recently, Fan *et al*. (2013) studied the *in vitro* functional assembly of cellulosomes with two mini-scaffoldins on the yeast cell surface and recombinant cellulases intracellularly expressed in *E. coli* or secreted by yeasts. Although, according to Tsai, Goyal, & Chen (2010), the production method of ethanol presented when using *E.coli* lysate treated cells could not be considered BCP due to separate enzyme production, those cells whose anchoring mini-scaffoldins were optimized produced approximately 1138 mg/L of ethanol from microcrystalline cellulose within 4 days. However, it is to notice that the ethanol titer and cellulose consumption were lower than that obtained when using a yeast self-assembled mini-cellulosome (Fan *et al*., 2012). Attending to the length of scaffoldin II, Fan *et al*. (2013; 2012) reported a higher display level on yeasts with smaller scaffoldin II (lower amount of cohesin II), as well as a maximum ethanol production when the mini-cellulosome had only two cohesins II. Considering the number of cohesin I domains, Cha *et al*. (2007) obtained a little difference in the effects on cellulosic and hemicellulosic substrates produced by three different *C. cellulovorans* recombinant mini-cellulosomes containing either endoglucanase EngB or endoxylanase XynA bound to mini-CbpA with one (mini-CbpA1), two (mini-CbpA12) or four (mini-CbpA1234) cohesin domains when the cellulosomal enzyme concentration was held constant, regardless of the copy number of cohesins in the cellulosome. However, a synergistic effect was observed when the enzyme concentration was increased to be proportional to the number of cohesins in the mini-cellulosome. Comparing mini-cellulosome with free cellulosomal enzymes, Murashima, Kosugi, & Doi (2002) reported an enhanced activity towards crystalline cellulose when using mini-cellulosomes.

Cereal-based diets contain high levels of soluble non-starch polysaccharides (NSPs) that significantly affect the digestive process impairing animal performance. It is now well known that barley incorporation in poultry diets is limited by its high content in soluble β-1,3-1,4-glucans that upon solubilization lead to an increase in digesta viscosity, reducing digesta passage rate and affecting the interaction of the endogenous digestive enzymes with their target substrates (Smits & Annison, 1996; Pettersson & Aman, 1989; Jozefiak *et al*., 2006). In addition, higher viscosities also lead to prolonged digesta passage rates promoting a modification in gut physiology such as the enlargement of gastrointestinal organs, which are detrimental final carcass yields (Smits & Annison, 1996). High digesta viscosity also favors the proliferation of anaerobic microbes in the upper parts of the GI tract thus affecting animal's health (Jozefiak *et al*., 2006). To reduce the negative effects associated with the presence of barley β-glucans, commercial enzyme mixtures expressing high levels of β-glucanase activity are currently added to broilers diets (Bedford & Morgan, 1996).

Exogenous enzymes added to poultry diets reduce the degree of polymerization of soluble glucans leading to a reduction in digesta viscosity, an improvement of diet digestibility and feed intake, thus contributing to improve animal performance (Bedford & Morgan, 1996). However, little is known about the mechanisms affecting the efficiency of exogenous enzymes used to improve the nutritive value of cereal-based diets *in vivo*.

Here we aim to produce mini-cellulosomes expressing β-1,3-1,4-glucanase activity and use the engineered nanomachines to improve the nutritive value of barley-based diets for broilers. These mini-cellulosomes are based on *C. thermocellum* CipA scaffoldin and incorporate one or three *Ct*Glc16A enzymes in a single nanomachine.

7.2.2. Materials and Methods

7.2.2.1. Bacterial strains and Plasmid

Genomic DNA of *Clostridium thermocellum* ATCC 27405 was purchased from DSMZ (Germany). *Escherichia coli* DH5α cells were used for cloning and *E. coli* BL21(DE3) was used as the expression host. The plasmids used for cloning and expression were pET21a (+) derivatives. All the above mentioned items were procured from Novagen (Madison, USA).

7.2.2.2. Gene isolation and cloning

The thermostable β-1,3-1,4-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 followed by a C-terminal type-I dockerin. The gene encoding mature *Ct*Glc16A (residues 30-334) was amplified from *C. thermocellum* genomic DNA through PCR using the NZYProof DNA polymerase (NZYTech, genes & enzymes, Portugal) and the following primers: 5´-CACACACA**GCTAGC**ACTGTGGTAAATACGCC-3´ and 5´-CACACACA**GGATCC**TCAAAGTGACGGAATTG-3´ (NZYTech, genes & enzymes, Portugal). Primers included engineered NheI and BamHI restriction sites (in bold) that allowed the subsequent cloning of the resulting nucleic acid into similarly restricted pET21a, generating the plasmid pGH16-Doc21a. pGH16-Doc21a encodes the bi-modular *Ct*Glc16A, containing a N-terminal GH16 β-1,3-1,4-glucanase catalytic domain and a C-terminal type I dockerin. Previously we observed that dockerins are highly unstable when expressed in *E. coli* cells. However, cellulosomal enzymes are stabilized once dockerins bind their cohesin counterpart (García-Alvarez *et al*., 2011). Thus, cohesin-dockerin complexes are usually assembled *in vivo* in cells co-expressing the genes encoding the dockerin and cohesin containing proteins. Previously we produced a *C. thermocellum* mini-cellulosome containing three CtCel8A cellulases bound to a mini-scaffoldin comprising cohesin modules 3, 4 and 5 of CipA (residues 562–1030) (García-Alvarez *et al*., 2011). The mini-cellulosome was produced in *E. coli* cells transformed with a pET21a derivative containing the genes encoding the cellulase and the tri-cohesin mini-scaffoldin organized in tandem and under the control of separate T7

promoters (the plasmid was termed pG8D3C21a; see García-Alvarez *et al*., (2011) for details). Here we used pG8D3C21a construct and replaced the gene encoding *Ct*Cel8A by *Ct*Glc16A gene through restriction digestion. Thus, pGH16-Doc21a was digested with *Sph*I and *Bam*HI and the DNA fragment containing the β-glucanase gene and the 5´-region of pET21a was cloned into a truncated version of pG8D3C21a from which the *Ct*Cel8A gene was excised after a *Sph*I-*Bgl*II digestion, generating pG16D3C21a. pG16D3C21a contains the genes encoding the *Ct*Glc16A glucanase and the three-cohesin mini-scaffoldin organized in tandem. In order to prepare a protein:protein complex resulting from the binding of a single-cohesin domain protein to *Ct*Glc16A, to use as a control in the biochemical experiments described below, the same approach was employed to produce plasmid pGH16Doc-1Coh21a by replacing *Ct*Cel8A with *Ct*Glc16A in plasmid pG8D1C21a, which contain the gene encoding *Ct*Cel8A followed by the gene encoding a single cohesin, cohesin 3, of CipA scaffoldin. The two resulting plasmids used in this study, pGH16Doc-3Coh21a (GH16-3C) and pGH16Doc-1Coh21a (GH16-1C), were subsequently sequenced to confirm that no mutations were accumulated during gene amplification and the resulting sub-cloning steps. The engineered cohesin containing proteins (3Coh and 1Coh) contain a C-terminal His tag, allowing the direct purification of the assembled complex and unbound scaffoldins by immobilized metal affinity chromatography (IMAC). Since cohesin containing proteins are usually expressed at low levels in *E. coli*, it is expected that in both cases the majority of the proteins recovered by affinity chromatography will constitute the protein complexes.

7.2.2.3. Expression and purification of *Ct*Glc16A mini-cellulosomes

E. coli BL21(DE3) cells were transformed with pGH16Doc-3Coh21a and pGH16Doc-1Coh21a as described elsewhere (Das *et al*., 2012). Recombinant cells were grown in Luria-Bertani medium supplemented with ampicillin (100 μg ml^{-1}) at 37 ºC, 200 rev/min, till mid-exponential phase (A_{600nm} ≈ 0.4). Recombinant gene expression was induced by adding 1 mM isopropyl-1-thio-β-D-galactopyranoside (NZYTech, genes & enzymes, Portugal) and cells were further incubated at 19 ºC for 16 h. The cells were harvested at 5,000 xg and the resulting pellet was resuspended in 10 mM Imidazol buffer (10mM Imidazol, 50 mM NaHEPES, 1M NaCl, 5 mM $CaCl_2$, at pH=7.5), submitted to ultrasonication, centrifuged at 4 ºC, 17000 xg during 30 min and, finally, the obtained His_6- 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 protein complexes was analyzed by SDS-PAGE (Laemmli, 1970).

7.2.2.4. Analytical Procedures

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 substrates used in this study were barley β-glucan, hydroxyethylcellulose, carboxymethylcellulose, arabinoxylan, xyloglucan, galactomannan, glucomannan (Megazyme, Ireland) at a 0.3% (w/v) final concentration. Reactions were stopped by adding a DNSA based solution (1% DNSA, 1% NaOH and 0.2% fenol) following the method described by Miller (1959). To explore the pH profile of GH16-1C (pGH16Doc-3Coh21a) and GH16-3C (pGH16Doc-1Coh21) proteins and compare it with the *Ct*Glc16A, 50mM MES (2-N-morpholino- ethanesulfonic acid) (pH 4.5-7), 50mM Tris-HCL (pH 7-9.5), and 50mM $NaHCO_3$ (pH 9.5-11) buffers were used in enzyme assays employing 0.25% barley β-glucan as the substrate. For thermostability experiments, the three proteins were incubated at temperatures ranging from 60 to 90 ºC. After 20 min at the required temperature, samples were withdrawn and residual activity was determined at 50 ºC by measuring the amount of reducing sugar released from barley β-glucan as described above. Determination of temperature of maximal enzyme activity was performed by incubating the enzyme at temperatures ranging from 50 to 90 ºC and measuring reducing sugar release as previously described. One unit of catalytic activity is defined as the amount of enzyme required to release 1 µmol of product per min.

In a previous study we shown that levels of endogenous β-glucanase activity affect the efficacy of exogenous enzymes used to improve the nutritive value of barley-based diets for poultry (Ribeiro *et al*., 2011). To select barley batch expressing lower levels of endogenous β-glucanases, five different barley lots available commercially were selected. The barley lots were milled at 0.5 mm and enzyme extraction was performed by adding 1 mL of phosphate/citrate (PC) buffer into 0.75 g of barley. The supernatant was recovered and used to quantify the β-glucanase activity using a β-glucanase commercial assay kit (Megazyme, Ireland), following the manufacturer protocol. The incubation period was extended to 3 hours. The barley lot expressing the lowest endo-β-glucanase activity was used as the main component of the basal diet described below.

In order to standardize the number of enzyme units used to supplement the animal diets, the catalytic activity of the recombinant enzymes, including the commercial enzyme mixture (Rovabio™ Excel AP®), was determined in parallel at 40 ºC by measuring the release of reducing sugars (Fontes *et al*., 2000). The extract containing the commercial enzyme was prepared by ressuspending 250 mg of the enzyme in 10 mL of PC buffer, followed by an overnight incubation at room temperature with gentle agitation and centrifugation at 16000 g for 10 min.

To evaluate the levels of β-glucanase activity present in the four animal diets and in digesta samples collected in different parts of the broilers´ GI tract, samples were initially centrifuged at 16000 ×g for 5 min and the supernatants were recovered for analysis. The barley and feed samples (0.75 g) were previously mixed in 1 mL of PC buffer. The mixture samples were

subjected to vigorous stirring during 30 min, centrifuged and the supernatant was analyzed. Qualitative analysis of β-glucanase activity 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 8.0). Catalytic activity was detected based in the method described by Ponte *et al*. (2004) with some modifications. Briefly, a Pasteur pipette was used to create wells in the agar plates, which were filled with 20 µL of enzyme extract from digesta samples. Subsequently, the plates were incubated for 16 h at 37 °C and dyed with 1% Congo red (E.Merck AG®) in10mM Tris-HCl (pH=8) for 30 min. To observe regions where the polysaccharides were degraded, the plates were subjected to three 15 min washes with 1M NaCl in 10mM Tris-HCl (pH=8). For measuring the viscosity of small intestine contents, samples collected from the duodenum plus jejunum and ileum were centrifuged for 10 min at 10867 x*g* and the viscosity of sample supernatant was measured at 6 x*g* using a viscometer (Model LVDVCP-II, Brookfield Engineering Laboratories, Middleboro, MA) with a cup maintained at 24 ºC.

7.2.2.5. Incorporating mini-cellulosomes in poultry diets

Bird experiment was 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 (Direcção Geral de Veterinária, Lisboa, Portugal), following the appropriate European Union guidelines (Council Directive 86/609/EEC). 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 (treatment NC) or supplemented with the commercial enzyme Rovabio™ Excel AP (Adisseo, France; treatment Rov), or supplemented with the pGH16Doc-1Coh21a (treatment 1C) or supplemented with the pGH16Doc-3C21a (treatment 3C). All the three enzymes were incorporated at a calculated dose of 1500 U/kg of feed. The calculated 1500 U/kg of feed of the commercial enzyme corresponded to the manufacturers recommended dose of 50 g of enzyme per ton of feed. The duration of the trial was 28 days. The basal diet (Table 7.6) contained 615 g/kg of barley and was formulated to ensure a nutrient availability as defined by the NRC (N.R.C, 1994). 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 throughout the experiment, and mortality was recorded daily. At 28 days of age, one bird per pen was slaughtered by cervical dislocation. The weight and lenght of the crop, gizzard, liver, duodenum, jejunum, ileum and caecum were determined. Digesta samples were collected at duodenum and jejunum, and ileum compartments to determine contents viscosity as described above.

7.2.2.6. Zymogram analysis and resistance of mini-cellulosomes to proteinase attack

Zymogram analysis was performed as described by Fontes *et al*. (2004) and Ribeiro *et al*. (2008). Summarily, digesta proteins were separated through SDS-PAGE electrophoresis in 14% acrylamide gels containing 0.1% of barley β-glucan (Megazyme, Ireland), according to Laemmli (1970). After electrophoresis, polypeptides were renatured by subjecting the gel to five 30 min washes in 100 mM sodium succinate, pH 6.3, 10 mM $CaCl_2$ and 1 mM DTT. The gels were incubated for 16 h at 37 ºC in the same buffer and proteins were stained in a solution comprising 40% (v/v) methanol, 10% (v/v) glacial acetic acid and 0.4% (w/v) Coomassie Brilliant Blue R. After destaining, gels were washed in 0.1M Tris-HCl (pH=8) for 20 min. β-glucanase activity was detected using a 0.1% (w/v) Congo Red solution for 15 min and washing with 1M NaCl in 10mM Tris-HCl (pH=8) until excess dye was removed. After Congo Red staining the gels were counter stained with 1N HCl as described elsewhere (Ruijssennars & Hartmans, 2001). Areas of catalytic activity appeared as colourless zones in a grey background.

The ability of mini-cellulosomes to resist to protease degradation was evaluated by incubating the two protein complexes (approximately 30 µg) with a 50 µg of pancreatin from porcine pancreas (Sigma, #P-1500) in Buffer A (50 mM Hepes, 50 mM NaCl, 5 mM $CaCl_2$, pH 7.5) at 37 °C during 30 minutes. Negative controls consisting on the reaction without pancreatin were prepared, as well as a reaction only with pancreatin in Buffer A. Mini-cellulosome integrity was evaluated by SDS-PAGE, as described above.

7.2.2.7. Statistical Analysis

Data related to bird performance from each experiment were subjected to ANOVA according to the general linear models procedure of SAS (SAS, 2004). The Least Squared Means procedure was used to detect significant differences between treatment groups. The experimental unit was the cage of 4 animals (n=10). Differences were considered significant when $P < 0.05$.

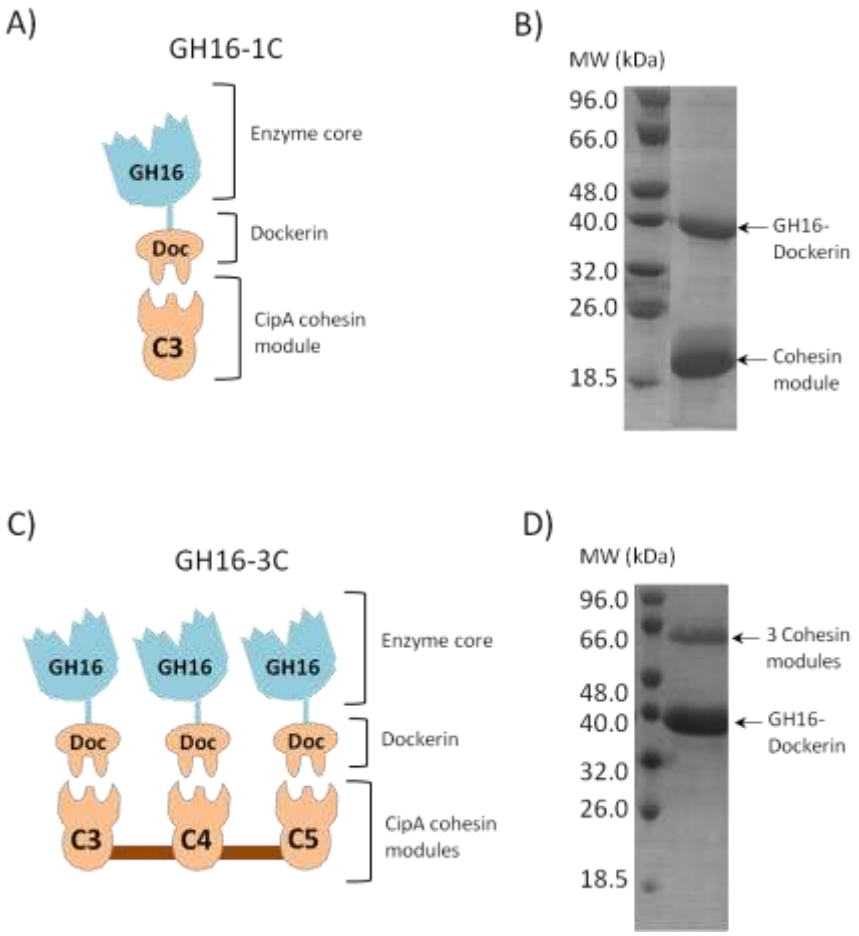
7.2.3. Results and Discussion

7.2.3.1. Construction of mini-cellulosomes expressing β-1,3-1,4-glucanase activity

The open reading frame of *Ct*Glc16A, amplified by polymerase chain reaction was used to replace the gene encoding *Ct*Cel8A in plasmids previously prepared to produce a mini-cellulosome containing *Ct*Cel8A bond to mini-scaffoldins either containing 1 or three cohesin domains (García-Alvarez *et al*., 2011). Two *C. thermocellum* mini-cellulosome protein complexes (Figure 7.5 A and C) of roughly 52 and 156 kDa were produced. The complexes comprising one (C3; 20 kDa) or three cohesin modules (C3–C4–C5; 52 kDa) of CipA bound

to one or three cellulosomal *Ct*Glc16A glucanases (34.9 kDa), respectively, were purified from *E. coli* via nickel affinity chromatography (Figure 7.5 B and D).

Figure 7.5| Molecular architecture of mini-cellulosomes (A & C) and protein purification (B & D).



The molecular architecture of mini-cellulosomes containing 1 cohesin fused to 1 *Ct*Glc16A enzyme or 3 cohesin domains fused to 3 *Ct*Glc16A enzymes are displayed in panels A and C, respectively. The two complexes were expressed and assembled *in vivo* in *E. coli* and purified through IMAC using a Poly-histidine-tag located at the C-terminus of the mini-scaffoldins as the affinity tag. SDS-PAGE was used to confirm complex homogeneity as observed in panels B (GH16-1C) and D (GH16-3C), respectively.

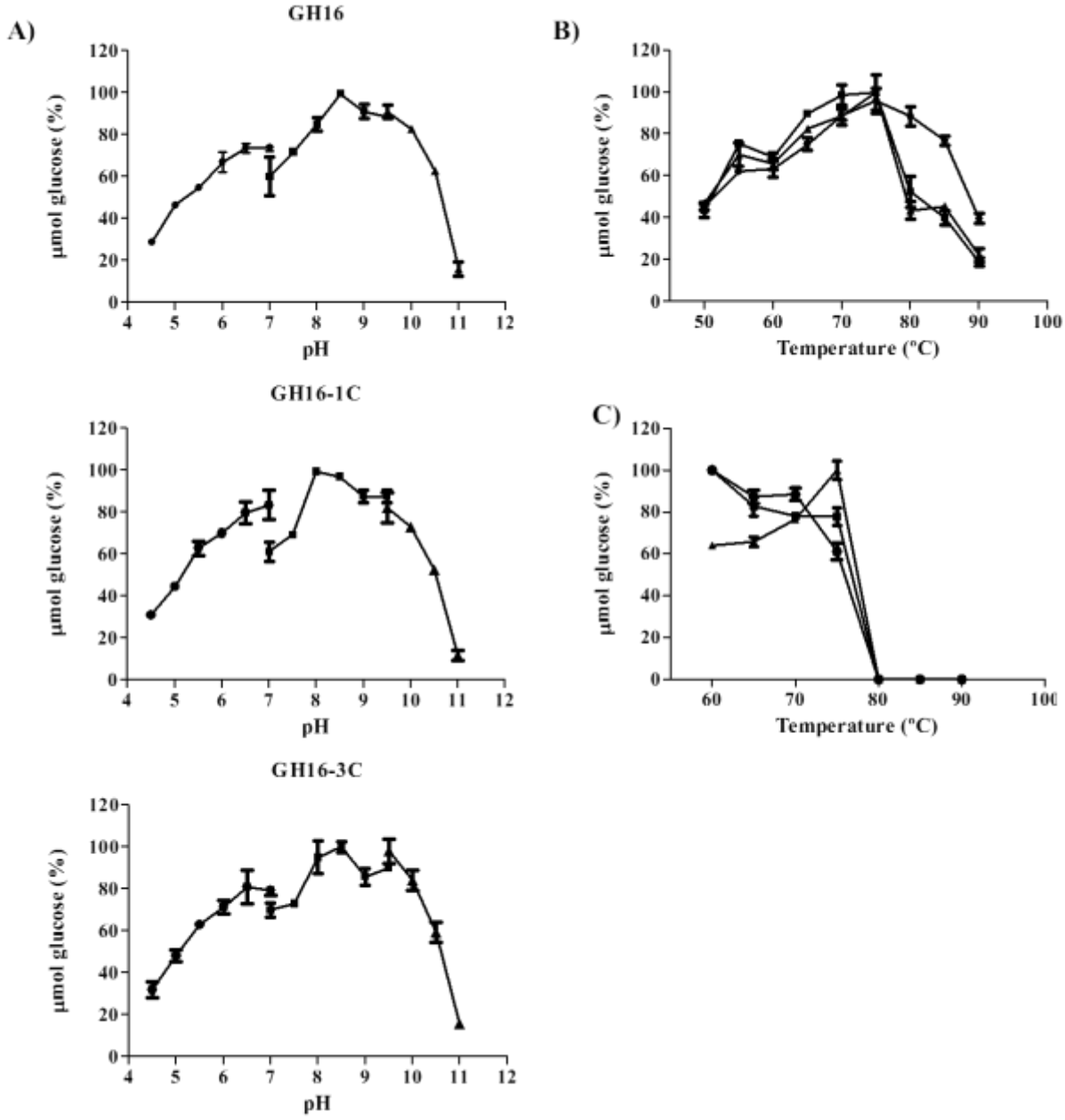
The genes encoding the truncated versions of CipA (C3 or C3–C4–C5) and *Ct*Glc16A were co-expressed in the bacterium under the control of separate T7 promoters (see Materials and Methods). The recombinant CipA derivatives, C3 and C3–C4–C5, were engineered to contain a C-terminal His_6-tag. The levels of expression of *Ct*Glc16A were estimated to be at least five times higher than the levels of expression of C3 or C3–C4–C5 mini-scaffoldins (not shown). Thus, during expression, the mini-scaffolds became saturated with the cellulosomal glucanase, allowing the purification of homogeneous complexes through affinity chromatography that targeted the C-terminal His_6-tag of C3 or C3–C4–C5, respectively.

7.2.3.2. Assembly of *Ct*Glc16A in mini-cellulosomes does not affect pH and temperature stability of the enzyme

*Ct*Glc16A is a typical β-glucanase as it does not exhibit any activity against soluble cellulosic substrates and hemicelluloses but displays specificity for β-1,3-1,4-glucans, such as lichenan and barley β-glucan (Ribeiro *et al*., 2012). The enzyme displays a broad pH optimum (pH 6-9), with maximal activity at pH 8.0, and is thermostable at temperatures of ~70 °C. In addition, *Ct*Glc16A was found to be completely resistant to proteolytic inactivation (Ribeiro *et al*., 2012). Thus, *Ct*Glc16A biochemical properties suggest that the enzyme is particularly suited to resist the harsh conditions observed during feed processing and poultry digestion (Ribeiro *et al*., 2012) and thus constitutes an ideal candidate to be assembled in mini-cellulosomes. In fact, *Ct*Glc16A is a cellulosomal enzyme isolated from the anaerobic thermophilic bacterium *C. thermocellum*.

Organization of cellulases and hemicellulases in mini-cellulosomes is supposed to confer additional stability into the enzyme complex and improved activity on the hydrolysis of recalcitrant substrates (Fontes & Gilbert, 2010). Here we interrogate the biochemical properties of *Ct*Glc16A when organized in mini-cellulosomes containing one (GH16-1C) or three-cohesin (GH16-1C) mini-cellulosomes. The data, presented in Figure 7.6, revealed that mini-cellulosomes GH16-1C and GH16-3C displayed optimum temperatures and pH similar to those expressed by uncomplexed *Ct*Glc16A; the mini-cellulosomes expressed a broad pH optimum (6-9) and displayed unaltered optimum temperatures and thermostability when compared with *Ct*Glc16A. The substrate specificities of GH16-1C and GH16-3C were also probed against a range of natural substrates at optimized pH and temperatures. The data, presented in Table 7.5, revealed that the two mini-cellulosomes retained a strict specificity for β-1,3-1,4-glucans. Thus, reflecting the specificity displayed by *Ct*Glc16A, the mini-cellulosomes were unable to degrade cellulosic substrates and a range of the most frequent hemicelluloses. In addition, the capacity of the enzyme to degrade β-glucans is not affected by the binding to the mini-scaffoldins although a slight reduction of the activity of GH16-3C during the hydrolysis of barley β-glucan is observed. This might reflect a restriction of *Ct*Glc16A flexibility upon its organization of mini-cellulosomes. This flexibility is of primary importance in the hydrolysis of soluble substrates. In contrast, organization of mini-cellulosomes is of primary importance to improve the catalytic efficiency of hydrolysis of recalcitrant substrates such as cellulose.

Figure 7.6| pH (Panel A), temperature (Panel B) and thermostability (Panel C) profiles of unbound *Ct*Glc16A (GH16) or the enzyme when organized in mini-cellulosomes GH16-1C and GH16-3C.



In **Panel A**, the enzymes were incubated at standard conditions in MES (●), Tris (■) or NaHCO3 (▲) buffers with different pH, and β-glucanase activity determined (% of glucose released). In **Panel B**, *Ct*Glc16A (●), pGH16Doc-1Coh21a (■) and pGH16Doc-3C21a (▲) activity were determined against barley β-glucan at different temperatures. In **Panel C**, thermostability of the enzyme and enzyme complexes was determined by incubating the proteins for 30 min at different temperatures, and residual activity determined at 55 ºC (figure labels are the same as panel B).

Table 7.5| Substrate specificity of *Ct*Glc16A, GH16-1C and GH16-3C, measured at 70 ºC and pH 8.0[a].

Substrate	*Ct*Glc16A	GH16-1C	GH16-3C
Barley β-glucan	2845.1 ± 35.98	2859.4 ± 98.39	2321.2 ± 139.50
Hydroxyethylcellulose	0	0	0
Carboxymethylcellulose	0	0	0
Arabinoxylan	0	0	0
Xyloglucan	0	0	0
Galactomannan	0	0	0
Glucomannan	0	0	0

[a]*Ct*Glc16A and *Ct*Glc16A mini-cellulosomes were incubated with a variety of carbohydrates (0,25%) and release of reducing sugars measured through the DNSA assay as described in Material and methods. Enzyme activity is expressed in number of enzyme units per mg of *Ct*Glc16A. Values are the average of three independent assays.

7.2.3.3. Using mini-cellulosomes to improve the nutritive value of barley based diets for poultry

The effect of assembling *Ct*Glc16A in mini-cellulosomes organized by scaffoldins containing 1 (GH16-1C) or 3 (GH16-3C) cohesins, respectively, in improving the nutritive value of barley-based diets for poultry was analyzed in an animal trial using 160 broiler chicks. Mini-cellulosome GH16-1C was used as a control mini-cellulosome as it contains exclusively one enzyme per protein complex and should thus behave as a typical single enzyme. Basal diet (Table 7.6) contained a major proportion of barley (>60%) and was formulated to ensure a nutrient availability as defined by the NRC (N.R.C, 1994). Barley contains a significant proportion of β-glucans which, due to its intrinsic viscosity, display considerable anti-nutritive properties for monogastric animals, in particular for poultry. Thus, the basal diet was either not supplemented with exogenous enzymes (negative control group, NC) or supplemented with three different β-glucanase preparations: the two mini-cellulosomes prepared above (GH16-1C and GH16-3C) or a commercial enzyme. The four diets were fed broiler chicks during a 28 day trial.

Table 7.6| Ingredient composition and calculated nutrient content.

Ingredients	g/ kg
Barley	615
Soybean meal 47%	294
Soybean oil	59
Salt	2.5
Calcium carbonate	8.4
Dicalcium phosphate 18%	17.4
DL-Methionine	1.7
Mineral and vitamin premix[a]	2
Nutrient content	
Energy (MJ ME/kg DM)	12.1
Crude protein	208
Ether extract	75
Crude cellulose	51

[a]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;

Animal performance data, presented in Table 7.7, revealed that as early as day 7 and along the entire trial animals from NC group had lower ($P<0.05$) body weight than animals fed on diets supplemented with the exogenous enzyme preparations. At days 7 and 14, animals from the Rov group presented the highest body weights that were similar to the animals from GH16-1C treatment ($P>0.05$). In addition, although GH16-3C animals had similar body weights to animals from the GH16-1C group ($P>0.05$), they presented lower weights ($P<0.05$) than animals supplemented with the commercial enzyme. These data suggest that on the initial period of the broiler growing period assembling *Ct*Glc16A in a 3 cohesin mini-cellulosome might restrict the efficacy of the exogenous β-glucanase. Weekly weight gains were only significantly different until day 14 (Table 7.7).

Table 7.7| Growth performance parameters of broilers fed on a barley-based diet non supplemented (NC) or supplemented with different exogenous enzymes. Rov, a commercial enzyme, GH16-1C, mini-cellulosome construct containing 1 cohesin, GH16-3C, mini-cellulosome construct containing 3 cohesins.

	NC	Rov	GH16-1C	GH16-3C	SEM	p(*F*)
Body Weight (g)						
0d	42.3	42.0	42.0	42.2	0.16	0.3059
7d	126 ^{c}	150 ^{a}	146 ^{ab}	140 ^{b}	2.32	0.0001
14d	298 ^{c}	393 ^{a}	394 ^{ab}	370 ^{b}	7.26	0.0001
21d	664 ^{b}	784 ^{a}	766 ^{a}	749 ^{a}	13.5	0.0001
28d	1105 ^{b}	1234 ^{a}	1229 ^{a}	1227 ^{a}	22.4	0.0004
Weight Gain (g)						
0-7d	81.6 ^{c}	108 ^{a}	104 ^{a}	96.5 ^{b}	2.33	0.0001
7-14d	174 ^{b}	243 ^{a}	245 ^{a}	232 ^{a}	5.60	0.0001
14-21d	363	386	375	379	9.17	0.3447
21-28d	443	450	465	478	12.8	0.1494
0-28d	1063	1192	1188	1185	42.7	0.0621
Feed Intake (g)						
0-7d	111 ^{b}	129 ^{a}	128 ^{a}	120 ^{a}	3.60	0.0028
7-14d	254 ^{c}	329 ^{ab}	331 ^{a}	312 ^{b}	5.99	0.0001
14-21d	529 ^{b}	575 ^{a}	558 ^{ab}	547 ^{ab}	10.4	0.0129
21-28d	746	765	786	774	17.5	0.3553
0-28d	1608 ^{b}	1798 ^{a}	1780 ^{a}	1753 ^{a}	33.7	0.0012
Feed Conversion Ratio						
0-7d	1.36 ^{a}	1.19 ^{b}	1.24 ^{b}	1.24 ^{b}	0.035	0.0110
7-14d	1.46 ^{a}	1.36 ^{b}	1.36 ^{b}	1.36 ^{b}	0.024	0.0048
14-21d	1.46	1.49	1.49	1.45	0.020	0.3044
21-28d	1.69	1.74	1.70	1.62	0.031	0.0866
0-28d	1.53	1.56	1.50	1.48	0.028	0.2389

During the first week of the trial, animals from NC group displayed the lowest weight gains while broiler chicks of the Rov and GH16-1C groups displayed the best performances. In addition during this initial trial period GH16-3C broilers presented an intermediate weight gain, suggesting a lower efficacy resulting from enzyme assembly. Overall the data suggest that exogenous enzymes act predominantly during the initial period of the broilers grow (before day 14). This has previously been observed in other studies using barley based diets for poultry (Jozefiak *et al*., 2006) suggesting that has bird ages the capacity of the endogenous enzymatic repertoire is improved and can face the inherent increase in digesta viscosity resulting from the presence of high levels of β-glucans (see below). Data presented in Table 7.7 also revealed that supplemented animals presented higher feed intakes ($P<0.05$) when compared with non-supplemented animals and this difference predominate at

the initial periods of the trial. In addition, Feed Conversion Ratios (FCRs) were lower for animals of the supplemented groups during the first two weeks of the experiment. Thus, the better performance revealed by animals supplemented with microbial β-glucanases results not only from a higher feed intake but also from an improved efficacy of nutrient digestibility as reflected by the improved FCR revealed by animals receiving exogenous enzymes. The mortality rate of experiment was 1.25 % and was not related to treatments. Taken together the data suggest that exogenous enzymes improve the nutritive value of barley-based diets for broilers. This effect operates during the first two weeks of broiler grow. In addition, assembling of β-glucanases in mini-cellulosomes seems to restrict their initial efficacy as result of a lower enzyme flexibility imposed by cellulosome assembly upon the hydrolysis of a predominantly soluble substrate.

The relative weight and length of the bird's GI compartments, as well as digesta viscosity are presented in Table 7.8. The relative weight of the bird's GI compartments was affected by treatment for the ileum ($P<0.05$) and jejunum ($P<0.1$). Ileum and jejunum from supplemented animals were lighter than from animals not subjected to feed supplementation. In addition, the relative length of bird's duodenum and jejunum was affected by enzyme supplementation. Animals from supplemented groups had shorter duodenum and jejunum when compared with the non-supplemented animals. Several studies have reported a lower overall size of the GI tract in birds subjected to supplementation with exogenous plant cell wall degrading enzymes (Choct & Annison, 1990; Choct, 1997). Exogenous enzymes contribute to decrease digesta viscosity that is the main factor contributing to the enlargement of the birds' GI tract. Hence, data presented in Table 7.8, revealed that digesta viscosity in the duodenum+jejunum and ileum was different among birds fed the different diets. Digesta viscosity in the duodenum and jejunum, and in the ileum was lowest ($P<0.05$) for animals of the GH16-1C and GH16-3C groups. However, duodenum and jejunum and ileum viscosity was also significantly lower in the animals from Rov group in comparison with the non-supplemented animals, where the viscosity reached the highest values.

Table 7.8| Relative weight and length of gastrointestinal tract compartments and viscosity of digesta samples of broilers fed on a barley-based diet non-supplemented (NC) or supplemented with different exogenous enzymes. Rov, a commercial enzyme, GH16-1C, mini-cellulosome construct containing 1 cohesin, GH16-3C, mini-cellulosome construct containing 3 cohesins.

	NC	Rov	GH16-1C	GH16-3C	SEM	p(*F*)
Relative Weight (g/kg BW)						
Crop	3.08	3.47	3.79	3.59	0.232	0.1904
Gizzard	11.8 ^{b}	15.3 ^{a}	12.4 ^{b}	13.4 ^{ab}	0.84	0.0315
Liver	27.5	30.0	29.1	28.4	1.06	0.3890
Duodenum	10.5	10.4	9.02	8.70	0.861	0.3477
Jejunum	17.0	14.6	14.5	14.5	0.77	0.0637
Ileum	13.5 ^{a}	10.6 ^{b}	11.6 ^{b}	11.6 ^{b}	0.66	0.0339
Caecum	3.14	3.08	2.94	2.98	0.178	0.8539
Relative Length (cm/kg BW)						
Duodenum	24.5 ^{a}	20.1 ^{b}	22.1 ^{ab}	20.9 ^{b}	1.08	0.0366
Jejunum	58.4 ^{a}	52.6 ^{b}	52.4 ^{b}	50.2 ^{b}	1.95	0.0327
Ileum	59.9	53.6	53.7	55.1	2.39	0.2197
Caecum	14.8	14.2	13.2	13.3	0.65	0.3032
Content Viscosity (cP)						
Duodenum+ Jejunum	9.08 ^{a}	7.27 ^{b}	5.50 ^{c}	5.07 ^{c}	0.676	0.0001
Ileum	14.9 ^{a}	11.4 ^{a}	7.03 ^{b}	7.47 ^{b}	1.396	0.0005

7.2.3.4. Molecular integrity of mini-cellulosomes upon passage through the GI tract

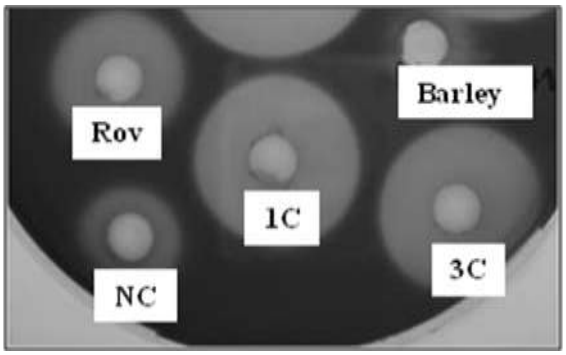
Data concerning the qualitative detection of β-glucanase activity in digesta samples collected in various GI compartments are presented in Table 7.9 and the qualitative β-glucanase activity present in the feed is present in Figure 7.7. The results revealed that β-glucanase activity could be detected in the crop and also in the caecum of animals of all treatments. However, only supplemented animals, in particular those of the GH16-1C and GH16-3C treatments, did presented significant levels of β-glucanase activity in the gizzard, duodenum, jejunum and ileum. This observation suggests an improved stability of *Ct*Glc16A when compared with the commercial enzymes upon passage through the GI tract. It is intriguingly that β-glucanase activity was detected in digesta samples collected in the crop of the non-supplemented animals. However, the intensity of the halos corresponding to β-glucanase crop activity of non-supplemented animals is considerably weaker when compared with the supplemented groups. Although a barley lot expressing lower levels of endogenous β-glucanase activity was selected for this study the sensitivity of the plate assay implemented to screen for exogenous enzymes was sufficient to detect it.

Table 7.9| Number of birds, out of 10 animals analyzed, fed on barley-based diet not supplemented (NC) or supplemented with different exogenous enzymes, presenting β-glucanase activity in digesta samples collected from various GI compartments. Rov, a commercial enzyme, GH16-1C, mini-cellulosome construct containing 1 cohesin, GH16-3C, mini-cellulosome construct containing 3 cohesins.

	NC	Rov	GH16-1C	GH16-3C	Chi-Square	P- value
Crop	9	7[1]	10	10	2.401	0.4934
Gizzard	0	2	9[2]	10	32.6	0.0001
Duodenum	0[2]	0	10	10	39.0	0.0001
Jejunum	2	0[2]	9[2]	10	31.5	0.0001
Ileum	0	0	10	10	40.0	0.0001
Caecum	8[1]	9[2]	8[1]	8[1]	-	-

[1] n=8; [2] n=9

Figure 7.7| Qualitative plate assay for detection of β-glucanase activity in the barley-based diet not supplemented (NC) or supplemented with a commercial β-glucanase or with the mini-cellulosomes GH16-1C or GH16-3C.



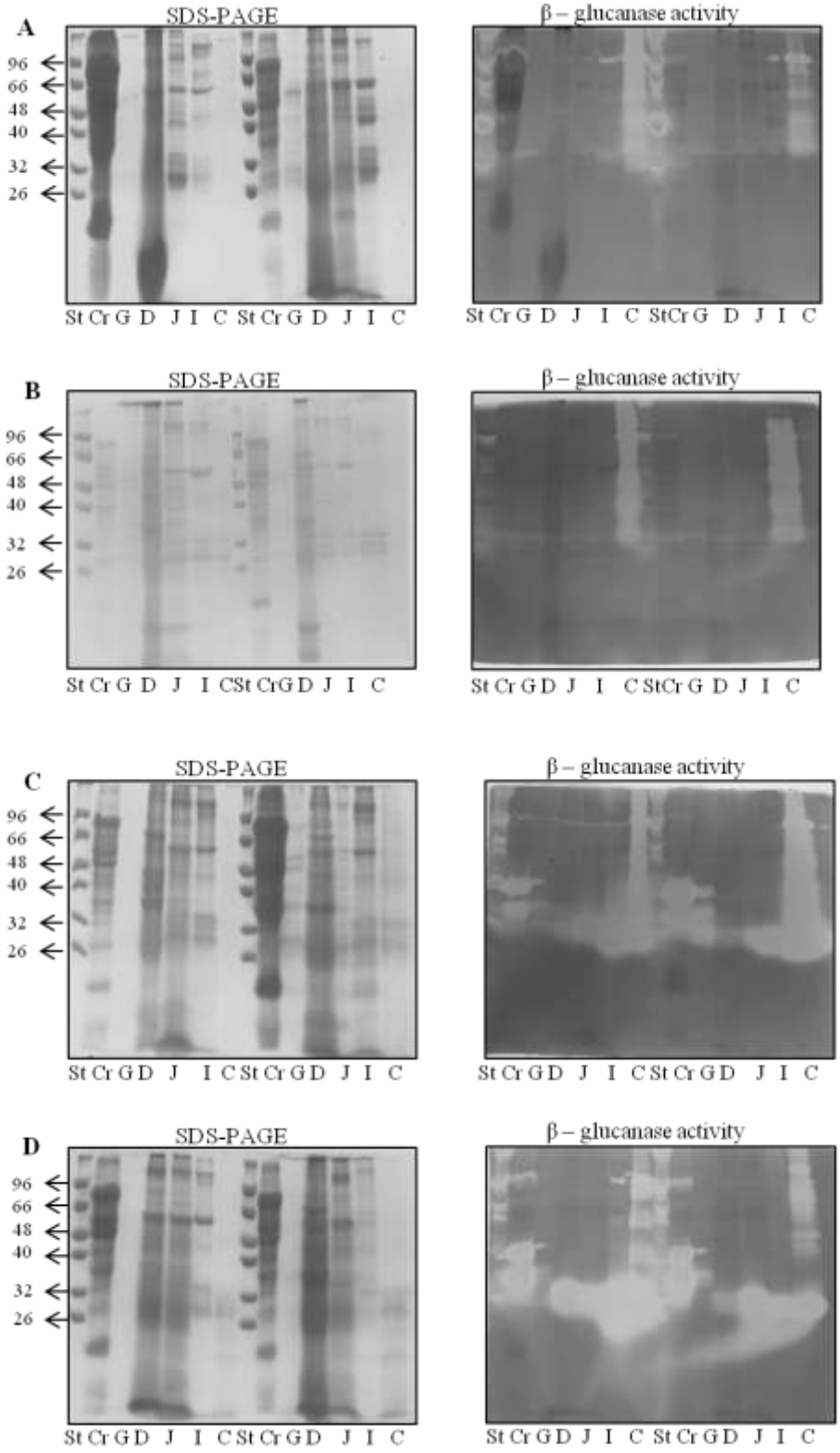
NC, barley-based diet not supplemented; Rov, barley-based diet supplemented with a commercial β-glucanase; 1C, barley-based diet supplemented with mini-cellulosome GH16-1C; 3C, barley-based diet supplemented with mini-cellulosome GH16-3C.

The integrity of exogenous enzymes during passage through the GI tract was verified through zymogram analysis. The data, presented in Figure 7.8, revealed that the caecum of all animals is populated with a large range of β-glucanases presenting molecular masses ranging from 20 to 100 kDa. In contrast, a limited number of polypeptides expressing β-glucanase activity are present in other portions of the GI tract. In particular, a prominent 96kDa protein is present in the majority of the digesta samples collected in both supplemented and non-supplemented animals and this might constitute an endogenous enzyme expressed in barley seeds. In addition, a second potentially endogenous but nonetheless less prominent polypeptide is identified in samples of the majority of the animals

and presents approximately 34kDa (Figure 7.8A). Digesta samples collected in the gizzard, duodenum, jejunum and ileum of supplemented animals displays the presence of highly active polypeptides. The majority of the β-glucanase activity expressed by birds supplemented with the commercial enzyme mixture seems to be represented by two polypeptides of 32 and 36 kDa, respectively (Figure 7.8B). These proteins are clearly visible in the small intestine of supplemented animals, although in some GI fractions they seem to be absent. This may result from a partial inactivation of the polypeptides due to pH denaturation or proteolysis.

With the exception of gizzard, zymogram analysis of digesta samples of animals supplemented with mini-cellulosome preparations revealed the presence of β-glucanase activity. Zymogram analysis is preceded by SDS-PAGE. During SDS-PAGE protein denaturation will lead to the de-assembling of mini-cellulosomes. Thus, zymogram analysis of mini-cellulosomes should reveal exclusively the polypeptide bands corresponding to the cellulosomal catalytic units. The data, presented in Figures 7.8C (GH16-1C) and 7.8D (GH16-3C), suggest that in the crop of animals receiving the GH16-1C and GH16-3C mini-cellulosomes, *Ct*Glc16A is present in two forms, with molecular masses of 28 kDa and 35 kDa, respectively. These two bands in the crop are present approximately in similar proportions. In contrast, when duodenum, jejunum and ileum digesta fractions are analyzed (Figures 7.8C and 7.8D), *Ct*Glc16A is observed exclusively in the 28 kDa form. *Ct*Glc16A used in this study is a bimodular enzyme containing an N-terminal GH16 β-glucanase catalytic domain and a C-terminal dockerin module, which is responsible for anchoring the enzyme into the cellulosome. The two modules are separated by a proline-rich linker region. Data presented in Figures 7.8C and 7.8D suggest that the linker sequence separating the two *Ct*Glc16A modules is prone to proteolysis in the avian GI tract, and the 28 kDa band corresponds in size to a *Ct*Glc16A truncated derivative lacking the C-terminal dockerin. Proteolytic cleavage of the bi-modular exogenous β-glucanase starts in the crop where only 50% of the enzyme was present in the intact form. Taken together data presented here suggests that proteolysis affects the integrity of mini-cellulosomes used to supplement barley-based diets for poultry by removing the dockerin sequence of *Ct*Glc16A most probably through cleavage of the linker region separating the two sequences.

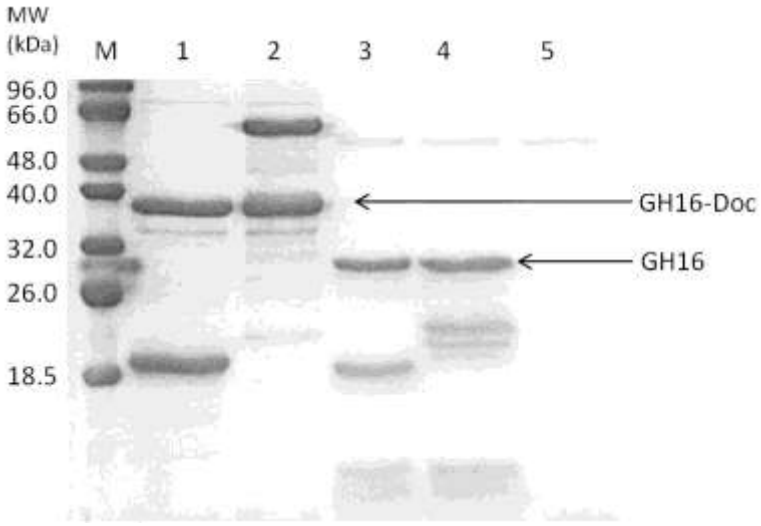
Figure 7.8| Zymogram analysis of digesta contents of broilers fed on a barley-based diet not supplemented or supplemented with different exogenous enzymes.



A) samples from birds fed on a diet non-supplemented; **B)** samples from birds fed on a diet supplemented with the commercial enzyme mixture; **C)** samples from birds fed on a diet supplemented with the GH16-1C; **D)** samples from birds fed on a diet supplemented with the GH16-3C. Proteins were fractionated through SDS-PAGE and stained for β-glucanase activity after enzyme renaturation. Abbreviations: St= Low molecular protein marker; Cr= crop; G= gizzard; D= duodenum; J= jejunum; I= ileum; C= caecum.

In order to confirm that endogenous proteases are contributing to unbound *Ct*Glc16A from mini-cellulosomes through the removal of the enzyme C-terminal dockerin, the two protein complexes were incubated *in vitro* with a pancreatic mixture of proteases (see materials and methods section for details). After 30 min incubation at 37 ºC the proteins were separated through SDS-PAGE to evaluate of mini-cellulosome integrity. The data, presented in Figure 7.9, confirms that *Ct*Glc16A is indeed susceptible to proteolysis and proteases cleave the full length enzyme to generate a truncated version of the recombinant β-glucanase consisting exclusively on the enzyme's catalytic domain (28 kDa). Linker regions play an important function in the efficacy of CAZymes and cellulosomes by contributing to the flexibility of the modular enzymes and the quaternary structure of protein complexes (Coutinho & Reilly, 1994). Glycosilation of protein linker regions is known to protect them from proteolytic cleavage (Tomme *et al*., 1995). *Ct*Glc16A was recombinantly produced in *Escherichia coli* that is unable to glycosylate proteins (Demain & Vaishnav, 2009; Kamionka, 2011). It is presently unknown if *Clostridium thermocellum* cellulosomal enzymes are glycosylated at their linkers. However, this study clearly indicates that protection of linker regions of cellulosomal enzymes from proteolytic cleavage is fundamental to maintain cellulosome integrity. Finally, cohesin domains seem to resist to the protease attack although mini-scaffoldin C3-C4-C5 was susceptible to proteolysis at its linkers (Figure 7.9, lane 4).

Figure 7.9| SDS-PAGE analysis of mini-cellulosomes GH16-1C and GH16-3 before and after incubation with a pancreatic mixture of proteases.



Lanes 1 and 2: mini-cellulosomes GH16-1C and GH16-3, respectively); lanes 3 and 4: mini-cellulosomes GH16-1C and GH16-3 after incubation with a pancreatic mixture of proteases, respectively. In lane 5 the pancreatic mixture of enzymes was analysed. M represents the low molecular protein marker.

7.2.4. Conclusions

In conclusion, this study reports the *in vivo* construction of two mini-cellulosomes based on scafoldins with 1 or 3 CipA-derived cohesin domains. The protein complexes were purified and protein stability studies revealed that *Ct*Glc16A biochemical properties are not significantly affected by protein complexation. However, the catalytic efficacy of GH16-3C mini-cellulosomes seems to be slightly reduced. This may represent a decrease in enzyme flexibility resulting from protein complexation. Protein flexibility is highly important for the hydrolysis of soluble carbohydrates. The two mini-cellulosomes are shown to be very effective in improving the nutritive value of barley-based diets for poultry. However, the efficacy of the recombinant nanomachines is not improved when compared with commercially available enzyme mixtures. Whether this represents a comparable efficacy of the three enzyme systems under analysis (commercial unbound enzymes, GH16-1C and GH16-3C) or a saturation of the animal feed with exogenous enzymes it remains to be established. In addition, endogenous proteases were shown to affect the integrity of the recombinant mini-cellulosomes by cleaving the linker region that separates *Ct*Glc16A catalytic and dockerin domains. Thus, in future studies protection of linker regions needs to be improved as the molecular integrity of mini-cellulosomes is critical to potentiate its use in a variety of biotechnological applications.

8. GENERAL DISCUSSION AND FUTURE PERSPECTIVES

The work presented in this thesis describes the discovery and characterization of novel carbohydrate-active enzymes (CAZymes) in anaerobic microbes producing cellulosomes, particularly in the thermophilic *Clostridium thermocellum* and in the important ruminal bacterium *Ruminococcus flavefaciens*. The former, although producing one of the best characterized cellulosomes to date, still displays many features that need to be elucidated concerning the complete repertoire of CAZymes that are organized in the multi-enzyme complex, and the last, producing one of the most intricate, and potentially versatile cellulosome described to date, represents an ideal resource for mining novel biocatalysts.

Here we have looked for open reading frames predicted to encode CAZymes bearing dockerins and thus constituting the catalytic members of bacterial cellulosome complexes. After this initial bioinformatic analysis we aimed at producing cellulosomal modules of unknown function that could support an enzyme discovery project. This was achieved through the implementation of novel high-throughput (HTP) methods for cloning, expression and purification of recombinant proteins. As a proof of concept that this approach would be effective to discover novel CAZymes in unexploited cellulosomes, the recombinant proteins of unknown function were screened for arabinofuranosidase activity. Chapter 3 reported the strategies developed for cloning 239 genes encoding cellulosomal modules of unknown function from *R. flavefaciens* (223 genes) and *C. thermocellum* (16 genes), and to express and purify their encoded proteins at high levels in *Escherichia coli*. A novel expression vector (pHTP) was constructed to afford cloning through a ligation-independent (LIC) system based on the complementation between nucleotides of single-stranded sequences from both the plasmid and the insert. The LIC-based method developed here is very efficient and allows obtaining high percentages of recombinant clones using vector to insert ratios of 1:2 to 1:10, even when large DNA fragments are intended to clone. The efficacy of the method is similar to that of other cloning systems that are commercially available. To support growth of *E. coli* to high cell densities, in order to obtain high-levels of recombinant protein expression from the *lac*-based promoter of pHTP vector, two different auto-induction media systems were developed. The media were tested for small, medium and large-scale production of recombinant proteins in different *E. coli* strains. The data revealed that recombinant protein overexpression was automatically induced as a result of serving lactose together with different metabolizable carbon sources other than glucose or glucose producing carbohydrates, to the recombinant bacteria. Similarly, a novel lysis buffer system was developed in this project and shown to be equally effective when compared with other commercially available systems and with the mechanical sonication of cells. In addition, cell lysis through the use of lysis buffers was shown to be more effective during the implementation of HTP projects, as it allowed the rapid disruption of dozens to hundreds of recombinant strains in a single step.

The novel HTP cloning, expression and protein purification platform was used to clone 239 genes and express them in *E. coli*. The data revealed that 184 cellulosomal proteins were produced in the soluble form (77% efficiency). The polypeptides were purified with yields of 60 µg to 2 mg of recombinant protein collected in a final volume of 0.15 mL of elution buffer, from a total culture volume of 5 mL. This yield was sufficient for the implementation of a functional screen based on a colorimetric substrate. In this work, we describe the identification of a novel α-L-arabinofuranosidase in *Ruminococcus flavefaciens* which constitutes the founding member of a novel CAZyme family. The novel CAZyme displays distant similarities with members of families 43 and 62 Glycoside Hydrolases (GH43 and GH62), thus belonging to Clan GH-F. The data confirms that modules of unknown function identified in cellulosomes could constitute novel CAZymes suggesting that other enzyme activities can be discovered in the pool of 184 proteins produced here if other enzyme screens are used. In addition, the HTP protocol described here could be adapted to recombinantly produce other groups of proteins of different origins and families and having different biological functions allowing the screening of novel enzyme activities and preliminary evaluation of their appropriateness for further structural and/or biochemical studies.

Although an effective system was developed in this work for the HTP cloning and expression of bacterial genes, it is well known that there are several difficulties related with the overexpression of recombinant proteins in *E. coli*. Among others, production of proteins in a soluble and active form remains, most of the times, unreliable to predict and constitutes a problem for some protein types. Thus, in Chapter 4 we described the generation of a comprehensive set of novel vectors to improve the levels of proteins expressed in the soluble form in *E. coli*. The vectors allow the fusion of the encoded recombinant protein to a protein/peptide partner (tag) that is known to be expressed at high levels and in the soluble form in *E. coli*. Although the fusion technology has already been extensively explored, the range of tags currently available is still limited while the efficacy of the different tags differs significantly with the protein type. The novel pHTP vector series created here contain engineered N-terminal fusion tags upstream of the vector´s cloning regions. The cloned genes will remain under the control of the T7 promoter and the tags selected for incorporating in the vectors were previously shown to be highly effective in raising levels of protein expression and solubility. In addition, in order to increase the repertoire of available tags, three novel fusion partners were tested in this work, including two recombinant proteins highly expressed in the soluble form in *E. coli* (Rf1 and Rf47, from Chapter 3) and a mini-cellulosome (CEL). Rf1 and Rf47 will allow the transcription of a single mRNA molecule encoding a soluble and a difficult to express protein to provide the required stability for efficient protein translation and at the same time protect the hydrophobic patches of the neighbour polypeptide, thus restricting the formation of inclusion bodies. The last system

(CEL) was developed to allow recombinant proteins to be integrated through an appended dockerin module into cohesin domains of the *C. thermocellum* CipA scaffoldin and thus precluding the formation of protein aggregates. After development of the different pHTP vector derivatives, a comparison study was implemented that compared the capacity of 12 different fusion systems to drive the expression of 8 different recombinant proteins in BL21(DE3). The cells were grown upon 5 different culture conditions. The data revealed that the recombinant proteins displayed different degrees of expression and solubility and that both the fusion partner and induction temperatures are the two major factors affecting these parameters. The percentage of proteins present in the soluble cell extract when using Rf1 and Rf47 as fusion partners was similar and higher than the percentage obtained with the remaining tags, including well-established solubility tags. Rf1 and Rf47 tags have an impressive performance at higher induction temperatures (37 ºC). Taken together the data suggest that the fusion tags may be ordered in the following order considering their capacity to enhance solubility of the fused proteins: Rf4 > Rf71 > Trx > MBP > NusA > His_6 > Fh8 > DsbA > SUMO = GST > CEL > Dsbc. A decrease of temperature during growth enhanced the percentage of soluble proteins, which is in agreement with previous data that suggest that a reduced rate of protein synthesis, conferring time for the proper fold and avoiding protein degradation, improves solubility (Chesshyre & Hipkiss, 1989; Spiess, Beil, & Ehrmann, 1999; Hunke & Betton, 2003; Sørensen & Mortensen, 2005). The use of either auto-induction media or the standard Luria-Bertani (LB) media had no effect in the capacity of *E. coli* to produce soluble recombinant proteins. Furthermore, the addition of IPTG to the auto-induction media at the mid-log phase did not improve the levels of soluble proteins. Protein purification allowed extrapolating both levels of protein solubility and protein expression and suggested that both Rf1 and R47 fusion tags perform among the best ones tested. Interestingly, spontaneous cleavage was observed for some large fusion systems after passage through the Ni^{2+}-columns, which could be attributed to the internal position of the affinity tag (His_6), between the two protein sequences (solubility tag and target protein). The CEL system tag failed to enhance protein solubility. Many factors could explain this result, including inadequate cohesin-dockerin interactions or even some conformation constraints during protein folding when cohesin-dockerin modules interact. However, the dockerin module alone could constitute a fusion partner to enhance protein solubility. Overall, these results are in close agreement with other comparative studies (Hammarström *et al*., 2002; Braun *et al*., 2002; Shih *et al*., 2002; Dyson *et al*., 2004; De Marco *et al*., 2004; Marblestone *et al*., 2006; Bird, 2011; Costa *et al*., 2013), and suggest that the efficiency of the fusion tags depends on the target protein. Thus, since there is no tag that display a high efficacy when applied to different types of proteins, the best combination of tag-fused protein needs to be experimentally determined. Taking this in consideration, we developed an ampicillin cloning vector, pHTP28, which will constitute the entry clone into the pHTP cloning

system. After cloning into this vector, the gene is sequenced to confirm that no mutations accumulated during the isolation of the nucleic acid. The LIC system supported by the pHTP vector series will allow the transfer of the gene from the entry clone into a range of different expression vectors, which are all kanamycin resistant, in a single step. After transfer into the destination vectors there is no need for further gene sequencing. Using this methodology a single gene can be transferred into a range of different vectors in a highly effective way and the most appropriate tag leading to the production of high levels of soluble protein easily identified and selected.

In Chapter 5, the effect of a CBM32 on the activity of one of the major cellulosomal enzymes of *C. thermocellum* ATCC 2740 (Cthe_0821, here named Man5A) was evaluated. Man5A is a typical modular enzyme comprising a signal peptide, a GH5 catalytic module, a family 32 CBM, and a C-terminal type I dockerin module. Some truncated versions of Man5A were engineered to generate GH5 protein derivatives with or without the CBM32 (rGH5-CBM32 and rGH5, respectively). The CBM32 module was also produced individually (rCBM32). Mannanase activity was evaluated for rGH5-CBM32 and rGH5 constructs while rCBM32 was evaluated for binding to mannans and glucomannans. Both rGH5-CBM32 and rGH5 were highly active toward konjac glucomannan and moderately effective toward β-1,4-D-mannan, ivory nut mannan, and carob galactomannan. No activity was detected against guar gum, either for rGH5-CBM32 or rGH5, suggesting that Man5A prefers unsubstituted mannans to highly galactosylated galactomannan. The crystalline structure of ivory nut mannan could explain the low activity of Man5A against this substrate, which is usually observed in GH5 mannanases. However, the enzyme activity toward ivory nut mannan and β-1,4-D-mannan (both insoluble polysaccharides) was higher in the presence of the CBM32 than when acting alone. Thus, CBM32 seems to play an important role in the hydrolysis of insoluble mannans by *C. thermocellum* Man5A. Besides its role in enhancing the proximity of the enzyme to the target substrate, CBM32 seems to decrease the resistance of the substrate to enzyme attack, as suggested by the weak biphasic action of rGH5-CBM32 toward the insoluble β-1,4-D-mannan when compared to the clear biphasic action showed by rGH5 against the same substrate. Activity-stimulating effects of CBMs toward insoluble substrates have been observed in several CAZymes (Maglione *et al*., 1992; Mangala *et al*., 2003; Sakka *et al*., 2011). rCBM32 showed specificity for the termini of mannans as confirmed by the limited retardation observed in the mannan-containing gels. The affinity profile of rCBM32 reflected the preference of the catalytic domains for less-decorated mannans once its electrophoretic mobility was not affected by the presence of guar gum, but was slightly retarded by inclusion of konjac glucomannan and carob galactomannan. ITC data suggested that rCBM32 contains two sugar binding sites that display a preference for mannose, although they can also recognize glucose-configured ligands. The preference of CBM32 from *C. thermocellum* Man5A for the non-reducing ends of β-manno-configured oligosaccharides was not

previously observed in family 32 CBMs. Thus, this work provides evidence that β-mannose-containing polymers should be included within the specificities known for this CBM family. In addition, the work provides novel insights into the capacity of CBMs to affect the activity of appended catalytic modules. *C. thermocellum* Man5A-CBM32 is presently commercially available from NZYTech. Other novel CAZymes identified in cellulosomal bacteria were investigated in this study and are also currently present in the Company portfolio. In Chapter 6, the biochemical properties of three *C. thermocellum* pectate lyases (PL) were evaluated aiming to investigate their role in the anaerobic conversion of biomass by *C. thermocellum* cellulosome. Homology searches using BLAST revealed that the protein ABN54148 includes a signal peptide followed by a putative family 1 PL (PL1A), a dockerin type I domain and a C-terminal family 6 CBM (CBM6). In addition, ABN53381 contains a signal peptide followed by a family 1 PL (PL1B), a dockerin type I domain, a family 35 CBM (CBM35) and a C-terminal family 9 PL (PL9). Although pectinases, in contrast to the majority of the other CAZymes, generally show a simple structure lacking CBMs, this work provides evidence that *C. thermocellum* cellulosome contains modular polysaccharide lyases belonging to PL families 1 and 9. The presence of CBMs in cellulosomal pectinases suggests that they potentiate the degradation of pectins. Prokaryotic pectinases containing cellulose-binding domains have previously been described (Pagès *et al*., 2003; McKie *et al*., 2001; Brown *et al*., 2001). More interestingly, families 6 and 35 CBMs display considerable promiscuity in ligand binding, and a recent work (Montanier *et al*., 2009) suggests that family 35 CBM may recognize the products of pectin hydrolysis. So the CBM found in the cellulosomal pectinases may bind the protein to the products of pectin hydrolysis thus targeting the enzymes to the regions of the cell wall that are being actively degraded, as pectins are of the initial targets of plant cell wall hydrolysis. All the three enzymes (PL1A, PL1B and PL9) were highly active towards polygalacturonic acid (PGA) and pectin (up to 55% methyl-esterified) both from citrus. Relatively high activities were observed for PL1B against 55% and 85% methyl-esterified pectins from citrus, and for PL1A and PL9 against rhamnogalacturonan from potato (RGAP) and soybean (RGAS). The data showed that PL1A, PL1B and PL9 are characteristic pectate lyases, as they catalyse the β-eliminative cleavage of glycosidic bonds with the production of Δ4,5 unsaturated galacturonates. An endo-cleavage pattern on PGA and pectin (citrus) was observed for all the three enzymes, by producing unsaturated di, tri and higher oligogalacturonates. From the current 23 families of polysaccharide lyases present in the CAZy data base (www.cazy.org), only families 1, 2, 3, 9 and 10 include pectate lyases. Thus, for the first time three cellulosomal pectinolytic enzymes from *C. thermocellum* were identified and characterized. However, the high heterogeneity found in pectic substrates evidences that within cellulosomes other enzymes presently of unknown function may target pectic polysaccharides. It is clear that the degradation of plant cell wall by cellulosomes implies a vast repertoire of distinct cell wall-degrading enzymes to achieve the hydrolysis of

the diverse structural carbohydrates. The possibility to use these enzymes in different industrial and agricultural processes is remarkable.

The experiments described in Chapter 7 explored the utilization of CAZymes in poultry nutrition. In recent years, the utilization of CAZymes, primarily glycoside hydrolases, carbohydrate esterases and pectate lyases, in poultry feed has increased significantly. This is mainly due to the rising costs of traditional raw-materials and the search for alternative ingredients (e.g. barley, oats, rice and wheat) that could improve the economic efficacy of feed production. However, a great majority of alternative feed ingredients contain significant levels of antinutritional factors that contribute to decrease their intrinsic nutritive value (Charlton, 1996). Since simple stomach animals do not produce an endogenous repertoire of enzymes to degrade plant cell wall structural polysaccharides, there is a need to integrate exogenous biocatalytic activities directly in the diets through the provision of microbial CAZymes. It was previously suggested that the use of a broad range of enzymes with different substrate specificities to ensure a large spectrum of action may be unnecessary in several cases since degradation of the antinutritive soluble non-starch polysaccharides (NSPs) can result from the action of individual enzymes expressing the required specificity (Ribeiro *et al.*, 2012). Thus, in Chapter 7, we investigated if the improvement of the nutritive value of barley-based diets for broilers results from the action of highly specific β-1,3-1,4-glucanases or β-1,4-glucanases displaying a broad substrate specificity that includes cleavage of β-1,3-1,4-glucans present in barley. The capacity of two *C. thermocellum* enzymes, β-1,3-1,4-glucanase 16A, termed *Ct*Glc16A, and β-1,4-glucanase 8A, termed *Ct*Cel8A, to improve the nutritive value of these diets was compared *in vivo*. The data suggested that although retaining its molecular integrity and catalytic activity during passage through the bird's gastro-intestinal (GI) tract, contrary to *Ct*Glc16A, *Ct*Cel8A did not contribute to improve the performance of broilers. Nevertheless, *in vitro* experiments revealed that *Ct*Cel8A effectively contributes to reduce the viscosity of a β-1,3-1,4-glucan preparation similarly to what was observed to *Ct*Glc16A. In addition, differences in the capacity of the recombinant exogenous enzymes to improve the nutritive value of barley-based diets cannot be merely explained by differences in enzyme stability and integrity. Thus, we postulated that the presence of β-1,4-glucans in barley based-diets may reduce the efficacy of *Ct*Cel8A to cleave β-1,4-linkages within β-1,3-1,4-glucans. This phenomenon was confirmed *in vitro* by observing a significant reduction of the β-1,3-1,4-glucanase activity of *Ct*Cel8A in the presence of the animal feed (only 15% of its β-1,3-1,4-glucanase activity was retained), compared to that for *Ct*Glc16A (β-1,3-1,4-glucanase activity reduced by only 54%). Thus, β-1,4-glucanase *Ct*Cel8A seems to be unable to affect the nutritive value of barley-based diets by undertaking non-productive interactions with its preferred substrate, cellulose. These observations suggest that current commercial enzyme mixtures used in poultry nutrition could be optimized by estimating the contribution of β-1,4-glucanases to the overall

β-1,3-1,4-glucanase activity expressed. Enzyme mixtures with predominant β-1,3-1,4-glucanases are most adequate to supplement barley-based diets.

Chapter 7 explored the use of mini-cellulosomes expressing β-1,3-1,4-glucanase activity for supplementation of barley-based diets for poultry. It is known that cellulosomal enzymes are stabilized once dockerins binds their complementary cohesin (García-Alvarez *et al*., 2011). Thus, an animal trial was conducted with one (GH16-1C) or three (GH16-3C) copies of *C. thermocellum* Glucanase 16A (*Ct*Glc16A), each of which appended to a *C. thermocellum* CipA cohesin via its dockerin domain. The data revealed that although protein complexation had no effect in the biochemical properties of *Ct*Glc16A, the catalytic activity of the cellulosomal enzyme seems to be slighted reduced in GH16-3C during the *in vitro* hydrolysis of soluble barley β-glucan. A restriction of *Ct*Glc16A flexibility upon its organization of mini-cellulosomes could explain this observation. Animals submitted to GH16-3C treatment had however similar body weights to animals from the GH16-1C group, though lower than animals supplemented with the commercial enzyme. In addition, supplemented animals presented higher feed intakes and lower Feed Conversion Ratios (FCRs) when compared with non-supplemented animals. A lower overall size of the GI tract in birds subjected to supplementation with exogenous enzymes was reported. In general, the data suggested that the two mini-cellulosomes were effective in improving the nutritive value of barley-based diets for poultry. However, their efficacy is still lower than the commercially available enzyme mixtures. In addition, the integrity of the recombinant mini-cellulosomes was affected during passage through the GI tract, due the presence of endogenous proteases. Thus, data presented here revealed that linker regions between the catalytic and dockerin domains constitute a target for protease attack. This work clearly suggests that further research should be achieved in order to develop mechanisms to protect the susceptible linker regions in mini-cellulosomes that might be used not only for feed supplementation but also in a variety of other biotechnological applications.

In conclusion, this thesis describes a totally novel approach for the discovery of novel CAZymes in a pool of dozens to thousands of open reading frames encoding proteins of unknown function. A HTP platform was developed while a novel series of cloning and expression vectors were constructed to support simple and automatable cloning methods coupled with high levels of expression of soluble recombinant proteins. In addition, novel biochemical capacities in cellulosomal CAZymes and CBMs were described in detailed here.

Future work should address the possibility of screening the proteins produced here for other enzymatic or CBM activities. In addition, studies in the pHTP expression vectors should be extended, in particular to analyse the effect of tag removal in recombinant protein stability and the efficacy of different cleavage systems. Furthermore, it is expected that next large-scale experiments include only three or four fusion partners that showed the best results in the small-scale screening.

9. BIBLIOGRAPHIC REFERENCES

Abbott, D. W., Hrynuik, S., & Boraston, A. B. (2007). Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterolitica*. *Journal of Molecular Biology*, *367*, 1023–1033.

Abdullah, J. M., Joachimiak, A., & Collart, F. R. (2009). "System 48" High-throughput cloning and protein expression analysis. *Methods in Molecular Biology*, *498*, 117–127.

Adams, J. J., Webb, B. A., Spencer, H. L., & Smith, S. P. (2005). Structural characterization of type II dockerin module from the cellulosome of *Clostridium thermocellum*: calcium-induced effects on conformation and target recognition. *Biochemistry*, *44*(6), 2173–2182.

Adams, E. A., & Pough, R. (1993). Non-starch polysaccharides and their digestion in poultry. *Feed Compounder*, *13*, 19–21.

Aden, A., Ruth, M., Ibsen, K., Jechura, J., Neeves, K., Sheehan, J., … Lukas, J. (2002). Lignocellulosic biomass to ethanol process design and economics utilizing co-current dilute acid prehydrolysis and enzymatic hydrolysis for corn stover. *National Renewable Energy Laboratory*, (June).

Adeola, O., & Cowieson, A. J. (2011). BOARD-INVITED REVIEW: Opportunities and challenges in using exogenous enzymes to improve nonruminant animal production. *Journal of Animal Science*, *89*, 3189–3218.

Alahuhta, M., Xu, Q., Brunecky, R., Adney, W. S., Ding, S.-Y., Himmel, M. E., & Lunin, V. V. (2010). Structure of a fibronectin type III-like module from *Clostridium thermocellum*. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications*, *66*(Pt 8), 878–880.

Albala, J. S., Franke, K., McConnell, I. R., Pak, K. L., Folta, P. A., Rubinfeld, B., … Clark, R. (2000). From genes to proteins: high-throughput expression and purification of the human proteome. *Journal of Cellular Biochemistry*, *80*(2), 187–91.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–402.

Anderson, V. E. (1998). *Comprehensive Biological Catalysis: A Mechanistic Reference*. (M. Sinnnot, Ed.) (Vol. 2., pp. 115–133). London: Academic Press.

Arai, T., Araki, R., Tanaka, A., Karita, S., Kimura, T., Sakka, K., & Ohmiya, K. (2003). . Characterization of a cellulase containing a family 30 carbohydrate-binding module (CBM) derived from *Clostridium thermocellum* CelJ: Importance of the CBM to cellulose hydrolysis. *Journal of Bacteriology*, *185:504 –5*, 504 –512.

Arechaga, I., Miroux, B., Runswick, M. J., & Walker, J. E. (2003). Over-expression of Escherichia coli F1Fo-ATPase subunit a is inhibited by instability of the *uncB* gene transcript. *FEBS Letters*, *547*, 97–100.

Aristidou, A., & Merja, P. (2000). Metabolic engineering applications to renewable resource utilization. *Current Opinion in Biotechnology*, *11*, 187–198.

Aslanidis, C., de Jong, P. J., & Schmitz, G. (1994). Minimal length requirement of the single-stranded Tails for ligation-independent cloning (LIC) of PCR products. *Genome Research*, (4), 172–177.

Attala, R. H., Vanderhart, D. L. (1989). Studies on the structure of cellulose using Raman spectroscopy and solid state (13)C NMR. In *Cellulose and Wood: chemistry and technology, Proceedings of the tenth Cellulose Conference (C. Schuerch, ed.)* (pp. 169–187). New York: John Wiley and sons.

Bahari, L., Gilad, Y., Borovok, I., Kahel-Raifer, H., Dassa, B., Nataf, Y., … Bayer, E. A. (2011). Glycoside hydrolases as components of putative carbohydrate biosensor proteins in *Clostridium thermocellum*. *Journal of Industrial Microbiology & Biotechnology*, *38*(7), 825–32.

Baker, R. T. (1996). Protein expression using ubiquitin fusion and cleavage. *Current Opinion in Biotechnology*, *7*, 541–546.

Balbás, P. (2001). Understanding the art of producing protein and nonprotein molecules in Escherichia coli. *Molecular Biotechnology*, *19*(3), 251–67.

Baneyx, F. (1999a). *In vivo* folding of recombinant proteins in *Escherichia coli*. In D. C. Davies J. E., Demain, A. L., Cohen, G., Hershberger, C. L., Forney, L. J., Holland, I. B., Hu, W-S, Wu J-HD, Sherman, D. H., Wilson R. C. Washington (Ed.), *Manual of Industrial Microbiology and Biotechnology* (2nd ed., pp. 551–565). American Society for Microbiology.

Baneyx, F. (1999b). Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*, *10*, 411–421.

Barras, F., Van Gigsegem, F., & Chatterjee, A. K. (1994). Extracellular enzymes and pathogenesis of soft-rot Erwinia. *Annual Review of Phytopathology*, *32*, 201–234.

Bayer, E. A, Belaich, J.-P., Shoham, Y., & Lamed, R. (2004). The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology*, *58*, 521–54.

Bayer, E. A, & Lamed, R. (1986). Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *Journal of Bacteriology*, *167*(3), 828–36.

Bayer, E. A, Lamed, R., White, B. A., & Flint, H. J. (2008). From cellulosomes to cellulosomics. *Chemical Record (New York, N.Y.)*, *8*(6), 364–77.

Bayer, E. A, Setter, E., & Lamed, R. (1985). Organization and distribution of the cellulosome in *Clostridium thermocellum*. *Journal of Bacteriology*, *163*(2), 552–9.

Bayer, E. A., Kening, R., & Lamed, R. (1983). Adherance of *Clostridium thermocellum* to cellulose. *Journal of Bacteriology*, *156*, 818–827.

Bedford, M. R. (2000). Exogenous enzymes in monogastric nutrition - their current value and future benefitd. *Animal Feed Science and Technology*, *86*, 1–13.

Bedford, M. R., & Cowieson, A. J. (2009). Phytate and Phytase Interactions. In *17th Eur. Symp. Poult. Nutr*. Edinburgh, Scotland: The World's Poultry Science Association European Federation-UK Branch.

Bedford, M. R., & Morgan, A. J. (1996). The use of enzymes in poultry diets. *World Poultry Science*, *52*, 61–68.

Béguin, P., & Aubert, J. P. (1994). The biological degradation of cellulose. *FEMS Microbiology Reviews*, *13*(1), 25–58.

Béguin, P., & Lemaire, M. (1996). The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. *Critical Reviews in Biochemistry and Molecular Biology*, *31*(3), 201–36.

Bernard, P. (1996). Positive selection of recombinant DNA by CcdB. *BioTechniques*, *21*(2), 320–323.

Berrow, N. S., Alderton, D., & Owens, R. J. (2009). The precise engineering of expression vectors using high-throughput In-Fusion PCR cloning. *Methods in Molecular Biology*, *498*, 75–90.

Bessette, P. H., Aslund, F., Beckwith, J., & Georgiou, G. (1999). Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 13703–8.

Bird, L. E. (2011). High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. *Methods (San Diego, Calif.)*, *55*(1), 29–37.

Blackwell, J. R., & Horgan, R. (1991). A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Letters*, *295*(1-3), 10–2.

Blommel, P. G., & Fox, B. G. (2007). A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expression and Purification*, *55*(1), 53–68.

Blum, D. L., Kataeva, I. A., Li, X. L., & Ljungdahl, L. G. (2000). Feruloyl esterase activity of the Clostridium thermocellum cellulosome can be attributed to previously unknown domains of XynY and XynZ. *Journal of Bacteriology*, *182*, 1346–1351.

Bolam, D. N., Ciruela, A., Mcqueen-mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., … Gilbert, H. J. (1998). Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochemical Journal*, *331*, 775–781.

Bolam, D. N., Xie, H., Pell, G., Hogg, D., Galbraith, G., Henrissat, B., & Gilbert, H. J. (2004). X4 modules represent a new family of carbohydrate-binding modules that display novel properties. *The Journal of Biological Chemistry*, *279*(22), 22953–22963.

Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical Journal*, *382*(Pt 3), 769–81.

Boraston, A. B., Kwan, E., Chiu, P., Warren, R. A. J., & Kilburn, D. G. (2003). Recognition and hydrolysis of noncrystalline cellulose. *The Journal of Biological Chemistry*, *278*(8), 6120–7.

Boraston, A. B., McLean, B. W., Kormos, J. M., Alam, M., Gilkes, N. R., Haynes, C. A., Tomme, P., Kilburn, D. G. & Warren, R. A. J. (1999). Carbohydrate-binding modules: diversity of structure and function. In E. Gilbert, H. J., Davies, G. J., Henrissat, B. and

Svensson, B. (Ed.), *Recent Advances in Carbohydrate Bioengineering* (pp. 202–211). Cambridge: Royal Society of Chemistry.

Bothast, R. J. (1998). Purification and characterization of a novel thermostable α-L-arabinofuranosidase from a color-variant strain of *Aureobasidium pullulans*. *Applied and Environmental Microbiology*, *64*(1), 216–220.

Brás, J. L. A., Alves, V. D., Carvalho, A. L., Najmudin, S., Prates, J. A. M., Ferreira, L. M. A, … Fontes, C. M. G. A. (2012). Novel *Clostridium thermocellum* type I cohesin-dockerin complexes reveal a single binding mode. *The Journal of Biological Chemistry*, *287*(53), 44394–405.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., & LaBaer, J. (2002). Proteome-scale purification of human proteins from bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(5), 2654–9.

Britton, H. T. K., & Robinson, R. A. (1931). Universal buffer solutions and the dissociation constant of veronal. *Journal of the Chemical Society*, *01931*, 1456 –1462.

Brown, I. E., Mallen, M. H., Charnock, S. J., Davies, G. J., & Black, G. W. (2001). Pectate lyase 10A from *Pseudomonas cellulosa* is a modular enzyme containing a family 2a carbohydrate-binding module. *Biochemical Journal*, *355*, 155–165.

Bruhlmann, F., & Keen, N. T. (1997). Cloning, sequence and expression of the pel gene from an Amycolata sp. *Gene*, *201*, 45–51.

Bruni, R., & Kloss, B. (2013). High-throughput cloning and expression of integral membrane proteins in *Escherichia coli*. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* (Vol. 74, p. Unit 29.6.).

Bryksin, A. V, & Matsumura, I. (2010). Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques*, *48*(6), 463–5.

Bucher, M. H., Evdokimov, A. G., & Waugh, D. S. (2002). Differential effects of short affinity tags on the crystallization of *Pyrococcus furiosus* maltodextrin-binding protein. *Acta Crystallographica Section D Biological Crystallography*, *58*(3), 392–397.

Butt, T. R., Edavettal, S. C., Hall, J. P., & Mattern, M. R. (2005). SUMO fusion technology for difficult-to-express proteins. *Protein Expression and Purification*, *43*(1), 1–9.

Campbell, J. M., Fahey, G. C., & Wolf, B. W. (1997). Selected indigestible oligosaccharides affect large bowel mass, cecal and fecal short-chain fatty acids, pH and microflora in rats. *Journal of Nutrition*, *127*, 130–136.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, *37*, 233–238.

Cardoso, V., Ferreira, A. P., Costa, M., Ponte, P. I. P., Falcão, L., Freire, J. P., … Ribeiro, T. (2014). Temporal restriction of enzyme supplementation in barley-based diets has no effect in broiler performance. *Animal Feed Science and Technology*, *in press*.

Carlson, E. D., Gan, R., Hodgman, C. E., & Jewett, M. C. (2011). Cell-free protein synthesis: applications come of age. *Biotechnology Advances*, *30*(5), 1185–94.

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

Carré, B. (1993). European Symposium Podtry Nutrition, WPSA, 5-9 September (pp. 148–163). Jelena, Poland.

Carrió, M. M., & Villaverde, A. (2002). Construction and deconstruction of bacterial inclusion bodies. *Journal of Biotechnology*, *96*(1), 3–12.

Carvalho, A. L., Dias, F. M. V, Nagy, T., Prates, J. A. M., Proctor, M. R., Smith, N., … Gilbert, H. J. (2007). Evidence for a dual binding mode of dockerin modules to cohesins. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3089–3094.

Carvalho, A. L., Dias, F. M. V, Prates, J. A. M., Nagy, T., Gilbert, H. J., Davies, G. J., … Fontes, C. M. G. A. (2003). Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(24), 13809–13814.

Carvalho, A. L., Goyal, A., Prates, J. A. M., Bolam, D. N., Gilbert, H. J., Pires, V. M. R., … Fontes, C. M. G. A. (2004). The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates beta-1,4- and beta-1,3-1,4-mixed linked glucans at a single binding site. *The Journal of Biological Chemistry*, *279*(33), 34785–34793.

Carvalho, A. L., Pires, V. M. R., Gloster, T. M., Turkenburg, J. P., Prates, J. A. M., Ferreira, L. M. A., … Gilbert, H. J. (2005). Insights into the structural determinants of cohesin-dockerin specificity revealed by the crystal structure of the type II cohesin from *Clostridium thermocellum* SdbA. *Journal of Molecular Biology*, *349*(5), 909–915.

Cha, J., Matsuoka, S., Chan, H., Yukawa, H., Inui, M., & Doi, R. H. (2007). Effect of multiple copies of cohesins on cellulase and hemicellulase activities of *Clostridium cellulovorans* mini-cellulosomes. *Journal of Microbiology and Biotechnology*, *17*(11), 1782–1788.

Chambers, S. P. (2002). High-throughput protein expression for the post-genomic era. *Drug Discovery Today*, *7*(14), 759–65.

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

Chen, X., Zhou, L., Tian, K., Kumar, A., Singh, S., Prior, B. a., & Wang, Z. (2013). Metabolic engineering of *Escherichia coli*: A sustainable industrial platform for bio-based chemical production. *Biotechnology Advances*, *31*(8), 1200–1223.

Chesshyre, J. A., & Hipkiss, A. R. (1989). Low temperatures stabilize interferon a-2 against proteolysis in Methylophilus methylotrophus and *Escherichia coli*. *Applied Biochemistry and Biotechnology*, *31*, 158–162.

Choct, M. (1997). Feed Non-Starch Polysaccharides: Chemical Structures and Nutritional Significance. *Feed Milling International*, (June Issue), 13–26.

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

Choct, M., & Annison, G. (1990). Anti-nutritive activity of wheat pentosans in broiler diets. *British Poultry Science*, *30*, 811–821.

Choct, M., & Annison, G. (1992). Anti-nutritive activity of wheat arabinoxylans: Role of viscosity. *British Poultry Science*, *33*, 821–834.

Choct, M., Hughes, R. J., Wang, J., Bedford, M. R., Morgan, A. J., & Annison, G. (1996). Increased small intestinal fermentation is partly responsible for the anti-nutritive activity of non-starch polysaccharides in chickens. *British Poultry Science*, *37*, 609–621.

Choi, S. K., & Ljungdahl, L. G. (1996). Structural role of calcium for the organization of the cellulosome of *Clostridium thermocellum*. *Biochemistry*, *35*, 4906–4910.

Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., … Xu, M. Q. (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, *192*(2), 271–81.

Chou, C. P. (2007). Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. *Applied Microbiology and Biotechnology*, *76*, 521–532.

Coleman, T. a, Parmelee, D., Thotakura, N. R., Nguyen, N., Bürgin, M., Gentz, S., & Gentz, R. (1997). Production and purification of novel secreted human proteins. *Gene*, *190*(1), 163–71.

Collins-Racie, L. A., McColgan, J. M., Grant, K. L., DiBlasio-Smith, E. A., McCoy, J. M., LaVallie, E. R. (1995). Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner DsbA. *Biotechnology (N. Y.)*, *13*(9), 982–987.

Collmer, A., Riad, J. L., & Mount, M. S. (1988). Assay methods for pectic enzymes. *Methods in Enzymology*, *161*, 329–335.

Colwill, K., Wells, C. D., Elder, K., Goudreault, M., Hersi, K., Kulkarni, S., … Morin, G. B. (2006). Modification of the Creator recombination system for proteomics applications-improved expression by addition of splice sites. *BMC Biotechnology*, *6*, 13.

Cornet, P., Millet, J., Béguin, P., & Aubert, J.-P. (1983). Characterization of two Cel (cellulose degradation) genes of *Clostridium thermocellum* coding for Endoglucanases. *Nature Biotechnology*, *1*, 589 – 594.

Cosgrove, D. J. (1997). Assembly and enlargement of the primary cell wall in plants. *Annual Review of Cell and Developmental Biology*, *13*, 171–201.

Cosgrove, D. J. (2005). Growth of the plant cell wall. *Nature Reviews. Molecular Cell Biology*, *6*(11), 850–61.

Costa, S., Almeida, A., Castro, A., & Domingues, L. (2014). Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Frontiers in Microbiology*, *5*(February),

Costa, S. J., Almeida, A., Castro, A., Domingues, L., & Besir, H. (2013). The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. *Applied Microbiology and Biotechnology*, *97*(15), 6779–91.

Cote, G. L., & Leathers, T. D. (2005). A method for surveying and classifying *Leuconostoc* sp. Glucansucrases according to strain-dependent acceptor product patterns. *Journal of Industrial Microbiology and Biotechnology*, *32*, 53–60.

Coutinho, P. M. & Reilly, P. J. (1994). Structure–function relationships in the catalytic and starch binding domains of glucoamylase. *Protein Engineering*, *7*(3), 393–400.

Creighton, T. E. (1986). Detection of folding intermediates using urea-gradient electrophoresis. *Methods in Enzymology*, *131*, 156–172.

Creighton, T. E. (1997). How important is the molten globule for correct protein folding? *Trends in Biochemical Sciences*, *22*(1), 9–10.

Creswell, D. C. (1994). *Upgrading the nutritional value of grains with the use of enzymes. Technical bulletin, American Soybean Association* (341 Orchar.). Singapore.

Czjzek, M., Bolam, D. N., Mosbah, a, Allouch, J., Fontes, C. M., Ferreira, L. M., … Gilbert, H. J. (2001). The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved. *The Journal of Biological Chemistry*, *276*(51), 48580–7.

Daas, P. J. H., Grolle, K., van Vliet, T., Schols, H. A., & de Jongh, H. H. J. (2002). Toward the recognition of structure-function relationships in galactomannans. *Journal of Agricultural and Food Chemistry*, *50*, 4282– 4289.

Darvill, A. G., McNeil, M., & Albersheim, P. (1978). Structure of plant cell walls VIII. A new pectic polysaccharide. *Plant Physiology*, *62*, 418–422.

Das, S. P., Ravindran, R., Ahmed, S., Das, D., Goyal, D., Fontes, C. M. G. A, & Goyal, A. (2012). Bioethanol production involving recombinant *C. thermocellum* hydrolytic hemicellulase and fermentative microbes. *Applied Biochemistry and Biotechnology*, *167*(6), 1475–88.

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

Davies, G. J., Gloster, T. M., & Henrissat, B. (2005). Recent structural insights into the expanding world of carbohydrate-active enzymes. *Current Opinion in Structural Biology*, *15*(6), 637–45.

Davis, G. D., Elisee, C., Newham, D. M., Harrison, R. G. (1999). New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnology and Bioengineering*, *8*, 1668–1674.

De Marco, A., Vigh, L., Diamant, S., & Goloubinoff, P. (2005). Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. *Cell Stress & Chaperones*, *10*(4), 329–39.

De Marco, V., Stier, G., Blandin, S., & de Marco, A. (2004). The solubility and stability of recombinant proteins are increased by their fusion to NusA. *Biochemical and Biophysical Research Communications*, *322*(3), 766–71.

De Vries, R. P., & Visser, J. (2001). Aspergillus enzymes involved in degradation of plant cell wall polysaccharides. *Microbiology and Molecular Biology Reviews*, *65*, 497–522.

DeBoy, R. T., Mongodin, E. F., Fouts, D. E., Tailford, L. E., Khouri, H., Emerson, J. B., … Nelson, K. E. (2008). Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. *Journal of Bacteriology*, *190*(15), 5455–5463.

Demain, A. L., Newcomb, M., & Wu, J. H. D. (2005). Cellulase, clostridia, and ethanol. *Microbiology and Molecular Biology Reviews*, *69*(1), 124–154.

Demain, A. L., & Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances*, *27*(3), 297–306.

Derman, A. I., Prinz, W. A., & Beckwith, J. (1993). Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science*, *262*(5140), 1744–1747.

Di Guana, C., Lib, P., Riggsa, P. D., Hiroshi, I. (1988). Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene*, *67*(1), 21–30.

Dieckman, L., Gu, M., Stols, L., Donnelly, M. I., & Collart, F. R. (2002). High-throughput methods for gene cloning and expression. *Protein Expression and Purification*, *25*(1), 1–7.

Ding, S., Rincon, M. T., Lamed, R., Jennifer, C., Mccrae, S. I., Aurilia, V., … Crae, S. I. M. C. (2001). Cellulosomal Scaffoldin-Like Proteins from *Ruminococcus flavefaciens*. *Journal of Bacteriology*, *183*(6), 1945–1.

Doi, R. H., Kosugi, A., Murashima, K., Tamaru, Y., & Han, S. O. (2003). Cellulosomes from mesophilic bacteria. *Journal of Bacteriology*, *185*(20), 5907–5914.

Dong, H., Nilsson, L., & Kurland, C. G. (1996). Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *Journal of Molecular Biology*, *260*(5), 649–63.

Dubendorf, J. W., & Studier, F. W. (1991). Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *Journal of Molecular Biology*, *219*(1), 45–59.

Dunnington, E. A., & Siegel, P. B. (1995). Enzyme activity and organ development in newly hatched chicks selected for high or low eight-week body weight. *Poultry Science*, *74*, 761–770.

Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L., & McCafferty, J. (2004). Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC Biotechnology*, *4*, 32.

Edwards, C. A., Johnson I. T., & Read, N. W. (1988). Do viscous polysaccharides slow absorption by inhibiting diffusion or convection? *European Journal of Clinical Nutrition*, *42*(4), 307–312.

Eggers, C. T., Murray, I. a, Delmar, V. a, Day, A. G., & Craik, C. S. (2004). The periplasmic serine protease inhibitor ecotin protects bacteria against neutrophil elastase. *The Biochemical Journal*, *379*(Pt 1), 107–18.

Eisenberg, D., Marcotte, E. M., Xenarios, I., & Yeates, T. O. (2000). Protein function in the post-genomic era. *Nature*, *405*(6788), 823–826.

Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annual Review of Biophysics and Biomolecular Structure*, *29*, 213–238.

Epstein, W., Rothman-Denes, L. B., & Hesse, J. (1975). Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(6), 2300–4.

Fan, L. H., Zhang, Z. J., Yu, X. Y. ., Xue, Y. X., & Tan, T. W. (2012). Self-surface assembly of cellulosomes with two miniscaffoldins on *Saccharomyces cerevisiae* for cellulosic ethanol production. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(33), 13260–13265.

Fan, L. H., Zhang, Z. J., Yu, X. Y., Xue, Y. X., Wang, M. M., & Tan, T. W. (2013). *In vitro* assembly of minicellulosomes with two scaffoldins on the yeast cell surface for cellulose saccharification and bioethanol production. *Process Biochemistry*, *48*(3), 430–437.

Fernandes, A. C., Fontes, C. M. G. A., Gilbert, H. J., Hazlewood, G. P., Fernandes, T. H., & Ferreira, L. M. A. (1999). Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. *Biochemical Journal*, *110*, 105–110.

Ferrer, M., Chernikova, T. N., and Yakimov, M. M. (2003). Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nature Biotechnology*, *50*, 1579–1589.

Ficko-Blean, E., & Boraston, A. B. (2006). The interaction of a carbohydrate-binding module from a *Clostridium perfringens* N-acetyl-beta-hexosaminidase with its carbohydrate receptor. *Journal of Biological Chemistry*, *281*, 37748– 37757.

Fierobe, H. P., Mechaly, a, Tardif, C., Belaich, a, Lamed, R., Shoham, Y., … Bayer, E. a. (2001). Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional complexes. *The Journal of Biological Chemistry*, *276*(24), 21257–61.

Fierobe, H.-P., Bayer, E. A, Tardif, C., Czjzek, M., Mechaly, A., Bélaïch, A., … Bélaïch, J.-P. (2002). Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. *The Journal of Biological Chemistry*, *277*(51), 49621–30.

Fontes, C. M. G. A., Ponte, P. I. P., Reis, T. C., Soares, M. C., Gama, L. T., Dias, F. M. V, & Ferreira, L. M. a. (2004). A family 6 carbohydrate-binding module potentiates the efficiency of a recombinant xylanase used to supplement cereal-based diets for poultry. *British Poultry Science*, *45*(5), 648–56.

Fontes, C. M. G. A., & Gilbert, H. J. (2010). Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annual Review of Biochemistry*, *79*, 655–681.

Fontes, C. M. G. A., Gilbert, H. J., Hazelwood, G. P., Clarke, J. H., Prates, J. A. M., McKie, V. A., … Ferreira, L. M. . (2000). A novel *Cellvibrio mixtus* family 10 xylanase that is both intracellular and expressed under non-inducing conditions. *Microbiology*, *146*, 1959–1967.

Fontes, C. M. G. A., Hall, J., Hirst, B. H., Hazlewood, G. P., & Gilbert, H. J. (1995a). The resistance of cellulases and xylanases to proteolytic inactivation. *Applied Microbiology and Biotechnology*, *43*, 52–57.

Fontes, C. M., Hazlewood, G. P., Morag, E., Hall, J., Hirst, B. H., & Gilbert, H. J. (1995b). Evidence for a general role for non-catalytic thermostabilizing domains in xylanases from thermophilic bacteria. *The Biochemical Journal*, *307 (Pt 1*, 151–8.

Freier, D., Mothershed, C. P., & Wiegel, J. (1988). Characterization of *Clostridium thermocellum* JW20. *Applied and Environmental Microbiology*, *54*(1), 204–211.

Fridjonsson, O., Bahrani-mougeot, F. K., Hreggvidsson, O., Kristjansson, J. K., & Mattiasson, B. (2004). A new thermostable α-L-arabinofuranosidase from a novel thermophilic bacterium. *Biotechnology Letters*, *26*, 1347–1351.

Fry, S. (1989). Cellulases, Hemicellulases and Auxin-stimulated Growth - A possible relationship. *Physiologia Plantarum*, *75*(4), 532–536.

Fujino, T., Beguin, P., & Aubert, J. (1993). Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. *Journal of Bacteriology*, *175*(7), 1891–1899.

Gao, P. J., Chen, G. J., Wang, T. H., Zhang, Y. S., & Liu, J. (2001). Non-hydrolytic Disruption of Crystalline Structure of Cellulose by Cellulose Binding Domain and Linker Sequence of Cellobiohydrolase I from *Penicillium janthinellum*. *Acta Biochimica et Biophisica*, *33*(1), 13–18.

García-Alvarez, B., Melero, R., Dias, F. M. V, Prates, J. A. M., Fontes, C. M. G. A., Smith, S. P., … Llorca, O. (2011). Molecular architecture and structural transitions of a *Clostridium thermocellum* mini-cellulosome. *Journal of Molecular Biology*, *407*(4), 571–580.

Gasparic, J., & Churacek, J. (1978). *Laboratory handbook of paper and thinlayer chromatography* (pp. 153–154). Chichester, England: Ellis Horwood Ltd.

Geertsma, E. R., & Dutzler, R. (2011). A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry*, *50*(15), 3272–8.

Genomes OnLine Database. Retrieved from http://www.genomesonline.org/

Gilbert, H. J. (2007). Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. *Molecular Microbiology*, *63*(6), 1568–76.

Gilbert, H. J. (2010). The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiology*, *153*(2), 444–55.

Gilbert, H. J., Knox, J. P., & Boraston, A. B. (2013). Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Current Opinion in Structural Biology*, *23*(5), 669–77.

Gilkes, N. R., Warren, R. A., Miller, R. C., & Kilburn, D. G. (1988). Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. *The Journal of Biological Chemistry*, *263*(21), 10401–7.

Gold, N. D., & Martin, V. J. J. (2007). Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *Journal of Bacteriology*, *189*(19), 6787–95.

Grabski, A., McCormick, M., & Mierendorf, O. (1999). BugBuster and benzonase: the clear solution to simple, efficient extraction of *E. coli* proteins. *InNovations*, *10*, 17–19.

Gragerov, A., Nudler, E., Komissarova, N., Gaitanaris, G., Gottesman, M. E., Nikiforov, V. (1992). Cooperation of GroEL/GroES and DnaK/DnaJ heat shock proteins in preventing protein misfolding in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(November), 10341–10344.

Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B. M., Bray, J., Gileadi, O., … Gunsalus, K. C. (2008). Protein production and purification. *Nature Methods*, *5*(2), 135–146.

Gribskov, M., & Burgess, R. R. (1983). Overexpression and purification of the sigma subunit of *Escherichia coli* RNA polymerase. *Gene*, *26*(2-3), 109–118.

Guerreiro, C. I. P. D., Ribeiro, T., Ponte, P. I. P., Lordelo, M. M. S., Falcão, L., Freire, J. P. B., … Fontes, C. M. G. A. (2008). Role of a family 11 carbohydrate-binding module in the function of a recombinant cellulase used to supplement a barley-based diet for broiler chickens. *British Poultry Science*, *49*(4), 446–54.

Guillén, D., Sánchez, S., & Rodríguez-Sanoja, R. (2010). Carbohydrate-binding domains: multiplicity of biological roles. *Applied Microbiology and Biotechnology*, *85*(5), 1241–1249. doi:10.1007/s00253-009-2331-y

Ha, M. A., Apperley, D. C., & Jarvis, M. C. (1997). Molecular rigidity in dry and hydrated onion cell walls. *Plant Physiology*, *115*(2), 593–598.

Haag, A. F., & Ostermeier, C. (2009). Positive-selection vector for direct protein expression. *BioTechniques*, *46*(6), 453–7.

Halstead, J. R., Vercoe, P. E., Gilbert, H. J., Davidson, K., & Hazlewood, G. P. (1999). A family 26 mannanase produced by *Clostridium thermocellum* as a component of the cellulosome contains a domain which is conserved in mannanases from anaerobic fungi. *Microbiology*, *145*, 3101–3108.

Hammarström, Martin, Hellgren, Niklas, van den Berg, Susanne, Berglund, Helena and Härd, T. (2002). Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Science*, *11*, 313–321.

Han, S. O., Cho, H.-Y., Yukawa, H., Inui, M., & Doi, R. H. (2004). Regulation of expression of cellulosomes and noncellulosomal (hemi)cellulolytic enzymes in *Clostridium cellulovorans* during growth on different carbon sources. *Journal of Bacteriology*, *186*(13), 4218–4227.

Han, S. O., Yukawa, H., Inui, M., & Doi, R. H. (2003). Transcription of *Clostridium cellulovorans* cellulosomal cellulase and hemicellulase genes. *Journal of Bacteriology*, *185*(8), 2520–2527.

Harholt, J., Suttangkakul, A., & Vibe Scheller, H. (2010). Biosynthesis of pectin. *Plant Physiology*, *153*(2), 384–95.

Hartley D. L., & Kane, J. F. (1988). Properties of inclusion bodies from recombinant *Escherichia coli*. *Biochemical Society Transactions*, *16*(2), 101–102.

Hartley, J. L. (2006). Cloning technologies for protein expression and purification. *Current Opinion in Biotechnology*, *17*(4), 359–66.

Hartley, J. L., Temple, G. F., & Brasch, M. A. (2000). DNA cloning using *in vitro* site-specific recombination. *Genome Research*, *10*(11), 1788–1795.

Hasegawa, S., & Nagel, C. W. (1966). A new pectic acid transeliminase produced exocellularly by a Bacillus. *Journal of Food Science*, *31*, 838–845.

Hayashi, T. (1989). Xyloglucans in the primary cell wall. *Annual Review of Plant Physiology and Plant Molecular Biology*, *40*, 139–168.

Henrissat, B. (1991). A classification of glycosyl hydrolases based sequence similarities amino acid. *Biochemical Journal*, *280*, 309–316.

Henrissat, B., Heffron, S. E., Yoder, M. D., Lietzke, S. E., & Jurnak, F. (1996). Functional implication of structure-based sequence alignment of proteins in the extracellular pectate lyase superfamily. *Plant Physiology*, *107*, 963–976.

Henrissat, B., Teeri, T. T., & Warren, R. A. (1998). A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. *FEBS Letters*, *425*(2), 352–4.

Henshaw, J. L., Bolam, D. N., Pires, V. M. R., Czjzek, M., Henrissat, B., Ferreira, L. M. A., … Gilbert, H. J. (2004). The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. *The Journal of Biological Chemistry*, *279*(20), 21552–9.

Herron, S. R., Scavetta, R. D., Garrett, M., Legner, M., & Jurnak, F. (2003). Characterization and implications of Ca^{2+} binding to pectate lyase. *Journal of Biological Chemistry*, *278*, 12271–12277.

Hervé, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., & Knox, J. P. (2010). Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(34), 15293–8.

Heyman, J. A., Cornthwaite, J., Foncerrada, L., Gilmore, J. R., Gontang, E., Hartman, K. J., … Hoeffler, J. P. (1999). Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation. *Genome Research*, *9*, 383–392.

Hinton, J. C. D., Sidebotham, J. M., Gill, D. R., & Salmond, G. P. C. (1989). Extracellular and periplasmatic isoenzymes of pectate lyase from *Erwinia caratovora* subspecies carotovora belong to different gene families. *Molecular Microbiology*, *3*, 1785–1795.

Horton, N., Lewis, M., & Lu, P. (1997). *Escherichia coli lac* repressor-*lac* operator interaction and the influence of allosteric effectors. *Journal of Molecular Biology*, *265*(1), 1–7.

Hu, L.-L., Zhang, S.-S., Li, X.-X., & Wang, B.-L. (2010). The use of the *ccdB* lethal gene for constructing a zero background vector in order to clone blunt-end PCR Products. *Molecular Biology*, *44*(1), 161–164.

Huang, C.-J., Lin, H., & Yang, X. (2012). Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *Journal of Industrial Microbiology & Biotechnology*, *39*, 383–399.

Hugouvieux-Cotte-Pattat, N., Condemine, G., Nasser, W., & Reverchon, S. (1996). Regulation of pectinolysis in Erwinia chrysanthemi. *Annual Review of Microbiology*, *50*, 213–257.

Hunke, S., & Betton, J.-M. (2003). Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli*. *Molecular Microbiology*, *50*, 1579–1589.

Hwang, P. M., Pan, J. S., & Sykes, B. D. (2014). Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in *Escherichia coli*. *FEBS Letters*, *588*(2), 247–52.

Integrated Microbial Genomes with Microbiome Samples. Retrieved from https://img.jgi.doe.gov/cgi-bin/m/main.cgi

Jamar, C., Jardin, P. Du, & Fauconnier, M. L. (2011). Cell wall polysaccharides hydrolysis of malting barley (Hordeum vulgare L.): a review. *Biotechnology, Agronomy, Society and Environment*, *15*(2), 301–313.

Jamroz, D., Orda, J., Wiliczkiewicz, A., & Skorupińska, J. (1996). The apparent digestibility of structural carbohydrates and the intestine fermentation of different kinds of grains in three poultry species. Part 4. *Wiener Tieraerztliche Monatsschrift*, *83*(7), 210.

Jana, S., & Deb, J. K. (2005). Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Applied Microbiology and Biotechnology*, *67*(3), 289–98.

Jenkins, J., Shevchik, V. E., Hugouvieux-Cotte-Pattat, N., & Pickersgill, R. W. (2004). The Crystal Structure of Pectate Lyase Pel9A from *Erwinia chrysanthemi*. *Journal of Biological Chemistry*, *279*, 9139–9145.

Jenny, R. J., Mann, K. G., & Lundblad, R. L. (2003). A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expression and Purification*, *31*(1), 1–11.

Jindou, S., Borovok, I., Rincon, M. T., Flint, H. J., Antonopoulos, D. A., Berg, M. E., … Lamed, R. (2006). Conservation and divergence in cellulosome architecture between two strains of *Ruminococcus flavefaciens*. *Journal of Bacteriology*, *188*(22), 7971–6.

Johnson, I. T., & Gee, J. M. (1981). Effect of gel-forming gums on the intestinal unstirred layer and sugar transport *in vitro*. *Gut*, *22*(5), 398–403.

Jozefiak, D., Rutkowski, A., Jensen, B. B., & Engberg, R. M. (2006). The effect of beta-glucanase supplementation of barley- and oat-based diets on growth performance and fermentation in broiler chicken gastrointestinal tract. *British Poultry Science*, *47*, 57–64.

Jurnak, F., Kita, N., Garrett, M., Heffron, S. E., Scavetta, R., Boyd, C., & Keen, N. (1996). Functional implications of the three-dimensional structures of pectate lyases. In J. Visser & A. . G. J. Voragen (Eds.), *Pectin and Pectinases* (pp. 295–308). Amsterdam: Elsevier.

Kamionka, M. (2011). Engineering of Therapeutic Proteins Production in *Escherichia coli*. *Current Pharmaceutical Biotechnology*, *12*, 268–274.

Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*, *8*, 494–500.

Kang, S., Barak, Y., Lamed, R., Bayer, E. A., & Morrison, M. (2006). The functional repertoire of prokaryote cellulosomes includes the serpin superfamily of serine proteinase inhibitors. *Molecular Microbiology*, *60*(6), 1344–54.

Kapust, R. B., & Waugh, D. S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science : A Publication of the Protein Society*, *8*(8), 1668–74.

Kapust, R. B., & Waugh, D. S. (2000). Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expression and Purification*, *19*, 312–318.

Keegstra, K. (2010). Plant cell walls. *Plant Physiology*, *154*(2), 483–6.

Khattak, F. M., Pasha, T. N., Hayat, Z., &, & Mahmud, A. (2006). Enzymes in poultry nutrition. *Journal of Animal and Plant Sciences*, *16*, 1–7.

King, B. C., Waxman, K. D., Nenni, N. V, Walker, L. P., Bergstrom, G. C., & Gibson, D. M. (2011). Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. *Biotechnology for Biofuels*, *4*(1), 4.

Kirjavainen, P. V., & Gibson, G. R. (1995). Healthy gut microflora and allergy: factors influencing development of the microbiota. *Annals of Medicine*, *31*(4), 288–292.

Kleine, J., & Liebl, W. (2006). Comparative characterization of deletion derivatives of the modular xylanase XynA of *Thermotoga maritima*. *Extremophiles: Life Under Extreme Conditions*, *10*(5), 373–381.

Koehn, J., & Hunt, I. (2009). High-Throughput Protein Production (HTPP): A Review of Enabling Technologies to Expedite Protein Production. In S. A. Doyle (Ed.), *Methods in Molecular Biology: High Throughput Protein Expression and Purification, vol. 498* (Vol. 498, pp. 1–18). Totowa, NJ: Humana Press.

Koshland, D. E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews of the Cambridge Philosophical Society*, *28*, 416–436.

Kruus, K., Lua, A. C., Demain, A. L., & Wu, J. H. (1995). The anchorage function of CipA (CelL), a scaffolding protein of the *Clostridium thermocellum* cellulosome. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(20), 9254–8.

Kurland, C., & Gallant, J. (1996). Errors of heterologous protein expression. *Current Opinion in Biotechnology*, *7*(5), 489–93.

Kyte, J., Doolittle, R. F., Diego, S., & Jolla, L. (1982). A Simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*, 105–132.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, *227*, 680–685.

Lamed, R., Setter, E., & Bayer, E. A. (1983). Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *Journal of Bacteriology*, *156*(2), 828–836.

Lamed, R., Setter, E., Kenig, R., & Bayer, E. A. (1983). The cellulosome: a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose- binding and various cellulolytic activities. *Biotechnology and Bioengineering Symposium*, *13*, 163–181.

LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant K. L., Schendel, P. F., McCoy, J. M. (1993). A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology (N. Y.)*, *11*(2), 187–193.

Leibovitz, E., & Béguin, P. (1996). A new type of cohesin domain that specifically binds the dockerin domain of the *Clostridium thermocellum* cellulosome-integrating protein CipA. *Journal of Bacteriology*, *178*(11), 3077–3084.

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. *Journal of Bacteriology*, *177*(9), 2451–2459.

Lesley, S. A. (2001). High-throughput proteomics: protein expression and purification in the postgenomic world. *Protein Expression and Purification*, *22*(2), 159–64.

Lesley, S. A, Kuhn, P., Godzik, A., Deacon, A. M., Mathews, I., Kreusch, A., … Stevens, R. C. (2002). Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(18), 11664–9.

Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., & Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*.

Li, L. L., McCorkle, S. R., Monchy, S., Taghavi, S., & van der Lelie, D. (2009). Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. *Biotechnology for Biofuels*, *2*, 10.

Li, M. Z., & Elledge, S. J. (2005). MAGIC, an *in vivo* genetic method for the rapid construction of recombinant DNA molecules. *Nature Genetics*, *37*, 311–319.

Li, M. Z., & Elledge, S. J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature Methods*, *4*, 251–256.

Lietzke, S. E., Yoder, M. D., Keen, N. T., & Jurnak, F. (1994). The three dimensional structure of pectate lyase E, a plant virulence factor from *Erwinia chysanthemi*. *Plant Physiology*, *106*, 849–862.

Lilie, H., Schwarz, E., & Rudolph, R. (1998). Advances in refolding of proteins produced in *E. coli*. *Current Opinion in Biotechnology*, *9*(5), 497–501.

Linhardt, R. J., Galliher, P. M., & Cooney, C. L. (1986). Polysaccharide lyases. *Applied Biochemistry and Biotechnology*, *12*, 135–176.

Liu, Z. L., Saha, B. C., & Slininger, P. J. (2008). Lignocellulosic biomass conversion to ethanol by *Saccharomyces* spp. In *Wall JD, Harwood CS* (Demain A (., pp. 17–36). ASM Press, Washington, DC.

Lojkwoska, E., Masclaux, C., Boccara, M., Robert-Baudouy, J., & Hugouvieux-Cotte-Pattat, N. (1995). Characterization of *pelL* gene encoding a novel pectate lyase of *Erwinia chysanthemi* 3937. *Molecular Microbiology*, *16*(6), 1183–1195.

Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P. M., & Henrissat, B. (2010). A hierarchical classification of polysaccharide lyases for glycogenomics. *The Biochemical Journal*, *432*(3), 437–444.

Lorimer, H. G. (1996). A quantitative assessment of the role of the the chaperonin proteins in protein folding in vivo. *The FASEB Journal*, *10*, 5–9.

Lynd, L. R., Currie, D., Ciazza, N., Herring, C., & Orem, N. (2008a). Consolidated bioprocessing of cellulosic biomass to ethanol using thermophilic bacteria. In *Wall JD, Harwood CS* (Demain A (.). Washington, DC: ASM Press.

Lynd, L. R., Cushman, J. H., Nichols, R. J., & Wyman, C. E. (1991). Fuel ethanol from cellulosic biomass. *Science (New York, N.Y.)*, *251*(4999), 1318–1323.

Lynd, L. R., Laser, M. ., Bransby, D., Dale, B. E., Davison, B., Hamilton, R., … Wyman, C. E. (2008b). How biotech can transform biofuels. *Nature Biotechnology*, *26*(2), 169–72.

Lynd, L. R., van Zyl, W. H., McBride, J. E., & Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. *Current Opinion in Biotechnology*, *16*(5), 577–83.

Lynd, L. R., Weimer, P. J., Zyl, W. H. Van, & Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, *66*(3), 506–577.

Lytle, B. L., Volkman, B. F., Westler, W. M., & Wu, J. H. (2000). Secondary structure and calcium-induced folding of the *Clostridium thermocellum* dockerin domain determined by NMR spectroscopy. *Archives of Biochemistry and Biophysics*, *379*(2), 237–244.

Lytle, B., Myers, C., Kruus, K., Wu, J. H., Lytle, B., Myers, C., … Wu, J. H. D. (1996). Interactions of the CelS binding ligand with various receptor domains of the *Clostridium thermocellum* cellulosomal scaffolding protein , CipA. *Journal of Bacteriology*, *178*(4), 1200–1203.

M. A. O'Neill, P. Albersheim, & A. D. (1990). The pectic polysaccharides of primary cell walls. In *P.M. Dey (Ed.) Methods in Plant Biochemistry* (vol. 2., pp. 415–441). London: Academic Press.

M. A. S. A. Samir, F. Alloin, & A. Dufresne. (2005). Review of recent research into cellulosic whiskers, their properties and their application in nanocomposite field. *Biomacromolecules*, *6*(2), 612–26.

M. McNeil, A. G. Darvill, S. C. Fry, P. A. (1984). Structure and function of the primary cell walls of plants. *Annual Review of Biochemistry*, *53*, 625–663.

Maglione, G., Matsushita, O., Russell, J. B., & Wilson, D. B. (1992). Properties of a genetically reconstructed *Prevotella ruminicola* endoglucanase. *Applied and Environmental Microbiology*, *58*(11), 3593–3597.

Malça, J., & Freire, F. (2006). Renewability and life-cycle energy efficiency of bioethanol and bio-ethyl tertiary butyl ether (bioETBE): Assessing the implications of allocation. *Energy*, *31*(15), 3326 – 3380.

Mancia, F., & Love, J. (2011). High throughput platforms for structural genomics of integral membrane proteins. *Current Opinion in Structural Biology*, *21*(4), 517–22.

Mangala, S. L., Kittur, F. S., Nishimoto, M., Sakka, K., Ohmiya, K., Kitaoka, M., & Hayashi, K. (2003). Fusion of family VI cellulose binding domains to *Bacillus halodurans* xylanase increases its catalytic activity and substrate-binding capacity to insoluble xylan. *Journal of Molecular Catalysis B: Enzymatic*, *21*, 221–230.

Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., & Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Science : A Publication of the Protein Society*, *15*(1), 182–9.

Marquardt, R. R., Brenes, A., & Zhang, Z. (1996). Use of enzymes to improve nutrient availability in poultry feedstuffs. *Animal Feed Science Technology*, *60*, 331–330.

Marsischky, G., & LaBaer, J. (2004). Many paths to many clones: a comparative look at high-throughput cloning methods. *Genome Research*, *14*(10B), 2020–8.

Martemyanov, K. a, Spirin, a S., & Gudkov, a T. (1997). Direct expression of PCR products in a cell-free transcription/translation system: synthesis of antibacterial peptide cecropin. *FEBS Letters*, *414*(2), 268–70.

Mathlouthi, N., Lalles, J. P., Lepercq, P., Juste, C., & Larbier, M. (2002). Xylanase and {beta}-glucanase supplementation improve conjugated bile acid fraction in intestinal contents and increase villus size of small intestine wall in broiler chickens fed a rye-based diet. *Journal of Animal Science*, *80*, 2773–2779.

Mavromatis, K., Land, M. L., Brettin, T. S., Quest, D. J., Copeland, A., Clum, A., … Kyrpides, N. C. (2012). The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. *PloS One*, *7*(12), e48837.

McCarter, J. D., & Withers, S. G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Current Opinion in Structural Biology*, *4*(6), 885–892.

McDonough, M. A., Kadirvelraj, R., Harris, P., Poulsen, J. N., & Larsen, S. (2004). Rhamnogalacturonan lyase reveals a unique thee-domain modular structure for polysaccharide lyase family 4. *FEBS Letters*, *565*, 188–194.

McKie, V. A., Vincken, J.-P., Voragen, A. G. J., Van Den Broek, L., Stimson, E., & Gilbert, H. J. (2001). A new family of rhamnogalacturonan lyases contains an enzyme that binds to cellulose. *Biochemistry Journal*, *355*, 167–177.

McNeil, M., Darvill, A. G., & Albersheim, P. (1980). Structure of plant cell walls X. Rhamnogalacturonan I, a structurally complex pectic polysaccharide in the walls of suspension-cultured sycamore cells. *Plant Physiology*, *66*, 1128–1134.

Miller, G. L. (1959). The use of dinitrosalicylic acid reagent for the determination of reducing sugar. *Analytical Chemistry*, *31*, 426–428.

Miller, M. E. B., Antonopoulos, D. A., Rincon, M. T., Band, M., Bari, A., Akraiko, T., … White, B. A. (2009). Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PloS One*, *4*(8), e6650.

Missiakas, D. and Raina, S. (1997). Protein Folding in the Bacterial Periplasm. *Journal of Bacteriology*, *179*(8), 2465–2471.

Mizutani, K., Fernandes, V. O., Karita, S., Luís, A. S., Sakka, M., Kimura, T., … Sakka, K. (2012). Influence of a mannan binding family 32 carbohydrate binding module on the activity of the appended mannanase. *Applied and Environmental Microbiology*, *78*(14), 4781–7.

Moffatt, B. A., &, & Studier, F. W. (1987). T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell*, *49*(2), 221–227.

Montanier, C., van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., … Gilbert, H. J. (2009). Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 3065–3070.

Morag, E., Bayer, E. A., & Lamed, R. (1990). Relationship of cellulosomal and noncellulosomal xylanases of *Clostridium thermocellum* to cellulose-degrading enzymes. *Journal of Bacteriology*, *172*(10), 6098–105.

Moran, F. S., Nasuno, S., & Starr, M. P. (1968). Extracellular and intracellular polygalacturonic acid trans eliminase of *Erwinia carotovora*. *Archives of Biochemistry and Biophysics*, *123*, 298–306.

Morgan, R. J., & Bedford, M. R. (1995). Advances in the Development and Application of Feed Enzymes. In *Proceedings of Australian Poultry Science Symposium, 7* (pp. 109–115). Sydney: University of Sydney; Poultry Research Foundation ; World's Poultry Science Association; Australian Branch.

Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y. Y., Holtzapple, M., & Ladisch, M. (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology*, *96*(6), 673–86.

Munir, K., & Maqsood, S. (2012). A review on role of exogenous enzyme supplementation in poultry production. *Emirates Journal of Food and Agriculture*, *25*(1), 66–80.

Munir, R. I., Schellenberg, J., Henrissat, B., Verbeke, T. J., Sparling, R., & Levin, D. B. (2014). Comparative analysis of carbohydrate active enzymes in *Clostridium termitidis* CT1112 reveals complex carbohydrate degradation ability. *PloS One*, *9*(8), e104260.

Murashima, K., Kosugi, a., & Doi, R. H. (2002). Synergistic effects on crystalline cellulose degradation between cellulosomal cellulases from *Clostridium cellulovorans*. *Journal of Bacteriology*, *184*(18), 5088–5095.

N.R.C. (1994). *Nutrient requirements of poultry* (9th ed.). Washington, DC: Natl. Acad. Press.

Nahas, J., & Lefrancois, M. R. (2001). Effects of feeding locally grown whole barley with or without enzyme addition and whole wheat on broiler performance and carcass traits. *Poultry Science*, *80*, 195–202.

Newcomb, M., Chen, C.-Y., & Wu, J. H. D. (2007). Induction of the *celC* operon of Clostridium thermocellum by laminaribiose. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(10), 3747–52.

Newman, R. K., & Newman, C. W. (1988). Nutritive-value of a new hull-less barley cultivar in broiler chick diets. *Poultry Science*, *67*, 1573–1579.

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

Nikaido, H. (1994). Maltose transport system of *Escherichia coli*: an ABC-type transporter. *FEBS Letters*, *346*, 55–58.

Nishihara, K., Kanemori, M., Yanagi, H., & Yura, T. (2000). Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Applied and Environmental Microbiology*, *66*(3), 884–9.

Nishihara, K, et al. (1998). Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, in *Escherichia coli*. *Applied and Environmental Microbiology*, *64*, 1694–1699.

Nishiyama, Y. (2009). Structure and properties of the cellulose microfibril. *Journal of Wood Science*, *55*(4), 241–249.

Nitsan, Z., Benavraham, G., Zoref, Z., & Nir, I. (1991). Growth and development of the digestive organs and some enzymes in broiler chicks after hatching. *British Poultry Science*, *32*, 515–523.

Noach, I., Frolow, F., Alber, O., Lamed, R., Shimon, L. J. W., & Bayer, E. A. (2009). Intermodular linker flexibility revealed from crystal structures of adjacent cellulosomal cohesins of *Acetivibrio cellulolyticus*. *Journal of Molecular Biology*, *391*(1), 86–97.

Nordon, R. E., Craig, S. J., & Foong, F. C. (2009). Molecular engineering of the cellulosome complex for affinity and bioenergy applications. *Biotechnology Letters*, *31*(4), 465–476.

Nozach, H., Fruchart-Gaillard, C., Fenaille, F., Beau, F., Ramos, O. H. P., Douzi, B., … Dive, V. (2013). High throughput screening identifies disulfide isomerase DsbC as a very efficient partner for recombinant expression of small disulfide-rich proteins in *E. coli*. *Microbial Cell Factories*, *12*(37), 2–16.

Numan, M. T., & Bhosle, N. B. (2006). Alpha-L-arabinofuranosidases: the potential applications in biotechnology. *Journal of Industrial Microbiology & Biotechnology*, *33*(4), 247–60.

O Neill, M. A., Warrenfeltz, D., Kates, K., Pellerin, P., Doco, T., Darvill, A. G., & Albersheim, P. (1996). Rhamnogalacturonan-II, a pectic polysaccharide in the walls of growing plant cell, forms a dimer that is covalently cross-linked by a borate ester. *Journal of Biological Chemistry*, *271*, 22923–22930.

Ochiai, A., Itoh, T., Maruyama, Y., Kawamata, A., Mikami, B., Hashimoto, W., & Murata, K. (2007). A novel structural fold in polysaccharide lyases: *Bacillus subtilis* family 11 rhamnogalacturonan lyase yesw with an eight-bladed β-propeller. *Journal of Biological Chemistry*, *282*, 37134–37145.

Pace, N. R., Sthal, D. A., Lane, D. J., & Olsen, G. J. (1986). The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences. *Advances in Microbial Ecology*, *9*(1-55).

Pagès, S., Valette, O., Abdou, L., Bélaïch, A., & Bélaïch, J. P. (2003). A rhamnogalacturonan lyase in the *Clostridium cellulolyticum* cellulosome. *Journal of Bacteriology*, *185*(16), 4727–4733.

Pagès, S., Bélaïch, A., Bélaïch, J. P., Morag, E., Lamed, R., Shoham, Y., & Bayer, E. A. (1997). Species-specificity of the cohesin-dockerin interaction between *Clostridium thermocellum* and *Clostridium cellulolyticum*: prediction of specificity determinants of the dockerin domain. *Proteins*, *29*(4), 517–527.

Pan, S. H., and Malcolm, B. A. (2000). Reduced Background Expression and Improved Plasmid Stability with pET Vectors in BL21 (DE3). *BioTechniques*, *29*(6), 1–4.

Pauly, M., & Keegstra, K. (2008). Cell-wall carbohydrates and their modification as a resource for biofuels. *The Plant Journal : For Cell and Molecular Biology*, *54*(4), 559–68.

Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P., Landau, E. M. (1997). X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science*, *277*(5332), 1676–1681.

Peer, A., Smith, S. P., Bayer, E. A., Lamed, R., & Borovok, I. (2011). Noncellulosomal cohesin- and dockerin-like modules in the three domains of life. *FEMS Microbiology Letters*, *291*(1), 1–16.

Peleg, Y., & Unger, T. (2012). Chemical Genomics and Proteomics. In E. D. Zanders (Ed.), *Chemical Genomics and Proteomics: Reviews and Protocols, Methods in Molecular Biology, vol. 800* (Vol. 800, pp. 173–186). Totowa, NJ: Humana Press. doi:10.1007/978-1-61779-349-3

Percival Zhang, Y.-H., Himmel, M. E., & Mielenz, J. R. (2006). Outlook for cellulase improvement: screening and selection strategies. *Biotechnology Advances*, *24*(5), 452–81.

Pettersson, D., & Aman, P. (1989). Enzyme supplementation of a poultry diet containing rye and wheat. *The British Journal of Nutrition*, *62*(1), 139–49.

Pickersgill, R., Jenkins, J., Harris, G., Nasser, W., & Robert-Baudouy, J. (1994). The structure of *Bacillus subtilis* pectate lyase in complex with calcium. *Nature Structural Biology*, *1*, 717–723.

Pissavin, C., Robert-Baudouy, J., & Hugouvieux-Cotte-Pattat, N. (1996). Regulation of *pelZ*, a gene of the pelBC cluster encoding a new pectate lyase in *Erwinia chrysanthem*i. *Journal of Bacteriology*, *178*, 7187–7196.

Poletto, M., Pistor, V., & Zattera, A. J. (2013). Structural characteristics and thermal properties of native cellulose. In *Cellulose – Fundamental Aspects* (pp. 45–68).

Ponte, P. I. P., Ferreira, L. M. A., Soares, M. A. C., Gama, L. T., & Fontes, C. M. G. A. (2004). Xylanase inhibitors affect the action of exogenous enzymes used to supplement Triticum durum-based diets for broiler chicks. *Journal of Applied Poultry Research*, *13*, 660–666.

Ponte, P. I. P., Lordelo, M. M. S., Guerreiro, C. I. P., Soares, M. A. C., Mourão, J. L., Crespo, J. P., … Fontes, C. M. G. A. (2008). Crop β-glucanase activity limits the effectiveness of a recombinant cellulase used to supplement a barley-based feed for free-range broilers. *British Poultry Science*, *49*, 347–359.

Ponte, P. I. P., Mendes, I., Quaresma, M., Aguiar, M. N., Lemos, J. P., Ferreira, L. M., … Fontes, C. A. M. (2004). Cholesterol levels and sensory characteristics of meat from broilers consuming moderate to high levels of alfalfa. *Poultry Science*, *83*(5), 810–814.

Poole D. M., Morag E., Lamed R., Bayer E. A., Hazlewood G. P., Gilbert, H. J. (1992). Identification of the cellulose-binding domain of the cellulosome subunit S1 from *Clostridium thermocellum* YS. *FEMS Microbiology Letters*, *78*, 181–186.

Popper, Z. A. (2008). Evolution and diversity of green plant cell walls. *Current Opinion in Plant Biology*, *11*(3), 286–92.

Quan, J., & Tian, J. (2009). Circular polymerase extension cloning of complex gene libraries and pathways. *PloS One*, *4*(7), e6441.

Quiocho, F. A. (1986). Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Annual Review of Biochemistry*, *55*, 287–315.

Rais-Beghdadi, C., Roggero, M. a, Fasel, N., & Reymond, C. D. (1998). Purification of recombinant proteins by chemical removal of the affinity tag. *Applied Biochemistry and Biotechnology*, *74*(2), 95–103.

Raman, B., Pan, C., Hurst, G. B., Rodriguez, M., McKeown, C. K., Lankford, P. K., … Mielenz, J. R. (2009). Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. *PloS One*, *4*(4), e5271.

Raran-Kurussi, S., & Waugh, D. S. (2012). The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. *PLoS ONE*, *7*(11).

Ribeiro, T., Lordelo, M. M. S., Ponte, P. I. P., Macas, B., Prates, J. A. M., Aguiar Fontes, M., … Fontes, C. M. G. A. (2011). Levels of endogenous β-glucanase activity in barley affect the efficacy of exogenous enzymes used to supplement barley-based diets for poultry. *Poultry Science*, *90*(6), 1245–1256.

Ribeiro, T., Lordelo, M. M. S., Prates, J. a. M., Falcão, L., Freire, J. P. B., Ferreira, L. M. A., & Fontes, C. M. G. A. (2012). The thermostable β-1,3-1,4-glucanase from *Clostridium thermocellum* improves the nutritive value of highly viscous barley-based diets for broilers. *British Poultry Science*, *53* (February 2015), 224–234.

Ribeiro, T., Ponte, P. I. P., Guerreiro, C. I. P. D., Santos, H. M., Falcão, L., Freire, J. P. B., … Lordelo, M. M. (2011). A family 11 carbohydrate-binding module (CBM) improves the efficacy of a recombinant cellulase used to supplement barley-based diets for broilers at lower dosage rates. *British Poultry Science*, *49*(5), 600–608.

Ribeiro, T., Santos-Silva, T., Alves, V. D., Dias, F. M. V, Luís, A. S., Prates, J. A. M., … Fontes, C. M. G. A. (2010). Family 42 carbohydrate-binding modules display multiple arabinoxylan-binding interfaces presenting different ligand affinities. *Biochimica Et Biophysica Acta*, *1804*(10), 2054–2062.

Ridley, B. L., O Neil, M. A., & Mohnen, D. (2001). Pectins: structure, biosyntesis and oligogalacturonic-related signaling. *Phytochemistry*, *57*, 929–967.

Rinas, U., & Bailey, J. E. (1993). Overexpression of bacterial hemoglobin causes incorporation of pre-beta-lactamase into cytoplasmic inclusion bodies. *Applied and Environmental Microbiology*, *59*(2), 561–6.

Rincon, M. T., Cepeljnik, T., Martin, J. C., Barak, Y., Lamed, R., Bayer, E. a, & Flint, H. J. (2007). A novel cell surface-anchored cellulose-binding protein encoded by the sca gene cluster of *Ruminococcus flavefaciens*. *Journal of Bacteriology*, *189*(13), 4774–83.

Rincon, M. T., Cepeljnik, T., Martin, J. C., Lamed, R., Barak, Y., Bayer, E. A., … Flint, H. J. (2005). Unconventional mode of attachment of the *Ruminococcus flavefaciens* cellulosome to the cell surface. *Journal of Bacteriology*, *187*, 7569–7578.

Rincon, M. T., Dassa, B., Flint, H. J., Travis, A. J., Jindou, S., Borovok, I., … White, B. a. (2010). Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, *Ruminococcus flavefaciens* FD-1. *PloS One*, *5*(8), e12476.

Rincon, M. T., Ding, S., Mccrae, S. I., Martin, C., Aurilia, V., Lamed, R., … Martin, J. C. (2003). Novel Organization and Divergent Dockerin Specificities in the Cellulosome System of *Ruminococcus flavefaciens*. *Journal of Bacteriology*, *185*(3), 703–713.

Rincón, M. T., Martin, J. C., Aurilia, V., Mccrae, S. I., Rucklidge, G. J., Reid, M. D., … Bayer, E. A. (2004). ScaC , an adaptor protein carrying a novel cohesin that expands the dockerin-binding repertoire of the *Ruminococcus flavefaciens* 17 cellulosome ScaC , an adaptor protein carrying a novel cohesin that expands the dockerin-binding repertoire of the *Ruminoco*. *Journal of Bacteriology*, *186*(9), 2576–2585.

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, *5*(April), 172.

Rotter, B. A., Neskar, M., Guenter, W., & Marquardt, R. R. (1988). Effect of enzyme supplementation on the nutritive value of hulless barley in chicken diets. *Animal Feed Science and Technology*, *24*, 233–245.

Ruijssennars, H. J., & Hartmans, S. (2001). Plate screening method for the detection of polysaccharide- producing microorganisms. *Applied Microbiology and Biotechnology*, *55*, 143–149.

Sa, M., & Sleytr, U. W. E. B. (2000). S-Layer Proteins. *Journal of Bacteriology*, *182*(4), 859–868.

Saha, B. C. (2000). α-L-Arabinofuranosidases : biochemistry, molecular biology and application in biotechnology. *Biotechnology Advances*, *18*, 403–423.

Saha, B. C. (2003). Hemicellulose bioconversion. *Journal of Industrial Microbiology & Biotechnology*, *30*(5), 279–91.

Sahdev, S., Khattar, S. K., & Saini, K. S. (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry*, *307*(1-2), 249–64.

Sakka, M., Goto, M., Fujino, T., Fujino, E., Karita, S., Kimura, T., & Sakka, K. (2010). Analysis of a *Clostridium josui* cellulase gene cluster containing the *man5A* gene and characterization of recombinant Man5A. *Bioscience, Biotechnology, and Biochemistry*, *74*, 2077.

Sakka, M., Higashi, Y., Kimura, T., Ratanakhanokchai, K., & Sakka, K. (2011). Characterization of *Paenibacillus curdlanolyticus* B-6 Xyn10, a xylanase that contains a family 3 carbohydrate-binding module. *Applied and Environmental Microbiology*, *77*, 4260–4263.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning; a Laboratory Manual* (2nd edition.). New York: Cold Spring Harbor Laboratory.

Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F., & Petersen, G. B. (1982). Nucleotide Sequence of Bacteriophage Lambda. *Journal of Molecular Biology*, *162*(4), 729–773.

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*(1), 89.

Sarkar, P., Bosneaga, E., & Auer, M. (2009). Plant cell walls throughout evolution : towards a molecular understanding of their design principles. *Biotechnology for Biofuels*, *60*(13), 3615–3635.

SAS. (2004). *SAS User´s Guide: Statistics*. SAS Institute In Cary NC.

Satakarni, M., & Curtis, R. (2011). Production of recombinant peptides as fusions with SUMO. *Protein Expression and Purification*, *78*(2), 113–9.

Scavetta, R. D., Herron, S. R., Hotchkiss, A. T., Kita, N., Keen, N. T., Benen, J. A. E., … Jurnak, F. (1999). Structure of plant cell wall fragment complexed to pectate lyase C. *The Plant Cell*, *11*, 1081–1092.

Scheich, C., Sievert, V., & Büssow, K. (2003). An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography. *BMC Biotechnology*, *3*, 12.

Schein, C. H. (1989). Production of Soluble Recombinant Proteins in Bacteria. *Nature Biotechnology*, *7*, 1141 – 1149.

Scheller, H. V., & Ulvskov, P. (2010). Hemicelluloses. *Annual Review of Plant Biology*, *61*, 263–89.

Schubert, C. (2006). Can biofuels finally take center stage? *Nature Biotechnology*, *24*(7), 777–84.

Schuster, M. (2000). Protein expression strategies for identification of novel target proteins. *Journal of Biomolecular Screening*, *5*(2), 89–97.

Schwarz, W. H., & Zverlov, V. V. (2006). Protease inhibitors in bacteria: an emerging concept for the regulation of bacterial protein complexes? *Molecular Microbiology*, *60*(6), 1323–6.

Shaw, a J., Podkaminer, K. K., Desai, S. G., Bardsley, J. S., Rogers, S. R., Thorne, P. G., … Lynd, L. R. (2008). Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(37), 13769–74.

Sheridan, C. (2009). Making green. *Nature Biotechnology*, *27*, 1074–1076.

Shevchik, V. E., Kester, H. C. M., Benen, J. A. E., Visser, J., Robert-Baudouy, J., & Hugouvieux-Cotte-Pattat, N. (1999). Characterization of the exopolygalacturonate lyase PelX of *Erwinia chysanthemi* 3937. *Journal of Bacteriology*, *181*, 1652–1663.

Shevchik, V. E., Robert-Baudouy, J., & Hugouvieux-Cotte-Pattat, N. (1997). The pectate lyase PelI of *Erwinia chrysanthemi* belongs to a new family. *Journal of Bacteriology*, *179*, 7321–7330.

Shih, Y., Kung, W., Chen, J., Yeh, C., Wang, A. H., & Wang, T. (2002). High-throughput screening of soluble recombinant proteins. *Protein Science*, *11*(7), 1714–1719.

Shuman, S. (1994). Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *The Journal of Biological Chemistry*, *269*(51), 32678–84.

Sinnott, M. L. (1990). Catalytic mechanism of enzymic glycosyl transfer. *Chemical Reviews*, *90*, 1171–1202.

Smith, D. B, & Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*, *67*(1), 31–40.

Smits, C. H. M., & Annison, G. (1996). Non-starch plant polysaccharides in broiler nutrition-towards s physiologically valid approach to their determination. *World Poultry Science*, *52*, 203–221.

Somerville, C. (2006). Cellulose synthesis in higher plants. *Annual Review of Cell and Developmental Biology*, *22*, 53–78.

Sørensen, H. P., & Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology*, *115*, 113–128.

Soriano, M., Blanco, A., Diaz, P., & Pastor, F. I. (2000). An unusual pectate lyase from a *Bacillus* sp. with high activity on pectin: cloning and characterization. *Microbiology*, *146*, 89–95.

Soriano, M., Diaz, P., & Pastor, F. I. (2006). Pectate lyase C from *Bacillus subtilis*: a novel endo-cleaving enzyme with activity on highly methylated pectin. *Microbiology*, *152*, 617–625.

Spiess, C., Beil, A., & Ehrmann, M. (1999). A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell*, *97*, 339–347.

Spinnler, H. E., Lavigne, B., & Blachere, H. (1986). Pectinolytic activity of *Clostridium thermocellum*: its use for anaerobic fermentation of sugar beet pulp. *Applied Microbiology and Biotechnology*, *23*, 434–437.

Staden, R. (1979). A strategy of DNA sequencing employing computer programs. *Nucleic Acids Research*, *6*(7), 2601–2610.

Ståhlberg, J., Johansson, G., & Pettersson, G. (1993). Trichoderma reesei has no true exo-cellulase: all intact and truncated cellulases produce new reducing end groups on cellulose. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1157*(1), 107–113.

Stewart, P. D. S., &, & Baldock, P. F. M. (1999). Practical experimental design techniques for automatic and manual protein crystallization. *Journal of Crystal Growth*, *196*(2-4), 665–673.

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–234.

Studier, M. F. (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of Molecular Biology*, *219*(1), 37–44.

Sukhumsiirchart, W., Kawanishi, S., Deesukon, W., Chansiri, K., Kawasaki, H., & Sakamoto, T. (2009). Purification, characterization, and overexpression of thermophilic pectate lyase of *Bacillus* sp. RN1 isolated from a hot spring in Thailand. *Bioscience, Biotechnology, and Biochemistry*, *73*(2), 268–73.

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

Swartz, J. R. (2001). Advances in Escherichia coli production of therapeutic proteins. *Current Opinion in Biotechnology*, *12*(2), 195–201.

Tachibana, A., Tohiguchi, K., Ueno, T., Setogawa, Y., Harada, A., & Tanabe, T. (2009). Preparation of long sticky ends for universal ligation-independent cloning: sequential T4 DNA polymerase treatments. *Journal of Bioscience and Bioengineering*, *107*(6), 668–9.

Tailford, L. E., Ducros, V. M., Flint, J. E., Roberts, S. M., Morland, C., Zechel, D. L., … Gilbert, H. J. (2009). Understanding how diverse beta-mannanases recognize heterogeneous substrates. *Biochemistry*, (48), 7009 –70018.

Talbott, L. D., & Ray, P. M. (1992). Molecular size and separability features of pea cell wall polysaccharides: implications for models of primary wall structure. *Plant Physiology*, *98*(1), 357–368.

Tamaru, Y., & Doi, R. H. (2001). Pectate lyase A, an enzymatic subunit of the Clostridium cellulovorans cellulosome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(7), 4125–9.

Tamaru, Y., Miyake, H., Kuroda, K., Ueda, M., & Doi, R. H. (2010). Comparative genomics of the mesophilic cellulosome-producing *Clostridium cellulovorans* and its application to biofuel production via consolidated bioprocessing. *Environmental Technology*, *31*, 889–903.

Taylor, L. E., Dai, Z., Decker, S. R., Brunecky, R., Adney, W. S., Ding, S.-Y., & Himmel, M. E. (2008). Heterologous expression of glycosyl hydrolases in planta: a new departure for biofuels. *Trends in Biotechnology*, *26*(8), 413–24.

Taylor, N. G. (2008). Cellulose biosynthesis and deposition in higher plants. *The New Phytologist*, *178*(2), 239–52.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, *60*(5), 523–33.

Thomas, J. G., & Baneyx, F. (1996a). Protein folding in the cytoplasm of *Escherichia coli*: requirements for the DnaK-DnaJ-GrpE and GroEL-GroES molecular chaperone machines. *Molecular Microbiology*, *21*(6), 1185–96.

Thomas, J. G., & Baneyx, F. (1996b). Protein Misfolding and Inclusion Body Formation in Recombinant *Escherichia col*i Cells Overexpressing Heat-shock Proteins. *Journal of Biological Chemistry*, *271*(19), 11141–11147.

Tomme, P., Warren, R. A. J., Gilkes, N. R., & Poole, R. K. (1995). Cellulose Hydrolysis by Bacteria and Fungi. *Advances in Microbial Physiology*, *37*, 1–81.

Tsai, S.-L., Goyal, G., & Chen, W. (2010). Surface display of a functional minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production. *Applied and Environmental Microbiology*, *76*(22), 7514–7520.

Tunnicliffe, R. B., Bolam, D. N., Pell, G., Gilbert, H. J., & Williamson, M. P. (2005). Structure of a mannan-specific family 35 carbohydrate-binding module: evidence for significant conformational changes upon ligand binding. *Journal of Molecular Biology*, *347*(2), 287–96.

Van der Wielen, P. W. J. ., Biesterveld, S., Notermans, S., Hofstra, H., Urling, B. A. P., & Van Knapen, F. (2000). Role of volatile fatty acids in development of the cecal microflora in broiler chickens during growth. *Applied and Environmental Microbiology*, *66*(6), 2536–2540.

Vasina, J. a, & Baneyx, F. (1997). Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the *Escherichia coli cspA* and *tac* promoter systems. *Protein Expression and Purification*, *9*, 211–218.

Wadal, K., Wada, Y., Ishibashi, F., Gojobori, T., & Ikemura, T. (1992). Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Research*, *20*, 2111–2118.

Wang, L., Zhang, Y., & Gao, P. (2008). A novel function for the cellulose binding module of cellobiohydrolase I. *Science in China Series C: Life Sciences*, *51*(7), 620–629.

Warren, R. A. J. (1996). Microbial hydrolysis of polysaccharides. *Annual Review of Microbiology*, *50*(1), 183–212.

Weber, J. L., & Myers, E. W. (1997). Human Whole-Genome Shotgun Sequencing. *Genome Research*, *7*, 401–409.

White, G. W., Katona, T., & Zodda, J. P. (1999). The Use of High-Performance Size Exclusion Chromatography (HPSEC) as a Molecular weight screening technique for polygalacturonic acid for use in pharmaceutical applications. *Journal of Pharmaceutical and Biomedical Analysis*, *20*, 905.

Whitman, W. B. , Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(12), 6578–6583.

Wild, J., Hradecna, Z., & Szybalski, W. (2002). Conditionally Amplifiable BACs : Switching From Single-Copy to High-Copy Vectors and Genomic Clones. *Genome*, *12*(1), 1434–1444.

Willats, W. G., McCartney, L., Mackie, W., & Knox, J. P. (2001). Pectin: cell biology and prospects for functional analysis. *Plant Molecular Biology*, *47*(1-2), 9–27.

Williams, P. E. V, Geraert, P. A., Uzu, G., & Annison, G. (1997). Factors affecting non-starch polysaccharide digestibility in poultry. *Cahiers Options Mediterranéennes*, *26*, 125–134.

Williams, D. C., Van Frank, R. M., Muth, W. L., Burnett, J. P. (1982). Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. *Science*, *215*(4533), 687–689.

Xue, G. P., Patel, M., Johnson, J. S., Smyth, D. J., & Vickers, C. E. (2003). Selectable marker-free transgenic barley producing a high level of cellulase (1,4-beta-glucanase) in developing grains. *Plant Cell Reports*, *21*, 1088–1094.

Yang, J., Zhang, Z., Zhang, X. A., & Luo, Q. (2010). A ligation-independent cloning method using nicking DNA endonuclease. *BioTechniques*, *49*, 817–821.

Yaron, S., Morag, E., Bayer, E. a, Lamed, R., & Shoham, Y. (1995). Expression, purification and subunit-binding properties of cohesins 2 and 3 of the *Clostridium thermocellum* cellulosome. *FEBS Letters*, *360*(2), 121–4.

Yip, V. L. & Withers, S. C. (2006). Breakdown of oligosaccharides by the process of elimination. *Current Opinion in Chemical Biology*, *10*, 147–155.

Yoder, M. D., Keen, N. T., & Jurnak, F. (1993). New domain motif: the structure of pectate lyase. *Science*, *260*, 1503–1507.

Yoshida, S., Sako, Y., & Uchida, A. (1998). Cloning, sequence analysis, and expression in *Escherichia coli* of a gene coding for an enzyme from *Bacillus circulans* K-1 that degrades guar gum. *Bioscience, Biotechnology, and Biochemistry*, *62*, 514–520.

Zhang, X., & Studier, F. W. (2004). Multiple roles of T7 RNA polymerase and T7 lysozyme during bacteriophage T7 infection. *Journal of Molecular Biology*, *340*(4), 707–30.

Zhang, Y. P., & Lynd, L. R. (2005). Regulation of Cellulase Synthesis in Batch and Continuous Cultures of *Clostridium thermocellum*, *187*(1), 99–106.

Zverlov, V. V, Kellermann, J., & Schwarz, W. H. (2005). Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes. *Proteomics*, *5*(14), 3646–53.

Zverlov, V. V., Fuchs, K. P., & Schwarz, W. H. (2002). Chi18A, the endochitinase in the cellulosome of the thermoplylic, cellulolytic bacterium *Clostridium thermocellum*. *Applied and Environmental Microbiology*, *68*, 3176–3179.

Zverlov, V. V., Fuchs, K. P., Schwarz, W. H., & Velikodvorskays, G. A. (1994). Purification and cellulosomal localization of *Clostridium thermocellum* mixed linkage β-glucanase LicB (1,3-1,4-β-D-glucanase). *Biotechnology Letters*, *16*, 29–34.

ANNEXES

Supplemental information – Chapter 3

Table S3.1| Primers used to construct the pHTP vector.

Primer	Sequence (5'→3')
p28F	TCCGTCGACAAGCTTGCG
p28R	GCCGCTGCTGTGATGATG
ccdbF	CATCATCACAGCAGCGGCCCTCAGCAAGGGCTGAGGGGATCCGGCTTAC TAAAAGCC
ccdbR	CGCAAGCTTGTCGACGGACCTCAGCGGAAGCTGAGGTTATATTCCCCAG AACATCAG

Table S3.2| Molecular architecture and primary sequences of the 184 expressing proteins encoding modules of unknown function from *Ruminococcus flavefaciens* (Rf) and *Clostridium thermocellum* (Ct). The modules studied in this work are highlighted in bold in the molecular architecture column. The unknown module here identified as an α-L-arabinofuranosidase is highlighted in grey (Rf137).

No.	Molecular Architecture	Primary sequence in amino acids
Rf 1	[(1-29)SIGN][(32-367)GH5_4][**(368-635)UNK**][(636-680)LNK][(681-765)DOC1]	GECHGYIDRTTGEWHEASLPIINKMMEIMDDESIQWASERH IPTVKHQSYAEGTTFLEGPYELDASKEKTYQNTTPGGDGE VEWSQLEGKEVAIKFTGSTPVLCFSDASYGGWTEMKPYDI DKENGIAYYNMAKVPDLWGDDPTTIAHMQAKTPKLTTVES VNILAAPEGEIKEPEATSKIKKINLKDAKNEDTLYVNLEGAPS TKTNGALGFMKGDEWTQIEWSGSTDADGKLTVEIPLADAV VGGTVEFQIWAGFKDLDVKDYSIV
Rf 2	[(1-27)SIGN][(28-271)GH11][(272-315)LNK][**(316-551)UNK**][(552-621)DOC1]	KYVQVVNNSDGGRGEIIPPYIYYSSTTNGSMYAAENGCFSA STDASEISRFYAGIEQPDGLNQYHIIGPDNTVTADYKFENTY KDSYQLSYNLSGKNDYKYTDIIIIENSNDHPLQYNFKDIRSS HRIPDILEPTLCKTYTVNGHEYDLYKDEYDVDGHWISNTYE TYIVVRKDQEEGPILEGSIDFKKHMKQIDEDLIGNFEADSVF CVLETDYTYGTLKALKNDIVFETDDYPY
Rf 3	[(1-26)SIGN][(31-118)DOC1][(119-156)LNK]**[(157-384)GH16_lic][(385-421)LNK][(422-670)GH16_lic][(671-699)LNK][(700-921)GH16_lic**]	DLGTPMNANATAVADFRKGSTPLFFASDGWENGDPFDCG WYKGQTSLDSGVLTLTIDKDKTGKYNYAGAEYRTSDHYGY GYYETSMQAIANDGVVSSFFTYTGPSEDNPWDEIDIEVLGK DPTKVQFNYYTNGQGNHEFMYDLGFDSSKAFHTYGFDWQ PDHITWYVDGKPVYTANQNIPKTEGRIMMNTWPGRGVDG WLNHYNGNTPLTARYQWVTYNNGGAGNNNQQTQTTQQP TTTSTTTTTTTTTQWQWQPTTTTTTQPVQQGAIADKGT PMDTSATMISDFRTGNAGDFFASDGWTNGKPFDCWWYK QNAQIKGDHLELSVDRKWTNDSNPDWNPAYSGGEFRTNK FYSYGYYETSMQAIKNDGVVSSFFTYTGPSDDNPWDEIDIE ILGKDTTKVQLNYYTNGVGNHEKMIDLGFDSSLEYHTYGFD WQPNYIAWYIDGKEVYRATENIPKTAGKIMMNAWPGKTVD DWLKAYNGNTPLTARYQWVTYKNSPKNGGNNNQWQNP WQPQQTTQTTTQGWQQPVTTTQAQVNVQNNGMNKNAT MVSDFTTGKAGDFFASDGWTNGKPFDCWWYKQNAQIKG DHLELSVDRKWTNDSNPDWNPAFSGGEFRTNNFYHYGYY ETSMKAIKNNGVVSSFFTYTGPSDDNPWDEIDIEILGKDTTK VQLNYYTNGVGNHEKMIDLGFDSSQDYHRYGFDWQPSYI AWYIDGKEVYRAYDNIPKTPGKIMMNAWPGKTVDDWLNAF DGRTPLTAYYQWVTYNKQ
Rf 4	[(1-38)SIGN][(39-333)GH30_8][(334-425)X92][(426-446)LNK][(447-590)CBM22][(591-596)UNK][(597-660)DOC1][(661-689)LNK][**(690-1039)UNK**]	QLDFYKSAIKNMGDTSRISAKLRAAENGSPLTIAYLGGSITE GKMYTSPFSSYVKNTFAKGGFTEINAGLSGTSSVVGLVRS EREIFSKKPDIVFLEFSVNDHEDISYKKCFESLVKKIIDQPQE PAVVILINRSKGGFSTQAQMAPIGQNANVAVISMDDALTKA FNSGFLQPGDYFNDEYHPHAKGGQLVADCLGYYFRQAMK TENATPAYTYPSKTVYGNEYSTCYNADPSTDLKNFNAGSF

		TKANGYSSGLYYTYNNSKNGNTPMTFKVDGKGIILVFKANS SGMGSVSVTVNGKTTKVNGNKQYTWGGPDAEVAYYQDT AGELDVSIKMDNAGSDFAIEGIGVIR
Rf 5	[(1-30)SIGN][**(31-504)GH30_5**][(508-652)CBM62][(653-796)CBM62][(797-860)DOC1]	AANVTISPNNTYEINKGIFQGWGSSLCWWANRVGYSDSLS QKAADAFYGDDGLRLNIARFNIGGGDDPSHHHITRTDSNM PGYTKYNNGVVTYDWTADANQRNVLDRCIKAAGDDMIVE MFSNSPPYYMCKSGCSTGNKNAGQNNLKDDQYTAFAEYL AEVCKHFEENWNVKVQSIDPMNEPYTNFWGAFSPKQEGC HFDIGNSESTILVELKKSLAKRGLNDIIISASDETSIDTQIEAY NALSADAKSVVGRIDTHTYGGSKRSQLKDTAIKAGKNLWM SEVDGKGTAGVNAGGMSAGLWLAQRITTDCNDLNASAWIL WQLIDNHVSSVGYNGNKDSGMPNINDGFWGVAVADHDKN EIILSKKYYAFGQYTRYIRPGMTMLKSSGSTMAAFDKKNNQ LVIVAYNTSGSASDINFDLSQFDELGASAQTIRTSPSENWA DIGKTAINGSSLKASLAANSVTTFIINGVKGG
Rf 6	[(1-22)SIGN][(23-142)UNK][(143-433)GH43][(434-587)CBM62][**(588-796)UNK**][(797-858)DOC1]	APEPIVKYDFETSSGGTVKDISGNGNDMRLSGNASVSADG ERGNVLVLDGSTDTYGAMPDGLLDGVRDYTVSMDIKSQSE GDFFTFAQGQDKEKYAFLKVAKDHFRFQTTTDTWRGESG FRYDLDGTRWHNYTLVVNGAVGKLYVDGELVKETSDLTTG PADMGSKQSCLIGKSYYPEDIMFSGSIDDLAIYKNALTESEV KAMLGKGK
Rf 7	[(1-29)SIGN][(30-61)UNK][(62-393)GH43_D][(**394-684)UNK**][(685-748)DOC1]	NGVFSPYEAHSAAELVGTAQIDYDMTDPYAPIVSAKQAGS WTAVRGVQFTESENASQPAAAELLQMNIDTIQYDLTVTSLD APTTITMYASAQNGVKNQSSVEVTGKGKYKVSVDMNSAK GFQIVGCFTAANDTPVTMEVDSITLNGKYNIAVAAELTNTRE WADGLRNIWNGFSDGDAVYTDDHAMLKYVKADDAIELFAA ENAGITNNAPLVEKPISFAASVKGKGSIDVHLDAPTGDLLTS IAFDSPSSFTTVYSDPISNIGGTHDLYFVYSNQGVSMQSWL FTESSE
Rf 8	[(1-24)SIGN][**(25-269)UNK**][(270-335)DOC1]	ADFINPVYEDSESYQKQISEWNTKATHAPELSKDFSDLCNG DIIDMWVNDNINDGIIYVIYQRPDLYSVSIAKDTTDSSEEIEK YLSAQLGSNTELTVTNEPYDEFIIIDIEFNGKDHELNYLLCDK ALSIVSQKYNVVSSAAFINRRFFGEYHISWDACTVNENNFS DFNEKQISDLNKAFTDNNIKAQYDPDSEKIVFSSDITAKEHL EYAAFIKDNYDQKVSMQGLVSAQNEYSQNLELKTD
Rf 9	[(1-25)SIGN][(26-329)UNK][(330-475)LNK][(476-524)UNK][**(525-1000)X139**][(1001-1025)LNK][(1026-1106)DOC1]	MSQANYNQKGPEWERPANIEIFENGKSVVSQNVGIRTKGA ASRAWAQKSFNIFTRMDYGKGEVEYDLFEGKSTKAKNNKV IDKFNGFTIRNGGNDNMAGFFRDSVNQSLVGDRDMATQAT SECILFIDGEFWGIYQLTEKYNSDYFKSHYGIKKNDVAYIKN NSLEEGSDQDLSDWNNLLSEISRADMTSDSAYQQISQKLDI QSFIDYFAAQIYWGNHDWPNNNTAAWRSNSVDPENPYSD GKWRMVLFDTEYSANLADKVNEVGPTFNSFSQFGGGGG WGGFMGGGGSLSGAFTALMKNAEFKKQFELSFMDMANY NFDTKNTTEAINYYKGFKQQIVDTYARFPSSKNIHNASTFDE DYKLLETFYNTRYGNITSMMKSYMGLTGSLASVSISNDGSK GSVKFNTIALDDSLSTWSGKYYTDYPVTVKATAKEGYTFDH WDVTGANVSNTTSDEITVPVSEGVSIKAVYKEGG
Rf 10	[(1-31)SIGN][**(32-549)UNK**][(550-626)LNK][(627-880)X161][(881-962)DOC1]	ADTTKEAVVDGLTYVYVPDSPNKNECTVQLIYDDANKQVT KHDTVSIPEKIGNYTVTTLGSDDKSIMQSKNADSVHVTTIKL PHSIKNIHKNAMLDVNLPLLQTLYVDLNNLEFVSEGVFGYLS AVSEIYVYDKADKAFYPTSEDLDKYRELVSIEGLKFEEIKDK LDWFMISKEEYEKNPNVNGKLEFINAVSTSTYTRKVGYMYA QEAVKKYGIDSDKLSVLQKSDKISNYISGHVRYSFLFPYAET IKDEQKCERLASTALSIIGFHTGVCGGYAHSFEMMARAAM GNDIVDKAADVQCVSVPGHALNAVRPKHSDDNSGYYLVD NTGSVFMQGQGKAVGEYDEIMDGYVYGLYATDQDTIDSN HDIKIVKAKNMFSEGVSQIYLRDETKTPINIELYDKNNSKDK YINFTSYPVTGSTFYLEELPHTKCGEINPLGAGLNLYVEPNI YHEYRISNSKGEAVFGGDGEHKFKLGNVEYVCTITTRDYN SESPYGKMAPHTANKNYFEVVIKQLTD
Rf 11	[(1-31)SIGN][(32-549)UNK][(550-626)LNK][**(627-880)X161**][(881-962)DOC1]	IPNIIPPKGKVLEYTGKPQELIVPGQTTGGTIVYRNGMLNGY YEEIPTGTEVGTYSIDYMIIGDENYYGANDDYAILNPRQGRK NYAFPYCFNCSREPAYPAKGENTNLFKISNPLVDFEKFGTT ASGKWTLEYVSAYGNFDMSLSKHSDILKSIAKALQQSYALT DKDINNVSIYELKDDGKHIAYGCIFSTAANDAEVLFIGDTWG NSGGGGYVLTNEVLSDRKTFTAAQDLIGISNQKMTLYGSVI AEIRE
Rf 12	[(1-21)SIGN][**(23-618)X134**][(619-826)UNK][(827-890)DOC1]	MQADAAEEEFAVRDKWGYCSTANYAESEHFVIFYGNNDTT GKVNDAFIKRNLEAYERLWHCYTEYLGMTNLNVDIYGKSTK KYKTNIYLTYTGLDQYKEGWAFMSSEDGYGIEIISPEAMLD DLTIAHEFGHVVHYQQHNWVDQEISGAWWEPMANWFRE MYLGSSYNPTDTKTGNFDPYLRNLSLALPHGRNYYETWPF LAYIAYNPDNLEGLGITSIHRLLSESKPDEYPLDMITRILGTD AHIVLGNYAKRMVTFDFGMKEAYREQFRKVMNQTPYYWN LFYTVPDQTAEGAYRVPEEEAPMQGGLNIIPLEIKGDDITVK LNGLSDDPNAGWEACLVTVDKDGNSSYSQLFTDGEVMSIA ANGAESAYITVIGTPKKFVRENAFHKEKDSSYKNGDERRRY

		PYEFTMTGADIVKSGGYSKSKGKAHPNGGGFVASTAKVDD SVYVGPDAMVLGNAVLTGNVRVEDHAVVANTVTASDNVVI SGHAVVDGGGWIYVDNGWKQGAVRLSDNAVISDSAVVAG GVTVSGNAKVLQKAYIADGVTLSENAVAKGMAYAYGKGGY SGQVILDGDYANEETLKSGIGFGWLDTANPKY
Rf 13	[(1-29)SIGN][**(30-578)UNK**][(579-645)DOC1]	EDFTDTSVTDINEEDNTIYGDVYFPMGRVNGCEMYAGDKL DLSNIPLELVTYSSYDNNYRNPILYHCEFTVGSGLYSDMYT LDTSQVDMNTPGDYKVIVRPKKGAVGTFTTKDNHTSLNPG YAPPDGDYDICMKGIESYIPVKVYDMEEAAESPLYLKFYTE AIEIRSGGGTMMELVGAKASKVKYEVADTSIANIRTANTSNK MLALDGLKEGETTVTVYASDGRILTEKIKVLPPVEVPEEPET DIRTGTTTTVAGGTQYVPTTTSKSTVSTVDKWWYDLETTTT TIAHTSMSETETTTTTTTIHIDYRTLAEYDKSPMKVGTTRKII FKHPETGTADDLYYVGDATDNIMVTHEKGTNYVIVTALSEG KASFYAAAKGCAFPVSVQLEITAADEFTGMPEKIVYQKGEE LDLSGIKTADGKDAEIAPEEITLTGPVSSVKKHTAKEFATLE DGKYIVRAGNLTFNVYIDDPEDPSRYVQLKNARVTEAEVKN GPVYVNFDGIDEAFYYDADAGMRADWTGVTMKKGDVVSG VLRTSERGSNATYIY
Rf 14	[(1-177)JUNK][(178-200)SIGN][**(201-772)UNK**][(773-852)DOC1]	KAETYEPEVSESALSIKQFVEESNIDLDINSDGKLDVFDVYA AYRCETGNVDGVPVNIREKYDALPQKWENASGEKGYLNY DSFAEYYFTYYDFRPEYFDPNYYIDNCPDTYDDELPLDVIK QAISGIEKWDDAAFTRTIWYVRNDDGTFRPFCEDDIDRAYF YDEESGKYTPDNNFFAFEAAASPIHEFIRNFKSYYTAELFLN NELMDLIINSDMTDVDINSDGVYDFDDIVLIEHYLNNFIFTGT YEYDSIIAALDYHETSSYQENYPAELREAAYCPLSEDEWNK AHDFIDTVRYYFAHDALIIKCMTENYLLENTVDPKYFDPVYY AKNHLAEYEFDKCWIGTKGELFGTLGYYEYFSLKYGPGSE KNKALFELADKREIYYEDEVNAAFPTYYKNVKTGKLPVPDM DLDGETGIADYLILEEIDDEYVSIGDRDPFAPLVRRYPEIKAE ISISPEARENYMTNFDFNNNGISCDFLETECMRMYIFGELES QYESREDLCDAIYNYLNTHPDMKYSRISAEKMEELNQKYG VNFTPVEDEEEEETYDLPESVDIIQNYSSFNQVFN
Ct 15	[(1-28)SIGN][**(29-470)COG5337**][(471-533)DOC1]	ASRPEGWTEETHGKKATPNYSVVFPEDKVNRIDIIISPENFQ RMENDVFKVFMMSNEDPIYVSATVKFNNHTWWHVGIRYK GQSTLTGAMMSMSHKYPFRLNFDKFEDDYPEIDNQRFYGF DELIFNNNWYDPSFLRDKLTSDIFRDAGIPAPRCAFYRVYV DTGNGPVYWGLYTVFEDPSDKMLEYQFENPNGNLYKGQQ APGGDLTIFDKRGYEKKTNEKADDWSDLQALVAALNAPKT DPAKWRADLEAVFNTDSFLKWLAINTTIVNFDTYGWVTKN HYLYQDLADNGRLVFIPWDYNLSLSSTNPWGIKPPSFSLDE IGRNWPLIRNLIDDPVYKHIYHTEIENTLNIYFREFNVIEKARR LHELIRPYTVGSEGEIKGYTYLTNGEAQFNQALTQLIEHIST RHREARSYLSSVNYYTPIPERTPTPFPSPTPKKP
Ct 16	[(1-30)SIGN][(31-176)X140][(185-330)X140][(331-490)UNK][**(491-973)X139**][(974-1051)DOC1]	YENRGKEWERPVHIEYFETDGKLGFSMDMGLRIHGGYTRK YPQKSFRLYADHNNDIGEIKYEIFPGLRGTGTGKKIKSFERLI LRNAGNDWTGALFRDEMMQSLVSHLKIDTQAFRPCIVFLN GEYWGIYHIRERYDDKYLKSHYGLDDDKVAILDVYQTPEVQ EGDSSDVLAYTNDVINYLKTHSITEKSTYDYIKTKIDIENYIDY YVAQIFFGNTDWPGNNVSIWRYKTDDGQYHPEAPYGQDG RWRWMLKDTDFGFGLYGKSPSHNTLAFAAGDIREGQANE EWAVFLFKTLLKNEEFRNEFINRFADQLNTSFVPSRVISIIDD IVATLEPEMKEHTDRWPFIKLTATSPWDTTWSQEVNRIRNY ANSRPSYVRQHILSKFRNNGVTGTALVTLNTDSTRGHIRIN SIDIVSDTPAVTNPNRWSGTYFKGVPITLKAIPKEGYVFDH WEGINGSVEASSDTITVNLSNDLNVTAVFRP
Rf 17	[(1-27)SIGN][**(28-413)UNK**][(414-704)GH2_6][(705-1363)UNK][(1364-1444)DOC1]	ADISFTHKEWTGQSGAEDIFAVNREAASVNPVPFHDDASA VNAVWDYNAREQSDYLQMLTGENEDWELNVVQNEEKAA PYRWGGFMNADYKGKDGDGWKTVQLPKSWTCLGFDFPI YDNVVMPWQSNYDKYVPCPTAPTNYNPVGLYRKKFTLDS SMKENGRRIYIQFDGVESAYYVYVNGKAVGYSEDTFSPHR FDITDYLKDGENLLAVEVHKFCDGTWFEGQDMIYDGGIFRD VFLVSSPSVQISDYTVRTDLDDSYTNAELQLSIDVKNTTGNT VSGWTLQADAYDENGNNILSGASTAVDKVNGWNKGTFNIK TKVMSPKLWSAEDPNLYALVLTLRDDKGNVQEKVSTQLGF REVGFTPTQVDNSYKVTTKQWQPITING
Rf 18	[(1-27)SIGN][(28-413)UNK][(414-704)GH2_6][**(705-1363)UNK**][(1364-1444)DOC1]	QSFWFSANSQQLAANTVSVYNENNFLDLSEFNVNWKLLKN GIAIGSGTIDDAQCAPLSKNSFTVPFRLPEKYYSGDEFILDIS VTTKKATDLLPVGTEVAYEQLNIDSSGSSAKYNSGDSSVTV VDTPDAYVPTNEHNDFNFSINKKTGLIEKYTYKGDLLIDKGP TPNFWRGNVENDGGSARSKLFDTAWENAMNGAEVIGIDT GEGSNGAKTVTSHLNLPKAGNTKVDIKYTIHPDGRVDVDFN VDATRSGLGNFIRVGSMMTLPEGSEQLSWYGNLTESFND RKSGGRQGVWESTVSEQFFPYMKADDTGNLTDVKWISVK NSSNSSGLLIAANGTVEASALHFYPEDLQKADHVYKLSPRK ETILSVDYGSMGTGSATCGQGTLEKYRLPSGRTYKWSYSII PVSSEADGKALSTTAAKLRSDGISVQDKSSNALTIPVKSPA

		VFKSTSEGNAVSGSLSIPSGNSIGKSLEGKNSFTVEAEFVP TGNPGFNMIASKGDHAFGLRTENGMLYFFIHAGGEWRTVS YKTGTDEASGWIGRKHQLAGIYDAENNMIKIYCDGKMVAE KSTGTSSGITSSSYDLTMGACPETGRTSMADFYEFRVYSK ALSESELASQRTASPAYAPDSPYVKLWLDFDNIAENEAIDDI PDDIPQVDP
Rf 19	[(1-32)SIGN][(**33-282)UNK**][(283-890)GH97][(891-906)UNK][(907-1028)CBM35][(1029-1095)DOC1]	DIYTELHAVSQGSNSAHAVLKGADASVITDNSIGSDVLYLK GSHNGGGWLQLPSLFESGCGGGFTLAMKFMLKEGASDYS RLFQFSPVPFASGNSSSYSSPDISIDLKDKTAFRASIFAGSG MDTENDKKHRAIYDLSAAPDTDKWHDLVLVCSPDGAGYYI DGQKLTYSSETVSDVVNSLFSENVLSSYVYNSLGRSLYND DDIAACFDDVAFYTRPLSGTEITSLPDDADYLYTFEKDTLEE GEAVPV
Rf 20	[(1-28)SIGN][(29-99)UNK][(100-120)X159][(121-195)UNK][(196-216)X159][(219-239)X159][(242-262)X159][(265-285)X159][**(286-501)UNK**][(502-583)DOC1]	HGAFENTGCIETVDGIDYVDNWAVDGDSNSLKDAAIREGT RGVAEFAFLLCNKTEHLSFPDSIMYTLPLCYASSKGPAVTID FSGHSIGERAFTGAKKLTDIYIYDRECDIFDDEKTIPETFKEP TELDDDLIIDSGSSDNDKISGGNSHSKSIQGPSGSELVIDEE LPEEMPYTASPVITADEETKDNRVTIHGYIGSTAEAYAKKYN RKFQPID
Rf 21	[(1-36)SIGN][**(37-244)UNK**][(245-398)COH2][(399-489)DOC1]	SDVKEYKLMGVTYSIYSDHCEVTDGSQASGDVFIPVSIGGQ TVTVIGGNAFKGSSITSVSMSSVTQISSGAFRGCQLLETVA FPSKLATIGSGAFADCPKLTEADLPQSVKSIGEDAFSGDKS LKTVTVRNPLCEIGDKSSTLSGTAVTISGYTDSTAQKYAEKY GFTFQSLGVSPLTTTTAASTTTRTTTTTAKPTTTSTTTKATT T
Rf 22	[(1-34)SIGN][(35-113)UNK][(114-134)X159][(137-157)X159][(160-179)X159][(181-201)X159][(202-216)UNK][(217-237)X159][(240-260)X159][(263-283)X159][(286-305)X159][(307-327)X159][(328-338)UNK][(339-359)X159][(362-382)X159][(385-405)X159][(408-427)X159][(428-464)UNK][(465-485)X159][(488-508)X159][(511-531)X159][(534-553)X159][(555-575)X159][(576-585)UNK][(586-606)X159][(609-629)X159][(632-652)X159][(655-675)X159][(678-697)X159][(699-719)X159][(720-737)UNK][(738-758)X159][(761-781)X159][(784-804)X159][(807-826)X159][(828-848)X159][(849-866)LNK][(867-887)X159][(888-912)UNK][(913-933)X159][(934-960)UNK][(961-981)X159][**(982-1714)UNK**][(1715-1791)LNK][(1792-1878)DOC1]	DSLRCSDSNSRISNSGYGYFTVYPESDELTVSFSCRIKSKK DINIDPTVTVTMDNEHHTPIMGSVELQASAISLGGPRYVSEL SAEVYGIAEKGQEVTILVNGKAADTVTTSDKTGKYRKVITLP DGNAGDEYTIAAKCGKNTSADIKTTYIKDSPVLKSADFSDS HFRTSHDITTVFTEGKSPVIVTYIGSSFSFRMKMANSDKIKH LYLTSTKGGEMKYIEAEYDKAKDEWTASGRFDESNRYYIP GYLNIAVVTENDYPTIDIDNDEPADSFRINNYTDDISKNSSA ETLFADDNKLLAKTTISNGKLSIDYGYFSASADSIKIGGKAVS AKDAAAAPDKNGFTKLPLNVIEEGEQSECYYRIMGADDSKT LVEGLFDKKDISQFSSHKSLVLIKKGSTEPSFMIHGVSDLST DDTDLFSFPMIGDMGTELAMYAGDTAAFSKLLAESGQDGL LDVMTMLYGSKFVSILAGSELKTSVNTVGGLTGPWTLAVD AAILLGEAENYNCYGRVMSEKYPLFTNPGCIRLIVDPSGKT YEAVKTNPVEDVQVTIYYKDENGKEVKWDPEEYDQENPLM TNSDGGYAWDVPEGEWKIKAVKEGYEDAESDWLPVPPA WTDVDIAMVSYEAPVLKSAECKDGKITVHFSKYMDIETVSS DNFTATGYSNISVVPVLDSKGDVYADTFEITGTVDKTAVKD GTVTIKASGKADSYAGTAMKASEVKAKVEGDITAVV
Rf 23	[(1-25)SIGN][**(26-232)UNK**][(233-253)X159][(254-448)UNK][(449-468)X159][(471-490)X159][(493-513)X159][(515-535)X159][(537-557)X159][(558-732)UNK][(733-803)DOC1]	DAESSDYHSQSSISFNYPYDKADLSLWQYKVLDSLDGVYD KPCIELTHCSSTDKTIVVPSEIEGLPVVSLGQGVFSSDPYLE ESTIYFPDSLQHFDRNFMPDENSILIYTESGDKYLCYSYFNE NTGADPKHLRLLQCGNRKNIVIPEAIGNLPVSETGIYLLQYA KDAESLELPDTITYFDEYLLGESTSLKKLKLPAHINILPSH
Rf 24	[(1-33)SIGN][(35-115)DOC1][(116-147)LNK][**(148-1104)UNK**]	ITYIHLKGSSITVDGDNATVSGTTVTISHSGTYMIDGTLDDG QINVNIPDETVDAETVKLFLNGVNITGKSAPAILVTNAENTSI NLVDGSANTISDGDTAYAGDYLGAAVIEAKDDLTIKGGDKG TGTLTVTANTQDGISCNNDIKLTGGIINVTTLNATDKTDAVK GKKSVTVKGGTVTVDAEGDGIKSSKGAVAVEGGNISIKAGN DAVQAETTIDISGGTLIAGGDRGLTAVTAVNITGGNVYATAT DNQADDKLIKSDAQPVILLNCKDDATNEKDGTWKKSNILQ WESMNGTGQNVTAEFTKKFKYVLISSETIKAGTTHFVNTAA GKWITHTNDQEGLFPVSNGVNIFENVNLAGAEAGVPVPPS TETPDTTTDGYTITLGSAMATNASAEVASVANNVCTIKQPG TFTVTGEMTGGQIVVDVDKTAYPDGVVELALSGMSLTNTS DSPIYVASIGDEVVISAKNGTENTISDGTSYTNADSDTGAIY SKDDIKFKGKGTLTVNGNAADAIVGKDDVKIYNGNLIVNAKD DGIRGKDSVTIGNTSSDGTEVDYSNLSVKVKTEGGDGIKAT STEASSTAKQVGIVTVNGGAVNIESYADGISAEQFFVMNGG DLNIKTYQGSGFTGSAAGGNTGGWGGGFGMGMDGNANK TDISAKGIKAVGLYDEAGTTWQSVGNIDINGGNITIDSSDDA VHCGGSMNLYGGTYTIASADDGFHSDHELNIGKTAANTFD DVQIYISKCYEGIEGVTINQNSGTVYIISGDDGYNAAGGADG SGFGNTGGGWGGGMMSSSTGTLNINGGLIVANSANGDHD

		AIDSNGDINLNGGYVCANGQEPLDCGDSGNTINYKGGSVIT MTAGNTNLSQRYSFVDNSGNVIVSFISASGNPGQNCTNCT AQSGGTVSGGKTVNAQSDKYSVTVGGTISGATQITAAASS GGGMGGPGGRQPGQPW
Rf25	[(1-29)SIGN][**(30-464)UNK**][(465-537)DOC1]	NTTLADPDLSYLYDEENNYYCIAVGMHDVECIVPTEYNGKK VGELNLDHVFIADYDFPPDVKTDTVILHVPDDIAVDGKYWL AEQTGVPCIMLAYGSGKTETFLSADYESLLEKIKTESENDIS LTEAEYKNMMLQIVPYNTGRENREYPILETGFKYDRGFSTY RENGHTYIKIMAFMAGKDIYLPEVLNGQKIDRLKLGDITDPR GGSPQIGKLVFPACPMEIERSTLNKPEIKEIVFGGDVTLPKM AFYGNEFLENVTFKGKAELDNTAFWKCEKLKNINISPDEFP AGVTFNQCRDLMTINGESPVNDDGSIKPEYEKIFKEKFYNT DGIGFVNKYVDYRVKQAVSEAVTDDMSDMEKVKALHDKLC SMTAYDHGNTEDPKNHVDVSVFLNDSTVCEGYARAMNLM LHEAGVESCYVDTDTHAWVIVKLGD
Rf 26	[(1-27)SIGN][(90-331)pfam00112][**(332-988)UNK**][(989-1001)LNK][(1002-1075)DOC1]	MGDKENYQYNFQHDSYIPVQTMAAAEDDADLAEGTPSYM ANVFKNDCQCQIEAVSTYFMNPSTDYEVTVYSGLQDPADP SSGTPSSVTKGHSDLTGYFTIPLDEAVPVGGDEYFSVVVKI SSAESAFVVPLETVLIAKDRETGEIENIGSYTTYDGICWYTG ENESFFSPDGNEWSSSDAGNYDYSEEEKEELLQIFSEELY DGLEEEDVEEKERADRQMAHYTELFEHSDVSIIMGNISLKA FGNSVDTVHFSHPSGAVPLNECIELTASGVDKILYHITDENG MSKEFEYTEPIPVKKDEVLVAHTPESGLSKRNYHPAKAEFF SFGYDVTPEYYSPKLSYAEKISASKYHIELPTANDKVRFFPV SDCDITYNGETVFNYQMTEQFDIPLGETVFEFELKKENALD NTVTVVVSRSPVSFDTETEKLKISGNSEVYAPDGTRLITGS DVGAYAGQRLTVKDGGNEFEISVPERRKIADRVIDYGREDI LFFEEFGEKDAQIKTGNSTEFVDLDGRISSHHTEEDGIVRT CVRIIPGETFTFRMKATDKLFASEELVVKVPEAPDFPEKMP AYTIQDGEPVFEDDTIRCIFIPEDEKQPIENYLEYYKYENDR EGFVKLMSDRYGVDNEEDLSTILWALNLINPTDVSKTQYIIM
Rf 27	[(1-24)SIGN][**(25-229)UNK**][(230-249)X159][(252-272)X159][(275-295)X159][(296-495)UNK][(496-516)X159][(519-539)X159][(542-561)X159][(564-583)X159][(586-606)X159][(608-628)X159][(629-700)UNK][(701-761)DOC1]	ADEETELWNKFLKYDLCITDYDSLTEKEQELCHFIYDTETRA EDTIVCNRARAILAGYDVGNRITVEQAEKYKHIVNPEDFLFY NGSINDYYEYEFPSLLTVPDISHIDEPDICNEYWLDDTKSSAI IYNSNGLYIQNYNNEGEIEYSELIDTAIKEKTDIEKNGLVFTVL PDDSLSLTEYKGADKEVKIPSEIDGHFVKSIDIG
Rf 28	[(1-25)SIGN][(26-48)UNK][(49-149)X142][(150-200)UNK][(201-310)X142][**(311-759)UNK**][(760-831)DOC1]	SCNMVWQDNAPNISLNKIAFTCSVWTDDSPVYAENLYIEVP ETINVIYDHWKTSITHIMQITLEYPSGKTVTYKSEDYEDYLQT SGRFKEYNRIMSDYDWMRQVDDVMGEDTAWILSSPDDYE RLPQYTDADKLVVGNISDYTSAPYIEELVFPDNIKKLTFGYS AFDNTRIGKLVLPDCPVNIDRTTWVNAEIKEIVFGGDATLPS MLFYGNQFLENVTFKGKAELENTTFWKCEKLKNINISPDVF PAGVSFNQCRDLMTINGESPVNDDGSIKPEYEKMFKEKFY NTDGIGFVNKYVDYSVKKAVSEAVTDDMSDMEKVKALHDK LCSMTRYDHGNTDDPKNHVDVSVFLNDSTVCEGYARAMN LMLHEAGVESCYVDTDTHAWVIVKLGDHYFHVDPTWDDN DEDITTYNWFMKADSEIKDDPSHSNWKMRCPSIMHNFQW EKMP
Rf 29	[(1-25)SIGN][**(26-329)UNK**][(330-475)LNK][(476-524)UNK][(525-1000)X139][(1001-1025)LNK][(1026-1106)DOC1]	APALISRAASSVTINEVCPKNTTYRAADGNYYDWVELYNSS GSSVDIGGWGITDKADKPYRFTLPSGTVIPAGGRKIIFFDAT AGETDTSIAPFGLSNSGETLTLTDASGNIASQITFEALASDN SYGQYPDGSGEFYTLSATPDSANTAPEGSNAVRTPGFSAE SGFYDNGFSLSLEVPEGTTVYYTTDGSDPTTESEKYSSPIT VKDMTSEPNKYSARTDITAYTDILAPDEGVLKAAVVRAMAV DGQGRTSDIITKTYFVGSANVEKYRNMKVISLVTDPDNLFD YEKGIYVKGKVYDDSN
Rf 30	[(1-26)SIGN][**(27-314)UNK**][(315-382)DOC1]	AAEISPPSPEAVSAEARIVYERFDSVTEAGLYVRENLKRHT EELHIILSPWSGSADILNDVLGVAFAETGRGDEGDYLRLSIE GYSSYTGYMLLDQVLDIRFNYNSTIEEEAAFAEKEAEFLAS MDIDRMDEYEKITAVYDYLVKNVDYAENFERSEVYTAYGAL VEKVAVCQGYIQAMYRILTDMGVSCRAVNGEGNGGDHVW GIAAINDTYYLLDPTWDSQFDGVFKIFFLKGYGDFDEYSSP VVHITGTGDERNSAFVPDCTSESFTMAYPVAESAFDPQTY YDS
Rf 31	[(1-35)SIGN][**36-620)X231**][(621-966)UNK][(967-1033)DOC1]	DGSRISDFSISDVKMTDDYCTNAFEKEMKYLLSFDTERLLA GFRENAGLSTNGAKRYGGWENTNIAGHCVGHYLTALAQA YQNPNVTSDQKDALYKRMKTLIDGMQACQQHPRGKKGFL WAAPVPSDGNVERQFDRVEIGKANIFDDAWVPWYTMHKLI AGIVDVYNATQYAPAKDVGSALGDWVYNRCSGWSQQTRN TVLSIEYGGMNDCMYDLYRITGKDSHAAAAHVFDEDALFQ KVSNGGRDVLNGRHANTTIPKFIGALKRYMVLDGKTVNGQ KVDASAYLKYAENFWDMVTTHHTYITGGNSEWEHFGKDDI LDAERTNCNCETCNSYNMLKLSRELFKITHDSKYMDFYEN TYYNSILSSQNPETGMTTYFQPMATGYFKVYSTQWDKFW CCTGSGMESFTKLGDTIYMHDNDSLYVNFYQSSVINWAEK

		NVSITQESTIPDGASVKFTIKGSSDLDLRFRIPDWIDGTMGV SVNGTKYSYKTVNGYADVSGSFSNGDVIELTVPSKVRAYP LPDSPDVYGFKYGPLVLSAELGKDDMKTDSTGMWVTIPKD KKVASETIKISKQGQSVASFMNEINE
Rf 32	[(1-620)X231][**(621-966)UNK**][(967-1033)DOC1]	HLVRGSNVLTFTLNDTNTKLTFTPHYKQYQQRYGIYWKFV PNGTVIEEKLPRAKTSVTDTVQPGYGQYESDNLHGMVEVG TVGVTNDSTYRYVKNGGWFTYRMAVDETAPQLRLRVKLR KTDNGKSLRVRVGDSVLWAGTLDYSGNKDVYDLLLTIPQD VRDRCTYSLTADGTEHSVLDVTFSPDEEGKESAKICDFIYM EAVTPAYETSSDIAYFVDCGDHKSDTATGKDRFGIYNSVTE QLYGPDEVSGKNWGLIDDSTDQYNGSSKSGGLYTANTWC DEANTADGADKSNSFRYTKNQYENNIARHLDYGFELPNGT YSVEMCFCDPWGCSKNPTAYANLGKD
Rf 33	[(1-30)SIGN][**(31-454)UNK**][(455-523)DOC1]	EIPSIRIQNTTYRDTGYVVQTSDNTADIFLHSGKNYVKSKSE DMCFYFNYDPYDKPDELLELNCVFSDQMLEELTIGQMRGE FDYDLNYALFNPPTLRKIIFKGEYQKFVVPFPAIHGTTELKTI VFPEKCDLIEIEEKGIFNVGLEEIEFPHYTRLYYRAFSNAPLL ENITFSNGCDIRSLAFDNCPSIKNLTFNGECNVGALTFTGNE NIENIDISDTNKTTFHWSAFNDCVNLMTINSECLFDSETHDF TGKYKQFIFDNFSAAENVGFINEYLRMQVGKIIDENITPDMS DIQKVRALHDWICNKVDYDHDDANAAKNHRDTSVLFNDRT VCEGYSRFYDILLREAGIESCYVDSSDHAWNIIKIGGHYFHS DTTWDDLDSSYKWFLRSDSEMKKEGGSHGEWTLREPSPL HAYDSTEVP
Rf 34	[(1-24)SIGN][(26-103)DOC1][**(104-381)UNK**]	EFFINNGEMPLDPFGSGLESDIDADRLLGGSWEFLSKGDPL STPSYTADFFNYDREFRISMDKTGIYGTGYYETSNYFDTPR GCYNLLRNTEEMNIKLSPNSPEMHFGGMIDTHFTAANVDG AYIMAISVIGNGISVFDQILGPETQAADGIWILRKNSNETHHII DDTEMASLREKGTTFYAYRWLDLGSEVYLQTVNAETKEVV LDDGMPHEAMFFSYANNGHALTAVKYKIAGAEDKANSGKY RPALVKVTTDPSGNVTEMEEVPRFAYGYYHVL
Rf 35	[(1-20)SIGN][(21-216)UNK][(217-297)DOC1][**(298-727)UNK**]	PATIQTPPKNNSYLFDSYDDLAEALTKADSFKTADSDSYGE LFNNTVSAFENKDITLYVPALNGKECELMNKEGFSNITLLTS ELYNLPWIWYHCKADNSDIDIKLAYHSIIENDALNSAKTYYDI LKLIAPEAPNPDNYTEFESYQKIYESKISLANDKNVDAMISEI KDSNKVYVMFNYDGMLVSIYADQNALTEEFWNSFSLNNLI WKLGKTDNDKSIDNSAIAGKIFAYEKEGAGGYCTLSFNENG RFLYYPGRLSSYMSGGDWKIDGDTVSLIGMVDKTIYLKITD DTLVYIAEGSDEFPYMEIKDGEKFAIYRPEISSDKFQLNSRY SEYGLGDPKVELNLIASELPEFCYVENVRLYDEDDNFIGMM SPAMDADIWSYLVDCNVTEECSKTYYTLTKIRCGAKNYLDD VRSEITVNFKVAPAP
Rf 36	[(1-27)SIGN][**(28-745)UNK**][(746-822)DOC1]	FPAARICTPITAEAAKNEVRIKVVDMATDMPIENVDAAASTV KDFSREDSRFSWNTSDAVEFPIEGAYNNDWYVKLYNVPEE YNYNAVYTIDRDTRTFGDYTIRLVKKSQDNNVTICSAHDEP VAKVLLKAADMDVYDKKGNWFCTIKSNEGTYLPDGEYSVK FKENVLENSGVCVKGEASQKLTIKDGVPTDTIFFLLASDEEK KPNVSFTIKNEGEDTSSTEDLGRVVITGDGFTLDTDGSAYL GDGHYTAHRYNFPKNGFNQTGNEERDQHNIDIIADSLLNKT EDIEFNVVDGKPDRDLVFVKVPVNRPAEEEGATAKIKIVDK ATGKNIEGVDVEVIAGINATGKSLAKWNTSDEEEKVLTGLS GNPNLAFGIDLSNVPEEYDFQKRYIFGFAKDSKEESWVVEL EKKNGPVAEEGATAKIKVVDKATGKNIEGIDVVAISGLNATA KSIAKWNTSDEAEKIIKGLAGDPRIVYGIELSNVPKEYEYNK QYIFSFAKDSKEESWVVELEKKNGPVAEEGATAKIKIVDKA TGKNIEGIDVVAISGLNATAKSIAKWNTSDEAEKIIKGLAGDP RIVYGIELSNVPKEYEYNKQYIFSFAKDSKEENWIVELEKKN GPVAEEGDIGFRLYSRYKGLWDRKDNEMVGYVIITDGADD FIGKYKLDEMISLPDGKYKAEIEVNSKGYSCFSEQKIQFTVD EGKAVENLDFNVERWN
Rf 37	[(1-19)SIGN][**(20-496)X160**][(497-571)DOC1]	YTVSNAEDDTALYKLASSLGADTDYLNIPNFKHHVQERAFS HEVYLRFLENCTNYEASNNPEGELFMGLSGGLCYGISAVQ MLSHNGVISPSDIVNGAETLSEISYSPETDVILSGYHTSQVY YDNNYLINYRPIEINSRTQCDELIKTAERNMAAKRYFVILYA GDNAHAVTGIGIADGSWKYGGVEFDKCILTLDSNSYIDENT ADPFRKESCIFINSKTKEYYVPKYKFGTISGSVQKQIITIDDD EIVNYHGVIAPTVSIERDLSETASMILEKSDLKRYDVTVKDK EGNEHSLSELGNKIGFIGNAMYYIQGRDFHIDVSDRTQFGK NYNDKFSISTQGWYFEGETVNEHGIFDVNGQNQSVSAKG GEKTGYIMTVKYNEGNYPCTPHFNWSFSGKTDSNLKTEIT EKGMILHSDGSIETKISTADVTFNEKGSISDAAAFPCEETITA VNDVLVTFDDNNKLCFYIDPDGD
Rf 38	[(1-29)SIGN][**(30-346)UNK][(**347-417)DOC1][(418-656)UNK]	AETQTSAVRTDYLDSIRSDIEAFMNENHISAQTYIAPISEGK DCLHVLFYADQDEEMKKTEEYLIQNDIPYLNTVIDTTTGRTK GFVILKPNYIDSFKSKVKAFMNENNISGKVYTDGFKESEKIIV DVMSDSDVQTVKNFIETEGKGAFYTDITPLGPDKGIRLCAG EESLMNIQGDVAQFMEENNIRGYTYSGDVITVVCVEKEDIE

		MIKSFVAEKGYRAERLEYTLPDFEVDPPETAPTELEDIRMA LDDFINQQGINGYTKIAPHKSKDMVWVILDPFGDEYQRKINL FMYLYGIDENSVLVTAMTSNAGITN
Rf 39	[(1-29)SIGN][(30-346)UNK][(347-417)DOC1][**(418-656)UNK**]	YYTQTIDWYMDDLPSGVYDICMGNGKSAVITNTDELKEYIA TVAPQKSIDSYQKKYSDSFFDENVLLINSVNQGAGTDVGYE FSNIDLSDKEFNISLKGTIGYDQPAASVMSLCIAQVTVPKTA YHGQPVNWVKDGNSNNISETEQLWCIFVSSLEWNGLTYH DNDSVDTGNYTKDAYIGKVSDFKGAYKDTVNYRINPDDSV YTTKESKDVLIVVKADAYSPYGAEVAMTSADYAGTS
Rf 40	[(1-29)SIGN][**(30-387)UNK**][(388-453)DOC1]	AETSTAAPTVIQTTTIPSNNQITTTKPESPQTITTTPAIKRPTL EEVVEFPTEGKFPMKFKVVDEASKEVVKGLDMELYTLDEY SVDANFIDKLAEWNTTDSETYSCELPYSFEDRNSYSAYGV VIKNMPEGYVYTFSGENSKCFPVDYRPMTIWADVINGNDT HEHEYVIRIEKEGTEHNYVTSTKTNTQTNVTTETSAIHQTLE EIIEFPTEGKFPMKFTVVDEASKEVVKGLDMELYTLDEYSV DANFIEKLAEWNTTDSETYSCELPYSFENRNSYSAYGVVIK NMPEGYMYTFSGENSKCFPVDYRPMTIWADVINGNDTHE HEYVIRIEKEGTIHNPVTSSVSDSDVITTTT
Rf 41	[(1-30)SIGN][**(31-553)UNK**][(554-620)DOC1][(621-803)UNK]	AAYNWDNRPDWTPDDFNSAMEFLNHHGTTYAEDGMICIV KHVPNGKHMAAKIEAVGNEGDQRTDSRYESKIFSFGFEVP DETQPENGEKTALQDYMEFHGYTDGSFVDYHYEALMIKEN QDSKFNIMTGIVGEDEELDESKAAVYTFDGDTETDIWGWL PDSYVEYRIFIDIYGNLSNHDGKLVYCHDVNYSTGATLKVD QQGEGKLKCTVNSSTYNEMVMLRAGDTSHLVQVYEGEEE GDVTVSFDSNIPWAPAETKDPVLTADLHVNADLTIRDKIYDV PEWIPQDSESLIDFYNKHGKIWIQDGLICTIRPVTDYKSERY SYSFGGSAADKIKQYTIFSKSINPFEEYSSILYDVNVYDIPKG TDLTINYDLTYAERTSLNSFVFEKDATGNVTQKDFYAWLPD CVEEYDAYYEKHGAFSIQDGYIMYCTECPIGTGYDLNVKQS GTGGVAEFHQETITPVENTVLDGGSNFVIKLFKPTKEGVVR LDLYKARIDTDVLPENGVDTAYFRIDKDMKIYEA
Rf 42	[(1-27)SIGN][(29-103)DOC1][**(104-401)UNK**]	SVEVFTAEDRPVVPKIPSYIFPQAVDISTIPSFTDKPYCILNN NIPAFDPDDLNGTPFESYEEPDEFGRCGVCVAAVGTELMP TEKRGEIGMIKPTGWHLDKYDFVDGKYLYNRCHLLGYQLT GENANPNNLVTGTRYLNTKGMLPFEDEIASYVKKTDNHVL YRVTPLFADDELVCRGVAMEGWSVEDSGASVCFNVFCYN VQPGVVIDYATGENHADESYIATTTVTETTVRTTVTTTVTQP PVEYDFVANKNSKVFHRPDCTSVDKMSEKNRWYFQGSRE ELIADGYKPCSNCQP
Rf 43	[(1-16)UNK][**(17-613)X134**][(614-843)UNK][(844-925)DOC1]	LSLCAVPCGYYAQAADEYAVRDPFFNFNKGYNYYESEHFQ FIWGNSGDSAKVNTAFLEGNAKNMEACWHVYMDELGMAP PCESVESYLRDGKKYKTNIYISGTGLEGMADDWAYMSWD SGGFAYMFCCVDSMRYDPPSWVFPHEFGHVMTAHQLGW NSNKYSYAWWEALGNWFREQYLYSDYSTDDTGHGTDFFE TYMKNLSFTFPLGRDYYAAWPFLQYLTENPENLEGYGSEF VKTMLQQGEKDEFPFDQVERLAPADMKDILGYFAAHMAGL DFKKGSSYRARLNELLAQGDWNWQQIYTMPEKITTPDGKG AKYRVPTERAPQFAGLNIIPVEINGSFTVKLSPETNVKGAD WRACVVQQTSDGKCTYSPLFGPDETITVEPVSGADAYISVI ATPDTDTVKKYGLPGIYDDKAMFSESNVPFSSKTQYPYSLV FDEDKDSTNVAVKARKVSTSSSDPWQRATYAPHPNGGGL VASTAKVDATVYVAPDAVVKGSATVKGNVKLLDHAVVEGN AVVSDNAVIAGYGMVAENASVSSNARVDDCGLVMGRAKIS GNAKVIESACVYDDVTMTDNSVAKGIAFAMAKGKLSGQGV
Rf 44	[(1-16)UNK][(17-613)X134][**(614-843)UNK**][(844-925)DOC1]	IDGDYYDDSGKSISKGTSYGWTSTQSYADSRPYTDKLMYA YDFDSDSTLSFSDRYTSTYGVNSGAEWENERTSAKGVLTF SGGNYADIDKGVLYSDSEEIQISFLDREGSGKAQNLLMLGD EKSHISVTLENNNITAEFAVDGKETKSVTAEKAYTSGEWTT LRLITDSGKAQLIVNGNKAADGDISASPKDVADALGYGEGV YRIGSGFNGSVDFVRFFSDEADAPSET
Rf 45	[(1-31)SIGN][**(32-561)UNK**][(562-639)LNK][(640-889)X161][(890-963)DOC1]	ADATKEAVVDGLTYVYVPDSPSKNECTVQLIYDDTNKQVTK HDTVSIPEKIGNYTVTAIGDDKQGIVISRNEDDIHVETIKLPN TIKEIHKRALVDIDMPYFSTLYVNINDLETVGEDAFGGYTRIT DIYAYDKIDKAYYQTNTDLDKFRELVGIDHLKFQEISKTSFDL VLSKSEYDKNKCVNGRLEFINEVSSSPYSRLVAAMYAKEIV KKYGFDDPKLTNLQKMEKIFNYIAMNSRYSTIYTYNADADK RRGMDNLKGTAMSNLGFHSGVCGSKAHAFEALCGAAMG YDIINKDRDILCVGIPAHALNAVRLEHNDKDEGYYLVDVTGD MFMQGVQKGLVDYTAFGDDWNYNGYICGEKGSRHDSDA ATRPMKLINDPNIFYKGISCAYIVDDTKGPIHIEMRDKNDKS NKYIDYTSAAPTSPDSYLGQLPYTKQYSIDSGKALYLEPNM YYEISISNSKGEAVFKEEGDHKFKLGDAEYEISFHTRDYGT ETPYGAVAPHTAFKNYFEIYIKQLSDDPKPDTYPAV
Rf 46	[(1-31)SIGN][(32-561)UNK][(562-639)LNK][**(640-889)X161**][(890-963)DOC1]	VKKETPNLILPKGKEMEYTGSLQELIEAGKASGGTLQYKIGE NGEWSEKIPTAAEVGDYTVYYRVVGNEKYYGADTDADIKG RKNYSFPYCYDCQRNPAHPSKGESCRAYQLKEPLIGDRVF GKNIIKGNWTMEYVSAYKDFESPSKYLIGLVEKTQEADSLS

		TPDDISIYELKRDGEHNSYGVVIAVSEGENELLYFGDTIKYG GGYILSKEAISDSKTFVIEKDLKEIIVETVKLSGTVNSKIFNTA T
Rf 47	[(1-23)SIGN][(24-43)UNK][(44-108)DOC1][(109-125)LNK][**(126-731)UNK**][(732-756)TM][(757-762)UNK][(763-784)TM][(785-1123)UNK]	PIETSLTKQMADKAKELGTAVNVYNYLYNNMRSEFYYGSR KGAIGTFEQGGGNDSDLSSLLIAMLRYLGYDANYVTDIVGF SAEQLMKWTNTDSLDAALAIYSCQGRENMPYEKAGVTYYF CDYTYVQVVDAGKTYYLDVCFKEYETQKNSIKTLDAGASA SDVERILQKTDLNYLDSISDSAYNNAMSNLDGQSYAFSSKK IVQRNITKLPTTSPHLFNVEPTVTEKLDDNRCDIIEIGFNNSR KKTYHASELYKKNITVEYVISDDTKETHEWVDFDASSIFNLP PYALGQALTVAPVIKIDGQAVLTGPAIDFESKQTLYINSKTG GKSEKFEEELCPGELCCIVFDVGTISPNELSEAYSKSIDQTT SANQKYQLNETTDASKVNEKNVYNANYLGSILRLTGVMYF SQLDIYTQTLAEKKSVNCEDTVKIGVFGFKPGVYPSKVQVA GEPYGIDKNGQFIVDILSNAVSTVSEVSNSEQLRAFNMERG YISSELESAILEQIVGVESLSTVQLFKRAQEKGINIVSLSKNT TKKVSDLKISDEDAKRLQAEIDAGNTIITMEQSITVGKWTGI GYIVETADQASQAFMISGKTNGGVCSSS
Rf 48	[(1-23)SIGN][(24-43)UNK][(44-108)DOC1][(109-125)LNK][(126-731)UNK][(732-756)TM][(757-762)UNK][(763-784)TM][**(785-1123)UNK**]	DIIETIANNYDYYMNGNEEAGTQVKINAAINVITFGVTKIGGAI ISQAKNATNCAKYGKNVITGLKNSGFTTAEVNAQISKFSKL GCSQTTIETLLKNPKSMFLGDDVLSFLGKQGGNQRILAELV LSNGDDFTKALMKTGAIDEFCDVIRKYGETANRDFLSVTTK DADLGKAIDTYKALDDIAGKYSEDAMHYDFVDGKYKSKYNL SADLVKLNKEAEEVMYPEIRTLTDDVATIAQNTGWSTEDITA IKNHLFNDIVLKNDGYGLLDSDYEIAVAWKRLIEGKFYDCDIL LLEHELFETTYYNYFHDVNSCTISEAHNFAEKYYNWRAMID ELMGF
Rf 49	[(1-25)SIGN][**(26-368)UNK**][(369-454)DOC1]	ADDDKKDFVVIGDSIAAGEIRDGFVEHNYGEILADYYGGTV ANYATSGMDSDVLLKSVKELSDEQKNAVKEAECVVISIGGN DIIHYFSKSMLTYFAEPNPANDFKNFLKAGYTEADIPEEPTID DLMKMVDADSVANFSKNMVNALELLGEIRGTASKLRNSSN GYIKTHIVKNLTDTIAEIKSINPDAEIVVQNIYQPVQLTPEYLA KTYGSGSKYIDIVGQVRDVAEGLTSSFDEQLAAVAADTGVK KADIRTDFTSMEDGVTQSDANPGHAAYFVDIATGSLSTGD VHPNQKGHLAIASKIITTLGDTHNDGGLLSDIYENLSDKASY PVAALKTYEAA
Rf 50	[(1-25)SIGN][**(26-277)UNK**][(278-343)DOC1]	LVPLSANAVYIPAASYDEVQKQLEGYTYITEFAVQESDPNY PSDSYDLYIRMPEEADTTCEVKLLMKENYDTVDVTFPEGG GRKAITEAMQKAGIDGKLLPGDNVNCECRLYDKVSADKVK KLYTILNDSGNITDFVYSTDLYTIRKGSANSESLSRFGFCTY DYDKGIWTPPDEQIAKRERFVELTKEILPDAAVELTSFSDKR EGFTAYEASVVPASSMTLKEWIDFSLEVNEKLGGFGGFGIA EEVEIP
Rf 51	[(1-24)SIGN][**(25-496)X160**][(497-570)DOC1]	VENPDLKALAESLGFESDHFSFGNFSRTEDILPVNYDYFDE YYNSLENLTKVRAEDAYKRSNETGAAFGLCTLEMLSHNGVI SPSDIFPEADKLSDIEYCDAVDKYISLYEAAQERFAIREYYR YLLYRTSTGEQADRLIEIAERSMAEGRYFLMISRLSMDDEN DNAPLMIVTVGSGIAEGEWEFDGKKYDKCVLTSDPNSRYT DSETIPNPRKPFEERLCVFINSETKDVTIPDYAKHGFNDLKI ASIDNDSLLNYNGFINPSTEIEGEDMSMYNCISDTTKYGMK YEGEIIDNNGVGEPFEFGPGAVIGKKFRIKSIRDFIPESYILK NVALPSDIYIIGVYGQFHLLSSKETVFDFDTEDNKAVFTSDG EFVFRGEYTFNEGYYKSNPYCDWGIEGSAVNQASIDFRDN GAVISSDSPIKDCLVSVQSFYFTDPDPKYDMLHFNSDNEVM ISLNEEEKLNLSIDPDND
Rf 52	[(1-28)SIGN][**(29-286)UNK**][(287-358)DOC1]	DGNEPPDYEALQENLEKMKENGFGTPATDSEICYKANSIEE VVTEHSGENYIALSGDIKDLHLQRYGENIYNMDDTGKLHIIS YFWEGSEIKIKKGTELPWDKLGRTYIQKTDDENYLLERGNE RNNVEKALATLRKCENVLAISNNFILYEDTKNSATIEGFYYK GTKSSDDIISANPDLKLNYVDLSDKDSKNYPLCDKYDSFFYI RSDNYKMTDIYGLFSEMKKNGDDFSCNISSTELVVLDKEAD VYSYCQDAV
Rf 53	[(1-25)SIGN][**(26-349)UNK**][(350-421)DOC1][(422-833)UNK]	EEPAETETVYTLGEKELTVADVITFSNMNRVLQWADFTDYS GEILHLKNKLWESEWQFAVADGEDGLMLRICGSTDKFPAHI YLSDSSKRQIDIRSDDVQSFLNKGAGEDSLTVSEEVKEHFR VYETYSDEFTKPERRLVQDFKSIMRYYNVPLRGAFTEFDNI DDVLASASVLKKYYIVEYEDGSVKSYDEHLNEMKSNIQTIH DGVQVDLPYLTIPEKAWNAFYNADFAKTYISPHAQVENVY WLSGESSMMGTAIYYRTSVGDYVYYYYHTIGEALFPVAEF CAYQQAIKDEIAKYPEAGGGINIADVWDLTDYRLKANIS
Rf 54	[(1-25)SIGN][(26-349)UNK][(350-421)DOC1][**(422-833)UNK**]	FKFHSVTDIRYNDDNHTKWTGFIARSENDLINILAENEGVSA DKATIEGIDSNTFKDKSIVIVYSICTAGNSYSIIDNISVKGTSID VSTISKKPMVPTPDMLFRRYVYVIDKNAVTNADSFNFTDES SYYQYDEENEAVAWFKKNGDIAGNNDGITNADALAVQKKL LGLDKTDNQSIDSSLIANKVFVYEKSADPGIYDDLCALSFGS NGMYTYHIGYYTSSNQDQGTWEISEDTLVLTGRYGTNKFR YEDKALIYIAEESDGFSKFTDKSTPKDGDKFNLAEEPDYAAI

		NNLSEIVTIKTDYTPIMSDWSGIGILLEFDSKDYSISLRTNDG HFTTWDIAKGSGPIKNAGVTYDIGNSGYIFWTPDGSEFDAD YQNEIVIIGEKDGKSVKLGSIIVTPSNNHTLTAALK
Rf 55	[(1-27)SIGN][**(28-282)UNK**][(283-357)DOC1]	ADGIIYDQHDAKWSEVKFDKYSTTASTMELSGCGIFSFCNA IYALNGTIADAYEVADWAVNIGGYRPGAGGTYRYPFYQQV EEAFGERYGFRVDGYYTGAVTDEVLIDHLKNGGVAVVNVP WHFMTITGYNEENETYHVLECAVDMPKRGLEADSWATAE TMSTGRTEVGWYVLLSDTGTSDRTIPASLDLNCDGMVNSI DASLITARYSEILNDTPLSNYGMTDDIRNTIDTCADIDHNGT VTQRDADILLRWI
Rf 56	[(1-24)SIGN][**(25-348)UNK**][(349-369)X159][(372-392)X159][(395-414)X159][(417-436)X159][(439-459)X159][(461-481)X159][(482-547)UNK][(548-614)DOC1]	AEAQPSDVSEIDSTASDNNNKLKKELTLEDVDNLSQKGYEL TVKDFDNYNYHDTGEDVLSGINREYIIDEEYSLVVIDKDSDV ETVPESIVLFSNKGFAADIRSSEYQDFFATPIMYLITEMPDA EYTAYINVILDNGSLYWCQIEKSYADTIVDPSEYTDDLAKSP RGFAKKDLKLNDDMLGSIKAVGKEIAKNSNQNELVECTKYT ITSERERLFLVNDGKSYAVADFGENCKIINNENVQKLIAELV ATGYSETVSAVDEKNGISGSYSENIKWTLDADGVLTLSGEG EIPDLTESAPWSSRRADIQTIIVEDGITSIGKD
Rf 57	[(1-25)SIGN][**(26-496)X160**][(497-580)DOC1]	AENPDLQALANSLGAETDHFNFTNFQRSEDFRPLNWDYFD NFYKTCTNWTAYSGESRYLTSGQTVALGMSILEVLSHNGVI RPSDILPNAQTLSEIDFCKEADKYITLYQMLQEHHEFNSAYR YKLFQWSPEQEVEDLCKIAENNMINNRYFLIFINCKSEENLP VMLASVGIGITDGEWEFNGKKYDKCILTLDPNGVAPDSSPE SPIPMPFKESVCIYVNSETNDFICPAYFEKNLSNFKIAAIDDD SILNYKGAINPSNEINEKYSTLNCVLKNGMEYEVEAQKSDG STQIIPDGYNIKGRYIKGENDNLKIKSTKIYEPEDKSEPLPSEI TYVDERGQISMASSKEAVYTIEKNKYTFEGDNNYSFWFFTA LNEGFYNYSPRYEWHIQGDVTDNICFEYLDDGILLRSDNEM KNIFFTSFDYKKGESGLPIKWQQNGKYLKFDSRNDVMISLD DNYELYLKIDPDND
Rf 58	[(1-20)SIGN][**(21-943)UNK**][(944-1024)DOC1]	QPGYVYDPGSITAFAEEQSDETAATVRAEDDSHEPETAYV ETTMNVIPPVTTTVPVTTGGIQICTTLVGNSKLTIEKMPDKIIY KKGEKLDLSGLVLKFSNDYTEYTYTDDDISDEFNISTEFDSS KPGRYIVHISDKYNPLFASFSVRVLAADRPDDYVSSVVFND TTGTLTLKGNVAVDDVIDFSENSAVKKIVAEKGTVFPADCS GMFRKFTVASIDLSNADTSNVTNMSEMFYDFTGCESIDLSG FDTSKVTDMSSMFEGCTASSIDVSGFDTSSVTDMSGMFSL CGNIESIATGRFDTSKVTDMSRIFNCCFNLKSVDISSFDTSN VTNMAYMFNCCGSLVSIDLSGLDTSKVTDMKHMFAWCECI AALDLSRFDTSKVTDMEDMFDGDDSLAVLDLSAFDTSNVT NFEGMFANCGSLRSIYVDKFDVLKAEKTEDMFYACQYLIG GNGTIWDPEKRDIKYARIDEKDAPGYFTKKTADTVITLPEVG TVTYDEETDTLTLKGNIVPAHVKRYRNAGTIIAEEGTVLPEN CDNLFNDSRAKKIDLSKADSSKVRTMKEMFFNCNNIKSLDL SGLYTSNVTDMNDMFGCCFALESVDLSGLDTSKVIDMSGM FRDTALKNIDLGKLNTSNVTDMSYMFDDCNKLETVVLKGLD TSKVKDMSCMFNLCSALRTVDLSGSDTSSVINMSGMFGW CESLEKLDLSGFDTSNAEDMTYMFTYCSELKTLDVSGFKT EKVKNMSGMFSGCELLQSLDLSSFATPDVVSMSKMFEGC SSLRSVDISKFNTSKVVDMSGMFDGCKNLEKLDLTGFDTS KVRKMDWMFNECSSLTELDLSSFDTSGVGNIEYMFDGDLF LETIWVNNFDLSKAENTTSMFGHCVNLTGGNGTKWDKTKN DSVYARIDEEGKPGFFTKKAASAEPATN
Rf 59	[(1-168)JUNK][(169-189)SIGN][**(190-728)UNK**][(729-794)DOC1]	ADGNLQYKRTDELPDYVKNFARCVEECDLDLDFNGDGKF DIFDDYAYYRCEYGSCPDYISGNVNACDEKMKAFSKETGIQ ARLSDLPEYFIFTQPMELEYFDPDYYIDNCPDTYDEIVPLEYI RHGDGKFGIKEYYDSMFYIQNEDGFVRIDYHTDNYAKVMS QVHYFIEYRLKGYMCFTNASYKMMCDLMDEKLVDTDINSD GEFNYDDIMLLYSFGDHFYDDYYEDIFGYWDENDDYHELD PAPIHAKKRTYAGYTEDYDKYIAFLEEFPWYNKDTDYQTNH KLTESEWNKATDFYDIASKYSLHCFKDIDQTLFIYYITHYNV DAEIFDESIYQNNKDRYLGDNLWGYMYTYKDICETYGVNA VRAPEDTPVDLRFARYDTEKMFPEYYSAVKNGELPEPDIN GNGRIDIQDYAFFQDLFNEIEIPGAPSYMIYVTIDAPQAVRD AFNNDYDFNGNGVSCDLGELECIELYIANELGAADYSDVG DMLKKYYEDNPDLDPMVKLDEIIRSMSGETEEKIRVNSDGE NMQGYMSSLE
Rf 60	[(1-28)SIGN][(29-99)UNK][(100-120)X159][(121-195)UNK][(196-216)X159][(219-239)X159][(242-262)X159][(265-285)X159][**(285-500)UNK**][(502-583)DOC1]	HGAFENTGCIETVDGIDYVDNWAVDGDSNSLKDAAIREGT RGVAEFAFLLCNKTEHLSFPDSIMYTLPLCYASSKGPAVTID FSGHSIGERAFTGAKKLTDIYIYDRECDIFDDEKTIPETFKEP TELDDDLIIDSGSSDNDKISGGNSHSKSIQGPSGSELVIDEE LPEEMPYTASPVITADEETKDNRVTIHGYIGSTAEAYAKKYN RKFQPID
Rf 61	[(1-23)SIGN][**(24-253)UNK**][(254-336)X72_dist][(337-411)DOC1]	DGTLTISGKGTMTNYDEMSIPFYPFQNNADIKKIIIGDSVTSI GSGAFTNTNKSFIIKGYSGSYAETYAKRKLITFEAIGTTQQIT TTATTTNKLTTTTTTEMPYSITTSYENRLEIVKLPDKTEFSIG EMYDLSGGLVRYKVINHCYDNGGYVVVKESENIEMISEDLD

		LTTDYNPNIKGKYTINLQYKKPYYYNEYSMPKASFCVNVVS SAVATIKTTTSASTAATTT
Rf 62	[(1-27)SIGN][(28-91)DOC1][**(92-482)UNK**]	SEAVNITSPACKAAAFACADDGELLYYDNINEHIAPASLTKL LTASVALHYLSPDTVVTVGSEQNLVRSGSSLCLIRPGHKLK LYDLLTGMLMASGNDAAYTVAVTTARAVKPDTAMTDAQAV SYFTELMNSYASSIGMRDSHFTTPEGWDDASQYTTVSDLL VLANHAFSIPEIKTITGTYQKKVYFVSGENITWTNTNALLDP NSAYYCADAIGIKTGTTASAGNCLIAAFERNGKTYLSVVVG CGTGNDRYELTLKMLSQFGVANEVKLSAAPNVTESVPSTS AEETTPVTTETTTVAPIVADKSEIFNRLDSLEYIPISCDGLPT HKLTAPDGTVYYLHLDENASYSYVWRRPSLIADADNEAPLT QEVIDAIYANWDQLNIVKTEW
Rf 63	[(1-23)SIGN][(24-198)UNK][(199-311)X142][(313-422)X142][**(423-879)UNK**][(880-952)DOC1]	IYSDDYVFDDLNAYVDLSKIKFRFADQEDMSEPTNVYSENL KLSVLDYAEVVNDKWLASDTNILQIEVCYPDGKKEIYKAED YDYYRKESGNFKEYNRILSNWNYSGHMEVFGEEAFNDEIL WLFSSCYDIDRIPEYTDGSRLVVYYLGPWRQNIDVDELKFP DNLKELYLEDSAFENVNINKLILPGCKLFYDDNTFAESNIKE VIFEGDVSLKHKTFRDNPNLENVTFNGNANLELTAFWNCN SLKNINISTDASINGIAFDHCNNFMTINGESPLGDDGSIKPEY ESFLRKNFNSAEGVGFIEKYVQASIKKAVSEAVTDGMSDM EKVKALHDKLCSMTRYDHGNTDDPKNHVDVSVFLNDSTVC EGYARAMNLMLHEAGVESCYVDTDTHAWVIVKLGDHYFH VDPTWDDNDEDITTYNWFMKADSEIKDDPSHSNWKMRCP SIMHNFQWEKM
Rf 64	[(1-21)SIGN][(**22-546)UNK**][(547-618)DOC1][(619-881)UNK]	MPVPAVYSSYTDEPLKKGDMRITLVDYYTGNPIEFDGNAEP YLWSDITYFTTHGKVSSGPIFYMHENSMIWENMADYFNAD SFEFGLNWDGLPKGYSIPDESVDQAGYFNGKSVPDNFVTV TKYDNGSADVEFRLINKNKSPAPKEAYESVIGTLPDWTPMD FADAMHFYNEHGKCYLKDNFICMVKPIHKSEIDKYGTRVSG SMTNVNTPAGTARKIYELEIKEKPDPSDEKSVHEYQDYLSR LEIVPRDYSLFEEYAQEEDPYVFEFQMFRVIEGYDLTIESYE KEGDEIKVLNTYTFENPDGDTIETDINKWLPDCKSECRFFA SPEIYGNYIAYHSTEKYYPGTALTVEQKGEGAVEKAYESEC SSFSLVPSDGDEPEYVRVYKPVADGRLNISFTVGKDGEEP FDQAKFDCEIKNNCSEIINYKGHTVFTFIDKDTGELITEPKSG ENFFFIGNYFREPLSGHIFNITSNPCAIKTFHTYNKNDNYTF NMKTASGRYDMPEFEVTYESSDCMDITCKLKW
Rf 65	[(1-21)SIGN][(22-546)UNK][(547-618)DOC1][**(619-881)UNK**]	NSYGYPLVVEKDNLKMYLGPDENYPCVAELPKNTELYEIGY NTDNDNWLFTEYKGKSGWVRTISENGEWNVRFLEMADKP VIYLYPEKETDVHVELELTESELSTTYPKYNNGWDVTAYPD GSLLNTADGTHHKYLFWDSKNCRTRFDLSQGFCVAGSDT EAFFKEKLSYMRMTEEEMNEFIVYWLPRMEHNRYNLISFQ GDVYTKSAKLNITPQPDSLCRIFMAYVPLENAVEIEPQQLDT FERKGFAVLEWGGCEIKAGEK
Rf 66	[(1-28)SIGN][**(29-498)UNK][(**499-562)DOC1][(563-713)UNK]	AGAYDIEKDIWKFRNTSDSFGQNYKFTDSDVAPFKNNLTN VDRRLASNLISGQFLGSCYGMAVTSILASYGLIDYNAYTEG ADSLYAMSGVASPDDMPSDEIQSLINYYSSLQFTEEVRQYA AYSMLEKTEKERLQQIIDAVEAGKPALVCYFGKINDSGDRY GHAVVAYNVEYGPFEVNYTEEVEQLALLKTAEFDGRIAVYN CNLDSQESSYIYFNNDGSWRTDKCSSDNEGNINLVISDIDL LNNKGLLGGTEKYKNDRDFLAMLNIRSLLPERTVDKIKFDN GKWDTISSNEDEIIELPPFLGDVPQIYANTYVLKDDTSGYML STDGLCTMNLEMNYQNCVLIADVNAANQIVFEPSGYIEMN GDTEYGFEIAMNDGYYNGSWYSFRIDQDHKTKHSALMRT DKGYILKSDNLNNTIITTCDGENVFRREIDTEFDIDSVLIFEKT DGNIGFAADLDNNGSYETE
Rf 67	[(1-29)SIGN][**(30-502)UNK**][(503-568)DOC1]	ESIETPAVMTNEESPVITTTSGPVEPVKTTVTESTTAVTTIES VKRECSYSFNIKFTDEDDEPIENINAKLVQHQIEWTDDEHY VCVGDAKTVAEWNTSEVNPFISEVLTANFTEYNYTVVTDEL PDGYVYYSDNKVEQGISGYLDGEVNITIKLKKGEITENTTPL SGTYSLKINVMDIVRNIPVEGLDCELFNIQTGDVAAKWNTS ETAEMYIENLEYSFDKPDSYNGNITYAIRITNLPENYRFFYG KTREQYGVSGFSLEEFKNGTDISCTVYLEDTSDDAPKYTYV TTPQAGTSLTETTTTTAEIGTSTTPIVTENNENNEDLDIIFKD NDLNIDNGTEKAFEYEITGTNDISFLSDSTGITVSNTFENGK GTLTVKAKGALEGKNFIKCFLGTGENSITKTITITVNKLTTFH CPECGRDVPIEDKVSGALRAVCKDCYEKGQSVGSTVPLET SETVTTTSTTVTST
Rf 68	[(1-27)SIGN][**(28-751)UNK**][(752-815)DOC1]	LSSDLFYNFEEDEHDIYIPDFPQEVLDAFDIDWTGLGFNRH VEPTEFTIDGAHHKFTYDGGYAMYNYQNELRTMCVLVNW YDCNSFSASQTKDYPTPEFFSYNHPKLSCRLNCDITKEGEL TAGTLLTLNGGKDELFIIEKHTDSTAFPEKEKIGSYSSEYSE YDLYKEASGDDSVSRYYAVNSASYGPFTEYCVDINDHLSA LSGNGVKVDTLDKYTSLVCGKEGRGDIFFESYIKTDPTPLP SEKLVTDEEGHPYTYHHNIVKNLDGTYYSFRTSDSSAPITE LENGRYEITPHEDNSWFTPHGNGGFTTQVTDGMTSTVLAG KEYDDGTSLTDHNFRLNYEYSISDPDIIPSAYVWLTDPSRR

		YIFSDRKIPDYDSLNQYLGTISLPEGDFDLYGTGDSNLEISP DGERPTFEYLFTRHTDDTAKPEPDAVVKGSYPIYALIQAAV KFGVDSGNVKRITFGTACYSKGYKLDVIKNEIAEDAVSVLD QYDPANSYADVPLFDTVNLPPYSYSMSGNQDSCTMTARD NGCFTGKAMGKKSSFIANYNNKVTYDPYHDVIMKYKADIK MNGEHNLVYSLYIPDKKDYNVTRNVHIIEENNELPISERTYN FSGMGNTLTDDDFELIKTYEANGHKYDLYYNYSKYYGCFN TNITESYICVRKDQPKGTETEGTINLSEHLKQIDALAEKSIDL NLIELTITGTDGTAELLMNEVENPVRSEPQ
Rf 69	[(1-24)SIGN][(**25-675)UNK**][(676-695)LNK][(696-766)DOC1][(767-782)UNK]	ADTASAPAVTEEAAEKNADKEGAAVEAMKNMNTIFSLTDG SDSDLEDYSKVGDERFTDIESVKKFIAETCTGSLKDEFIEKC EKSLVEKEDGLYKRNSGRFFFTFLTDEGVEIVDPAMDAFTA VTKKRDEMNDYGQAVFKADGSTWKISSYSFKSSPAETDKN ADLKELAGNWIYEDAEGGYTVDICSKYNGRVTVNEDGTFT FKNAEGIVSNGKLTTAAETYSDGSSIPYLLFSTGDTRTDFG GYHNEGSDIITLGNGGMARLVRDKNFDQDLNDLAVQRIEN YLLIEKITSGGLESDTEEFFKKDDILYYKVTNKEYTSIAAVKK LINDNTTGDMNKTLLEYCDERFIEKDGVLYESYAGRGSVGT DTAYGVIITDKTDKSFNATTIALNGIKGSGHTRAIFAADGDT WKISGIDYDTYTYNKSIDDYEICAESRVASMIYCMRYLEKGA DTDAKEKIKIDGVEYAKDADQMYTIDELKSLVANACTDQPR KMLFANIEKRLVEKDGTVYRIADEGQVSPDFNLNDGMKILGI TDKGFKAATVGFSQRDGYGIFEFVKDGDRFVLSSYSYSRF NEERLFGGYVDTQSGGLNLREKPDIKSAIIDEIPQGTQLDIY MCDTNGWYKTEFKGNTGYVSAEFIKKIPDSDIPD
Rf 70	[(1-24)SIGN][**(25-348)UNK**][(349-369)X159][(372-392)X159][(395-414)X159][(417-436)X159][(439-459)X159][(461-481)X159][(484-504)X159][(505-547)UNK][(548-615)DOC1]	AEAQPSDVSEIDSTASDNNNELKKELTLEDVDNLSQKGYEL TVKDFDNYNYHDTGEDVLSGINREYIIDEEYSLVVIDKDSDV ETVPESIVLFSNKGFAADIRSSEYQDFFATPIMYLITEMPDA EYTAYINVILDNGSLYWCQIEKAYADTIVDPSEYTDDRAKSP RGFAKKDLKLNDDMLGSIKAVGKEIAENSNQNELVECTKYT ITSERERLFLVNDGKSYAVADFGEKCKMINNENVQKLIAELI ATGYSETVSAVDEKNGISGNYSENITWTLDTDGKLTLSGEG EIPDRDGITPWFDQMVNIKTIVVGEGITKIGKG
Rf 71	[(1-22)SIGN][(**23-399)UNK**][(400-420)LNK][(421-489)DOC1][(490-624)UNK]	APVTISAETVETAVTQTTTSAETVETAVTQTTTSTEAEGLAP VIHDIDTAREKFEKYIEETDIDANIAEEGKYPQYDGKIVIECE PDTDAYARIFKFADENMILHELFETVIKFDTSKTVNYTVSFR FVDEQTGENIENIHAKLYRYNNKLEPNEDEPGSYFITRDRD VEPVLIAEWNSTETPVFTSETITSYLAQYGYLVITDKLPEGY NFYGKDHAESGCSGAIGNGNHSMDIRISKGEPTPETVDIPL EGTFSLDVRVVDQNRNIPLKGMKCEVFEKYSGEVVAEWNT SDTEVMHIEGLEYKFEGNKILEDFGKKIYQFRITNLPENYIYY GYQEDSLTLCGVYIDEFAEGNELSATAYLIDQSPDAPEIKYT T
Rf 72	[(1-21)SIGN][**(22-244)UNK**][(245-322)DOC1][(323-557)UNK]	VGTFTHAEDAETVTENYKGMTLEEAVLEMTGGDYLWLAGK RPVLRPIRGFTGKISTVTADNKTHNLEAAKAITSIRMKAGKE LPYDDIKAAVEAEGLTMPVIQKNGKDYEIISSISEKAYNFTLD QIKACPDVTAIDMKYKLYEDRANKLSIYSLLFENVAEEEISEK YPELTKNDTDANFPTEGTKLYYTFDNDKLYEIISKLTENGTP FTPCWSQTELGMK
Rf 73	[(1-21)SIGN][(22-244)UNK][(245-322)DOC1][**(323-557)UNK**]	LGMGDIPIPPKDADIDLVSYDPIDFTEEHLFRSMTMNGIDYW VSYVPVKKDLIGEELGEFTATGHKTDSISETETVEEPVTVYE MKGYAVDAMVAFRFKDSNNYYVFRNPKYLPKTVGELIEGM QLDEQLIFNRVYIKKFKPYETADYDPEKAFIWDALFGDKDIEI MELNDTNRYNNITHVPRPDNDFQLSIGCDMPITGKYNFGFT VYSTGFVTTNLVDYGLCFKISPEKAQA
Rf 74	[(1-25)SIGN][(26-96)DOC1][**(97-490)UNK**]	KYENSGTVNLCGGIESAHPDGKKVDESFTDAQTEFSLELFK HTIKEDKNVLISPYSVSMALGMTANGAAGQTKADMEKALG GLEIGELNKYLKTWRTGQPNKDNLKIKTANSIWVRDNEALIK PQPTFIQDTVNYYGADIFKAAFDQSTVNDINNWINEHTDGM IPEILKELKPTDMMALINAVTFDAKWSTPYDDYQVKPHKFTA YDGSVKETDMLYGMESSYIQDEDAVGMIKYYEGGRYAFAA ILPDEDITVTDFVNNLTAERLHKMFTEQHYADVTTILPKFKY ESSSDLNDPLKDMGMESAFLGTADFSKMVTPDSIPLYISSV KHKTFIDLNESGTTAAAATIVLMAGNGIAPERERKEVILDRP FVYAIVDTETDLPLFMGTVMDIE
Rf 75	[(1-28)SIGN][(29-42)UNK][(43-113)DOC1][**(114-569)UNK**]	RLKVVDMMSGEPLESVDVELFGLCDDYCYDVGRWTYTPE DDTYFTGLPTDGRYTYMVNLDNLPRGYGNGWGNWDQQL FFSYDGVTDKEVTVRVLADKSERNVNMDMYNWSIGLHQT YYGTVCITDKDGEVYYPQLRNEEFALPDGEYHAEFNGFDY PVTLIDPESDFGKHMKEVYPDVEFTDKSNGFDFTVKDGKA DKDLSFDFGPLPGKSNYITVNCIDISTGKPLPGVELSVIECP DTYAKTIAKWTSDETGTVKFDGLTMTGSNAYKLQLDKVPE GYVGGFDEYYHWGYVYEYEGEANLYFSPVTEEKNVSADV LSIYDKSVMNDLCTYDVYKEAGENFKLDHIYTNVKPGEKIAL QDGDYFAVLDLRKLREKGYDGILLYTERGKAVAGDIKPDEY MMDTAMLKFTVKDGKPDRDLYFYIKEYDEHDYDPEEIEHD DEIKFINDFIEAKDEE

Rf 76	[(1-37)SIGN][**(38-647)UNK**][(648-753)DOC1]	VMKNQDGVYLNEALSSAGDGSFTESRDQWLMTLDKYSVT DFYIGTPFSDWAYAASPYGDKWQFNEGYGWIMDKYGGDE VNGSMNCTGFLWHILSKSISDYTGVSMSVASEAVPILNHFN EQGFSCSSWNGGGNRWYDFISENNVHYYEFNTKAEMLSS GVLQKGDIIWCVDSNCGTLMKGLKTPSDDHHVGIYMGDGG DDLWWQSGPIYGNGDKRDQQNSINPIYGCSVSNTYIVLPW SGSDGASTPAKPPHEPYMYNSDGWYMGEAINMASLNTCA ADLGKWQNFIDDYTSNDYYKGTPYIDWVFASSPKGDRWQ MNEWSVSEMLKVGGTYEEGGLNCMGFVWHAIAKGLSVES GLDISQTGQYVPFSSYFNGLGLSRKCWATPGGSGGWTVF VDYYNLHYYEFPTKEEMLSSGVLQKGDIIWCVDGSVGLGM AGLRTIADNHHIGIYTGNGTSDSWWQSGPVKADGDLVNVG TDVCPIYGAAAKNTYVVLPWAKKAEPVVTTTTSTTTSTTTT TTTTSTTTTTAATTTSTAASETTTVSETVTSNTTVTETETTQ PVETTTEVPVQTDPNVVLRSDAGEKIFSSGVKAEDIAESAKI VDDNGFVSEFK
Rf 77	[(1-26)SIGN][**(27-367)UNK**][(368-427)DOC1][(428-699)UNK]	ESDLISDPATILEIPEQDPAPLQVKLRTDSNIVRSDSDLKIYV DIKGDIPDDAWLTYVPSDVPHTEKDSDEFNGNWVYLKDIKN GEVTLKAPKAVGNYDIRVFDSDNPDDAKEITYLPIILEYSPDL EVTIDVDSRTVKAKAPLDIKVDIKGYVPDDAWITFVPTDVPH TEKDGDEHNGNWKALKDIENGIATINTPATVGAYDIRVYNG EIPELASEIACKTIYLSKAPDLSVVVSTDNTQVKPSEEMEIKV NITGTIPDDAWITFVPSDIPHTEKDGDEHNGSWARLKDIEG GSVALKAPSETGNYDIRVYNGEDSIAEEIACLPVLVTENPAI VKGST
Rf 78	[(1-26)SIGN][(27-367)UNK][(368-427)DOC1][**(428-699)UNK**]	DVTYVEPDERFEWGVPFYVVKDGLKLYSGPDESYDVIASIP ADTRLTESGVKKNNNTWLFTEYNGQCGWIDTLDKDGKMVI HFEEGTKKPVIYLYPEKETDVHVELELTESELNTTYPKYNN GWDVTAYPDGTLLNKADSTHHKYLFWDAVNCRTRFDFSK GFCVAGSDTEAFLKEKLTYMGLTEQEMNEFIVYWLPLMEH NAYNLISFQDEAYTNSAKLTITPTPDSECRIFMAYVPLENAV DIEPQQLSAFERKGFSVVEWGGAEVRSRQ
Rf 79	[(1-25)SIGN][(26-100)UNK][(101-121)X159][(124-143)X159][(146-166)X159][(167-343)UNK][(344-389)LNK][**(390-616)UNK**][(617-686)DOC1][(687-701)UNK]	PLITWTATPEYIFVEVASYPTKTAYSDGEKLDFSGLNITVYS RTWHESYDFGEMVYGSDFDMGYEPSKDIVTVTSSDGRKY EGSEFPKLPGGNYKIKIQGNDYDTHPSIHEVNIEYDVTKQG ADPRFLGEWELYKLIDANGSEKYNDGSMYIKLTFNDDKTGT AVGYMGSGREEEAFTWYADGNGVTVKDNSGGSFRLAYN NGEIEAYVNGGTMTAYLKKSQASNDE
Rf 80	[(1-30)SIGN][**(31-586)UNK**][(587-654)DOC1]	AVVADSRSDEAETNGDQAKNGRYLTGTVTYSAPDKVIYKV GEELDVTGGICEGSTSVMAYNGLSSYIEDGGDIIDPDRPYTI DELDLSGFDNTKPGVYRIVPKVRCPEGVDENKIDYYGFSV MVVDGNTSEDNYVTAVGEKSFEFSYSYGLAVKTGDKLDSL GFKCMEKWHYEWHHDGEVEYLFGADNERYNSEGFIDTST VDTNKPGVYKVTASKISQPNLDRDEIKYSSGYIMVTENKPT DADEDEIWAINKNGGPITTTDVSDDPVTTTTVTTVSTGSTTS IIYTADQAFECLSVNTYPTQTEFMEGEEINTNGLNLRIRRYIP YEELGYDRSEMTIEPEFNSFKTFSPVNIDIVDKDNNVYNGT QFSQLPPGKYVAVILGDYRVWHDVQLIKNVDIAYEVEIKAD DRGLLPEKTELSDKPLWPDMSKYYQTFKGVAAYPDKKVYT KGEELDLAGLIIKAQRRYGEKPSSREIDKYAENPVYLKTVLP ETDIMKITDESGRIYNSNEFSMLPVGKYTVSNRKSSAEWG GYDSIVSDISLSYDIEIRNKNDIGE
Rf 81	[(1-25)SIGN][**(26-757)UNK**][(758-780)LNK][(781-850)DOC1_dist][(851-883)LNK][(884-999)UNK]	IAPELALTVNAADAAETFYTSLPIVSHKNVGIYKDDKSATFK MSGRIYNNGIVFREHTNTTSEITFDVSSAKSLSFTLGHIENA NKYDSGIKVYVDDKLTDTIELKWTMNAYKYDMDVSKAKQV SFIMENGYAANYALGDITTDTDKPAIPSTRTIYKDMSSVIGG AYDGDKIKLYPGTDETKSFNMNGRTYHEGITFSHGYKGDV ATIGFNVENCSKLNFTLGNIDGTESADGVFNFYIDGKLVDTK KVSYGEPLKDCSIAIPKGSSYLRMEFAGEGNCAYGAGDIQL DDLAVSKKAAVPEFKDSKSLLESAYDLNKATIYYGDEKGKS FNVNGRSYYQGINFDTSSSNPTGAVSFNVENHDKISFSVG RQDSQNAQGGSLGVFIDNKEIEKIPLSPDMLTTDYEFDVKD AKNIRFVCSAYASHYAMMDVKVDDLSAGLDSTVAETKDTA SLIKSAFNYEKDAFTIYSGESDKEAFNMNGRTYHDGFIVVG DYNPIVHNISLNVEDFEKITWDTGSLDSWENDHEGYVNVYL DNELKEKIDLTLNIPITETSLDVSKGKVLRFEFHLNSAYLKYG IANIRADKLAPVNAPSIPEYKDENEFIKSGFKAVNVTKYTGG SDEANSFTVGGKKYYSGFVFPRTGSVAPCTVSFNTENVDG MKFSIGVANRVYDSDSVTLNIFKDNVLYKTFSIKGNSEPFPF GLDTKDCKVVRFSVNSSPAVNPAIYDMELGEF
Rf 82	[(1-28)SIGN][**(29-254)UNK**][(255-326)DOC1][(327-683)UNK]	ANSLNSNDIAEVKITSSDLKECNLVTESKNDGCWINFDLVS GAFAMSGNNHQSFSIIGSFERKDNDLYLYPENGSAEFYILH REEDHFVSQSDEIGVQLKAGLVFYADNDAFWEILLDWTLS GNEPSNTGETAEQNTSFIMGDVNDDDTFEDAECAMTVTVS EINGETLLVKSSDGKGELLTLSTKYLDSSIQPKVGMKLEVVY TGGILLTYPCQFGNVKKVSVVS
Rf 83	[(1-28)SIGN][(29-254)UNK][(255-	APPEPSTYGFYSYDKYAKYVAENNLQDKIVTYEQISQFGDF

	326)DOC1][**(327-683)UNK**]	VQCFINCDWEYNHYFSLVYVLDDGSGKTFDLVIKALNCDKT LVNTDFDRNAQDNHELVSLTQEEMNLTNLRTVETNAEYAY FEFDDKQYNYYNGRLSQIVWYDNEHEYMLIGNPQLSDYPD VDNTYLAKLLKVSAVNPCTPPTQPINYQNFDSLVKALNSND LTSYPEEDRETYHQMFERFQNDGFIYQVTDNDLIKTNQER GITLFPSASYEDVGIGCYVTFKGNNYHIMFYSANADVLAET DGIAAYLKKRMGRSSDKEITVSDKTVSLLFHENGQCYANAF VDENHYFDVIGAVSEEEMTEFLNAFAYEKIVF
Rf 84	[(1-212)JUNK][(213-232)SIGN][**(233-957)UNK**][(958-1024)DOC1]	VTNDSNANDIKVKESSPFIYSYYYNEGEMFEALPKLHENES VLDVNKDGIFDIEDVYIMYSSWAYENFKTPEKYDFSGLRKT EIDDEGNEVNVPLYDLIDVYRYFATYRTVKPEHISKEPFRQF CMENLDLSSYRDEEEAVDAICKDFLDSFTTSISFTYQFYDM FRDKIANNEIDIDLNSDGVIDFGDYCCYNDFQIPYTDPFRLN KYLDSPDKISSNNPDADPETIEKCVQLIKALKFEDRYPMLSD TGIYILQYYMEEHPYNPEWSDDLYYQYLDYYVDHPYSLDYI LQPVQYNLGYGFHNMRYCYPDTSAENIAAEYETFKKSADS GKRKVPDLNSDGLVDVGDFRLAFLFFDDYRYMPGIPFPEE YRQSFLTDFDLNDNGMNGDINDLAVYQMYISDLIGGGEDE FRDELIKYYREHPGFDAKHAAYYLEDSVPDEYQGNIYDTYN KYLSDIKSGKKEKPDINMDGKVDMEDFIEGWLTMLSYRNN YGGWRMSLVSEETQNRFLESFDPDGDGNPATWDDENLL QLYVGDILGYDIMHVLDSEVLIDDAEYYISRYDDILPEDERTI YWTLDPETTKNMTRLQKEAILNGDIDENDKIKRNLSEIISSLS DEKKDALDINRNGQIDNEDFHLAHALKNLYFRDIPEDEALT GEIKEFFFSRFDFNDNGLYGDFADYLIADKYYESNMVNILP DEPAPLPDGHSSRQDIIDRANKLRESLA
Rf 85	[(1-27)SIGN][**(28-542)UNK**][(543-615)DOC1]	SQHEGGIPRSPEITEEYLQDKIPVQIYVWECCAEKEEDGTL FVGRSPYFKFRADVSITITSPDGEDVIDTIELGEFYTGTEQS LLFGIPTSIRDKYNDSSRYRIFVNIDPVDLPEDKVMVETITMT ATDSVSFEYDNQNELKGDIDITIAEKISVKQKGVIHLKAPEKR EYHIGEEFDITGGKISGFGEIFNENEQTSEDWRIEERELTID DLDISNFDNTKAGEYLITAKPVDLNTTDCSVFSDMIVYDSFY VTVVDDSAPETTEPSHEELKKGEQFRLTFNKFVSNDYDEE LQNNVLRTEELKGMEFDLEFDVFRYSDDGSERVDHMDMG HFTMGDASSITITIPDEILKKYSDTDIYQCDFTLTPTNLPDDK VLSDFSCLNSSGYSYYAFSPSFGETVFDILVRDKKTCISEG SISLEAPEKVIYRIGEELDLTGSKISGCGGCTLNGDTVLKWD NFAHQPSIDELDVSGFDNTKAGEYTIRPLKMNTTIPCDETV DKISYGSFTVTVVDDE
Rf 86	[(1-25)SIGN][(26-36)UNK][(37-120)DOC1][**(121-478)UNK**]	SNFKANVWDIYINTEENVKFTVDVTSDEKLAERAVALYDDS DNLVAYMNDDGKNGDEKANDGVYSVEKLLSADAPVNIDYY AAADNVKSDDYRIVFYRDLTKEDFEGFDAVSDDLVGLSFDE VCDYLEHSDKIESYSIDNDAGTIKYRTVYCISGLWEKPSSDS TKTWGGNGLTVPFNNEEKFDQAYKNAESAFTNGQLMPRD LNGKGVLLLQPAASTMGDEFGKEAANCLAKVLDCNVTTFQ DDKVDVELMTNLDGYDIVLVESHGIVHFGTEYIETGEKYNY FIDGIIMLESFSSAPFSGEIIVPSKEDKNSTEYRISINAKKFFD LVYGFQDEPFKDSLWFFSSCHGMEKGKMAQ
Rf 87	[(1-30)SIGN][(39-109)DOC1][**(110-549)UNK**]	NINVADMMTGELLEGADINLSGVIGQSSFPLGSFHFDPEDTI SIHGLPTSDKYKYVLEIKGLPDNYGNRFGGWDRSMKIGFD QDSDTKDLTVRLLSDDAELNIDAGCFDWTKGDSESSHGMF TVTSKDGEIFYQNIRANDFALPDGEYHIDMRPETQAPLNLL DPDSDFAKYIAEIYPDVSFTDKSDGIDITVKDGKPDKDVFFD FGPINNFANKLEINCFDADTYELVEGAEMTIIEAPDTYAKKIA DVVSNADDTIIIDSLYRAGENAYKVILNNIPEGYIGPDEVTINT GYLTNALSEVTIPLIPQQKAEDVTVKVHNISDNKEVDGIGIKI YDMDNKLFADVKSGEPFVLTDGIYTAEINAEDADSKHFKAL SLIRNVNTADPKYTWWTENHGLIMFKVSGGIPDTSIDLYIAD ADTSDEELESFVKELYPEANKD
Rf 88	[(1-25)SIGN][**(26-497)X160**][(498-574)DOC1]	TDIIDIKEIANSLGADTDYLNIPNFVYGEDNLRPYPMEVYEEF ISKCSNYEASIYNFEEYKYNLSTGSCLGLSLIEILSHNGIISPS DIKPGAKSLSEISYSSDLDRFILGYSATQLFYLNEYYLCSDIP NFDQRTQINELIRVAERNMAQNKYFLISYYGTTAHAIVGIGIA DGYWKYNEKDFDKCILTLDSNCVKDNNHASPFNENSCIYIN SESMDYYIPKYNYGSSSDNYKTVIVASDDDDILNYRGVINPT KQLKCDFSGTSDFVIPKNKFKNYVFTIKDKNGNIIDVSKNGK KFDCPTQDVFFVNSRDFHIEIDNKSIIGEAQNDSFEMSNIRY KFNGGTQKGHGIFDVNDLGCSVTAKNEKDMIYYVNIRYNE GSYPYTPHFNWKFNGLTNKNFKATITANGMHLSSDSIIETVI STADVKLNEDGTIADVNDNTNEVRITAVNDVLVSFGDNKKL CFFIDPDGD
Rf 89	[(1-23)SIGN][(24-311)X128][(312-340)LNK][(341-405)DOC1][**(406-819)UNK**]	SPAIEDGPQKQAEFITANLAKHGASLPTQGDAKLVVFYVDF PDCRYDYEPTTEQLNRITFGEADEKNVCYPFESISAFYGRA SKGSMNFSGQVFRYTTKEKQSAYDTDKVKIAEECYEAFKD KVDFSQFDGNGDGRIDATLFTTPAKAGDTNWWPCSGAFG DPKYRVDGVGVGHIITGNAQVESTENYVNYISTYCHELGHC TGLPDYYLFTTNDSEGMHGTAGAELMDTDAGSDFGAFSKL

		VEGWYTKDQVQLWYPDQGTKTFTLSNAQTNAGNCLIIPNG KLADDYFSEYFIVEYATKDGNNSGIGKNTAWWVKSGEGVR IYHIDATTEYGWNNYFRYASGSEFTNKDKGRRLIRIIDDREI DNLYHTGDIINGNISGFHWYDANGGQTVDTGFSIEIGENKN GTYSVTVKK
Rf 90	[(1-20)SIGN][(21-33)UNK][(34-118)DOC1][**(119-894)UNK**]	GRFKANVWDIYINEKETVTFTIDVTSDEKLAEKALALYDDSD TLVTYMNDDGKNGDVKADDGVYSAELTLSSAETKYTDYYA AVGDNKSNTRRVSFYRLLTEDDFKGFAELNEAMKDLSFEE VCEYVSNSDKIKTYEIYEKEQSVCFQTIYGFHSCWQPPIDE NAEYPTCGTGQYAIPDSTIQGMIGQDFMYRDVPFQTSFTD MASRLIDEYEFTPANHKHHNVAVLKQTKDPRFGDYCENLG DCIVQATNQNSPMFSSKEEIREKQLKIIDPAECNTLEELKHL DGYGTVIMDAHGAFYNKMPYITTGIKIPDSAVLAYMNEVLIY LANNKYEMSEFSKDIFSGNISSNIIFSEYMVGPGFFDKYYED NSLEDSLWLLGDCHSMQSDDLANVIRSKGATAVIGFSNSV SFLYRDEMFFEILINSMLFSADSLKGGIAEAEELFGKYDPVN DTEDSECYLRMIGDDYHYVTDVKARTTEGENVPDDDEPDD TPGGDTPVTPPTERFPDEIISTPGPIGISFSPLDAGTSSDDT TLKILEKLYNGLDDGNKATYTSYSGCTMFPFTPKYTKDGEE YGNYSSLGEILNSAKSIDDSYSLYISGLKPKTDIRHLYCKRD DFDTVTGKLIYPYGTTPAKKIYEHATTENETFNMYCAMNGG SYDLFQGKNQTKITIVKDDYYFYKWSTKGPVETYSTRMWY ENYGYNHKKNYFTSIHYIIDLRPTIDQRLNELCFESCGKYIQ YSEENLDKLITLINERRAYIDEYFEEDYHPDIEDFAG
Rf 91	[(1-22)SIGN][**(23-873)UNK**][(874-933)DOC1][(934-984)UNK]	FAEDTTNAVTAESRVTDGLAVYERELEHMKSVRSISKADTI NRWNAYIKEFNDRYLKELDADKQLKEYFIKRWKELSEKEKA EKAAEAAANDPTLTTTAQTTAPVATTTVVSTEYDDYETRDP DSISTEFVRQETIDAILKVMPPEKVDEMINNGFNLYYINGSVI VISDMPFQAGDLDESAQFQIGLFYTDPLKVDWNYDKIDNFT VTGKLQNVVLIEDERNRELAQIEYEMTEMGYWNVKEEEVAI PYDPQKGPTEIYKLPPDPSLEDTKKWLDSLPHLFYDLTSGS DTYHVYDVGDEPIYAIISERGVGFSGEDDTLFLDYDTAPDD LEFIRMGKDLFINDPEHEIYLICSKYFSEPSKRIEHITFKDGT DMSYADVCYITNSLVGTENDDTITGYPETNYIWGLGGNDYI CGNKTDDYLFGGDGDDTLTLPDGCFGLLSHNSGNNYAYG MDGNDTIRLGCGNDFIWGGKGDDIIKSGSGEDIIYYELGDG NDYIDDTTGKGSYPSSGKDVLWLGEGILPDQVQVSFSDKY YEYNLHIMKTGDTITIPGNMYSGITPVFPIEEIHFADGTTWG RLDLLERTRYLYGTDGDDTLTAVVDVDAEFKKEYVPDAILK GFGGNDVLTGAKGNDKIYGGKGDDVMRGGNGDDTFYYEL GDGNDLIDMGTGKSSYPQGGYNVVVLGEGINPDEVTIERS ADEYSYTLWFDKTGESLGMTGNVVSGLTNLFQIKEIQFADG TVWDHDYLDSHYVTWIKGTDGDDSLMDSTADDIVYCGKG NDYIRGRGGDDLYIYELGDGCDTIIDSTVWGNGYNTLKFGE GIVMENIYTETSVYEGTKVDRFYIGSKNSYVEVS
Rf 92	[(1-29)SIGN][**(30-244)UNK**][(245-316)DOC1]	VHINAENTAKTLTKDELINDMSFLYDEEDSVYYLGVGLRNIY FKVPTECNGRKIGKLDLGHVYFAEKYIPSDTDRVIIIVPDGVE VVNKHWNCNDTGIPFIELVYASGEVEDIKADDYEKVAEYLR SYYRSYHNEITDEELARKLIGYSINERNNIPYPITDKPAYEVK NETEYIVYTGEDGLTYLKVFIGTNGEKLIIPEEYNGVKIDRLN LKD
Rf 93	[(1-30)SIGN][(36-100)DOC1][(101-122)LNK][**(123-583)UNK**]	CCAATYVSPVTFDIEAYKQTHRATASAKPYRPSAQTYGPFL TDEDFITPPIKDMEGSLPTQGEAKIVMIYVDFPNCHYQWAP LSDTMYNITFGEEDPTNPNYPNESIRAFYQRSSKHSLDISG QVYRYTTEHDKEYYENDTHKRLFVNEVLDAMDDIIDYSQFD GNKDNVIDTVLISVPASAGDAEWWPCASNYYYDFDHLLDG TYLSYVIIGNDEIKSASDYEDFTTSYIHELGHCMGLPDYYLY DNRDSEGLHGSAGFDTMDELFSDFSCASKLMLGWYTGEQ IQVYDGSDEQTFTLTNGQSDTGNCLIIPCGELDDNYRSEFLI VEYTTLGGNNSKLREHYWWRNTGSGVRVLHVEATDEYTP SGSNFLYRSGNDSATNFDLGKRFIRIVNDSNYDNLFRTGNT IDHSSYGFGFYDDFGLEKIDPGVEISVGDLTDDTYTVTISKK GLINGTNDIQYN
Rf 94	[(1-20)SIGN][(21-49)UNK][(50-140)X142][(141-149)UNK][(150-240)X142][(241-313)UNK][(314-400)X142][(405-500)X142][**(501-881)UNK**][(882-956)DOC1]	DYDELVGYTKNEMMKNNYSSDIIEKVYPMYMRSHLTEHPD HKNTCYPLENIDSQYSSGVKYDIEVNKEYADNYDITITGEQL DEDPSRGIFEDKNIRDLNINISADEMFQDLFNPFCRVNNITIN SENSGKVIVHQIGASIVKEFKELTIGENVELGEGAFAWCEEL ADLNVDITKDICGQAFSDCPNLMKINNESPFNDDGSPKPGY KEFIEKNFYDADDIGFLNKYTAFLVKQTVNETITDDMPDIVKL KAIHDKVCSMVYYDLDNMPAQKNHVDQSVFLSDSTVCEGY ARAMNLMLHEAGIESCYVEYNSEYDDGHAWVIAKIGDHYF HVDTTWDDGDTVSYDWFMRSDSQIVDKDYHTKWEMRCP SSLHSFQWNELP
Rf 95	[(1-179)JUNK][(180-198)SIGN][**(199-695)UNK][(**696-763)DOC1]	VLPSNAETNPCAFLDFLRNDSAFYLTINKSAVNNVADKEAD LDLNEDGSFDMWDLYAFYRGMNDHNNHGEVIVHDDRTET FIPPVFTAPENITENVVKYGDLNGDKEIDRDDFEILAYYYSIN YNSTLTYDSVDPNNYYFNCPDDYDDTQYSYHFTTPDDKW

		TFTRFYDLDIYRNPDPIIQFVFDFVDCNSDAYGGYSIFCDMI DKGLIDLDVNGDGQFTIDELYDHIEAAHLCYWESLPEYGVD DSEEYETARNTYFTEEEWEKLRSNCNYARFSLNYAFEYDR YFVNYFFNHNEFKRAYGEDHYFDNMRGGFFSISAHNHMF DYMKYAMPGVYSARFDFTQEETKNDFISYFTRVKNGELPE PDINMNGTIEFEDYIYADLMLMKDSFKNDPKYPEFDQAIIDN FNKNCDFNGNEISGDLSDVISIQLYVVKELGIPEDGIKDEMA RYFKKHPEIDLFDYAHYVLPEEEELPEFVDKDMAKTGLTSIR VYMSNID
Rf 96	[(1-29)SIGN][(30-110)DOC1][**(111-549)UNK**][(550-880)pfam08757][(881-952)UNK]	YLNGTEPGTTGRQTTTAMPETTSETTTVTTAAPQPEITVEA NIKLGGGSVTSDSEYAKAQGSTLTITHSGTYNISGKLDNGQI CVDIPDENADPGTVKLIFSGVDISGKNAPAILVKNADKTSITV ADGTENTITDGDTAYSGDFLDNAVIEAKDDLTIKGGDAGTG KLTITANTQPAVVCNNDLKFTGGDITIQTLNAADKTDAVKGK TSVTVKGGRLTVDAEGDGIKSSKGSLAIEGGEVTVKSGKDA VQAETDLTVSGGKVIACGDRGLTCPGTIGISGCELFATATD NQCETLAATDAPALILNFTKEWAKNNPVAIVDGSGQTVFDV NNLKKFRYAIVASDSLNTDTQYKVFAGGIRVNHAGGDTFKA GFGAGNMTYNDVNNTDDAEVLYGKLFDQSMVHSVDVKMS ESDWQTFLAHADEEAYYPCDVVIDGEE
Rf 97	[(1-22)SIGN][(23-40)UNK][(41-318)GH43_C][(319-341)UNK][**(342-550)X19**][(560-724)CBM22][(728-795)DOC1][(796-813)UNK][(814-1075)CE1]	DFSYSSNDLKLEWQWNHNPDNKSWSVTERDGWLRLHNN TKATNLLNARNTLTMRTEGPACTSYIKLDTKGMKVGDYAGL SAFQFNYGNIGVYVNDSGQKKIYMARNGGSDIATSSNKIIA ETNMSGDEVYLKIDFKFNDVKSDMSSSNNIDKATFYYSTNG SDWKQLGEQLGMTYDLKLFTGYRSGIYSYATKNTGGYADI DFFEYSKA
Rf 98	[(1-38)SIGN][**(39-318)UNK**][(319-725)PL1_2][(726-773)X149][(788-931)X157][(932-995)DOC1]	ADTDSRIRVDINKNDGRKASYSKNANNWILEEGTAPTYKVG NVTFKLSNGGSAGGNVTGANNKKLQLQSGIYPQLTMDGAK IKDGDNGGVLKLEISGLSESEHSLQMWHCNTDGYTNSKLSI YVNGKKVLTGVNCPTNVTNENDAGISYVTWTGSSVTILISP EGGGKMDVAWLNGFELDGSDPFNGVSKMTPADKEDHLD RSQGLSWTAGKNALSHDVYIGTSYDAVFNANHNSAEFKGN QTATKYTIDDSYSSIPTYYWRVDEVSANGTVKGAVYSFM
Rf 99	[(1-28)SIGN][**(29-99)UNK][(100-120)X159][(121-195)UNK][(196-216)X159][(219-239)X159][(242-262)X159][(265-285)X159][(286-501)UNK**][(502-583)DOC1]	EFTYSADYIAYKDETLPDLEADHREFVKDGLVFNIYDDFAYL VSCEDTDITDAVIPEEADGVPVVGLTDTPFGYCRSLKSVTL PDTMKYIDWLDLAASSGKVSTDKTDGEILPTLEKVTVSENN PYFTSENGIIYSKDMKELIGCPPAMEMKELKISEKAEAIKDF AFAACYKLEKAVIPENIKHIHNSAFVACKNLKSVEIPSGVTTI SGDAFFGCSSLSEVKINSKLEKIGFGAFSGCTALKEFNIPET VSVIGH
Rf 100	[(1-20)SIGN][**(21-751)X141**][(752-888)CBM6][(889-952)DOC1]	AVPCDLSAVVSAEGDGTAFYVSPDGSDTNDGSLAHPFATL TAARDAVRKINGNMSSDITVYLRGGDYRITEPIVFDTRDSAT NGCHINYTAYEDEIPVINGAQQVTGWTKFNDKLYSATLDRD YKLRNLYVNDKRANMGSVTVGSKGGWGEYKVTAGQADW AWDSGTAKDGISYNAGDIPRIPSNFDDLEIINGTTWNENIVC TRDIKVDGNSLIMLLQQPYGAIAQTPGWGAGFNTGGTHTIY NSLSFVDSPGEFYFDKTDKKLYYYPRNGEDMSSADVEAPV AEQLIVVEGKDTSDRVENISFSGITFANTEYQLTNVAGSHG KTTCQAAQTYTAYADSNWHKRKYEMADTLPAAVHITNSKD ISVTGCVIKHTGADGLSMCNDVIDSEIKGNYITDITSSGITIGH PQHIYIGDASWDNHEKFPKGVEGICKNDIVSDNMLYDISVV HGFGGCAAITSYYVDTVKILNNTIRKTAYNGIHLGWGWCNF KDSTTCRNNMICYNRVIDSLNRLHDSGGIYTIGQMEGTVIN ENYIQGIPAAGSFQPTYGLHNDEGTAYIEENDNVLEISHNVT YTINCEDYGQKHHLKIKRTYATVKKMGKNPPDSDIDDPIVV SDNVWDLPQYKVCVNSGVSDEYRSLVPNYVISEADFVFPA SCRTTCSSSLPIRKGDGIVWIAPDGTDTFKAGADMTRASAN ATSIRTPSKEGEYRIYVTDKSGKILSKSGHILRLS
Rf 101	[(1-28)SIGN][**(29-239)X148**][(240-279)LNK][(280-424)COH1_b][(425-453)LNK][(454-598)COH1_b][(599-631)LNK][(632-718)DOC1]	DASKGFAIKAYAEAGSKYDAMGSKVTVSKDDIAAGDVVVP VAVYLDEATNDSEAVSVSVKLVSDSADVKNVTFKRVIPTDD YFTTAKEYTAGDKTFSSTRAVIFAGEVSRRGSFTPAGSWEI AADTSQKEAGADNAYIGCSWTNNGSAYEFTGSKSTDHPFF VFDVTLPKGTAAGDYKLEFCRYNTDTSGQHNNPTPMIETK AGRFNEDLK
Rf 102	[(1-25)SIGN][**(26-63)UNK][(64-82)X159][(85-105)X159][(108-128)X159][(131-150)X159][(153-173)X159][(176-196)X159][(199-218)X159][(221-241)X159][(242-280)UNK][(281-301)X159][(302-326)UNK][(327-347)X159][(350-370)X159][(371-393)UNK**][(394-479)DOC1]	ENVNSGKCGDNASWKYDGNGKLTISGSGKMYDAIDSWNS FSNNITEIEVKSGITYIGVHEFDRLENLKIVSLPNTINQIGDCA FSMCINLEEIKLPDSITSIGSYTFEGCNLKEIVLPQKLSSISDG LFSSCFDLSNIIIPDTITEIGHDAFGGCTSLKTIQLPSNLTSIG EFAFDSSGLTQIVLPESLQDIENCAFVECNNIKSITIPKNVRII GDIEGGKIFSQNTKIDVSHDNSYFVSENGILFDKNRTTLIHY PIDNSVKEYIIPDSVKKIYPCAFLGATNLEKVQIPNKISVINDS TFANCTNLVELDFPESVTEIKTAVFYGCSNLNSIIIRNPQCVI SDPYWESTFKDFQGTIYGYPNSIV
Rf 103	[**(26-493)X160**][(494-574)DOC1]	KETDLSSIASSLGADTDYLDIPNYKHSSELPFPKSVYKDFFS KCTNYEASSQSEEVFCLELASGTCYGISAIEVLSHNGIISPS DIMLNAQTLSEIQLSQEADHFISGYHASQLHYENDYYNRYL

		VTQLNHQSQCNELIRIAEKNMADGKYFIITYYGDTAHAVCGI GITDGNWNYNGVDFDRCILTLDSNAFDDDTNKGPFSERSCI YVNSQTCDYYIPKYNFGSIPTAQQVQILANDNINIINYKGVFE PTKELDFDFTDTARIFFTQSRFKKYDAIVKDKNGMYTDVIKE GRKYEFGNDVGYYIIGNEFHIETDNKLLLGESQNDKFNIMN QRYYFEGETENNHGIFDISDTEFSVATKNDKALVYDMKIKY NEGYYSCSPHFNWTFNGETDHNFKSILVDKGMILQSDGTIK TSISTSDVLLDGQDNITDININPNETTITAVNDVFVTFNENNK LCFFIDPDD
Rf 104	[(1-27)SIGN][**(90-331)pfam00112**][(332-988)UNK][(989-1001)LNK][(1002-1075)DOC1]	SAYDMREHGAVTSVKNQSGHGTCWAHSSAAAAETDIVRR MPDVDLSELHTAYYSYGGLGQIEPPSDDIDEILDYGGNSSIV VNLWSQWIGPEFESVMPYDSLDSLKDPFDVVVGNGSGVF HLENAVLLDYDTDRTNMDQVNAVIKQSVMDGKGVDVTFCS DSEKYFNGVYGTTNCNKKPRFANHAVMIAGWDDNYPAEN FNVRPEGNGAWLAKNSWGSAYGNDGYIWISYYDKSLSEF TTYE
Rf 105	[(1-29)SIGN][(30-165)X140][(166-177)UNK][(178-244)X135][(245-399)UNK][**(400-750)pfam08757**][(751-849)UNK][(850-911)DOC1]	KVVDQNVGIRIKGAYSRNSVQKSFNILARMDYGKAELEYDF FNGTATKAKNGKKIKAFDSVTIRNGGNDVGAAYFRDTINQS LVTDRAFTHQAMSECVLFIDGEFWGVYQLTEKVSDDFISSH YGIKKSDAVIVKNDELEEGTEADFNEWNSLVSQFASADMT NASNYSQFCEKFDLRSFVDYFAAQIYWSNSDWPQNNLAA WKTNALDETNPYADGKWRMALFDTESGQGLYNSANNNVN SNPFSRISMNNDNMSRLFNNLLKNDDFRKEFELTMMDLAN YNFAPEKVTPVIEHYKNTYKQQILDTYERFFSNNLSGQRGE EKLNNEYNTITNFYKGREQYITQNMKQAL
Rf 106	[(1-30)SIGN][**(31-251)UNK**][(252-309)UNK][(308-496)DOC1]	DNDNGFVSSEGAETAFDSVRNESNVTTTLQTQITTTSTTGV STTTVTTTANPRGYDFDGTIFRKDYEWQYYLNDELDLSKLT LAVAVIDPKGDYSTEIVESTFSYESGKYSDLYTLDTSEVDM STPGKYKIYIRSKKDAIGDFETFPSRFLSAGHYKVRMDGHE SYFTITVRDIERPVEVEDTDFRFVNHQGSNDCVSITKGMSA AVALYGYKLKEIMENY
Rf 107	[(1-25)SIGN][(26-48)UNK**][(49-149)X142][(150-200)UNK][(201-310)X142**][(311-759)UNK][(760-831)DOC1]	DLSFIEDAANNYYCVAVGMHGVECTVPSEYNGKKVGQIDL NHVYIADYDFPSDIKSDTVTIHIPDGMKVGGDFWLAEQTGV PCIRLVYESGKTEVYKSADYDDLLADVKTQTGSHMELTEEE YRDMMLQIVPFSPGRENVEYPILEIDPRYNCGLSTYKDEKG YTYLKVETLAMTGEIHLPAEFNGEKIDRLKLGDIHASEDGTM GWGNWGLTVYLPDSIWVSDLHWVDTDTGFSDIILVDEAGN EELFEAELQNAPYIDVN
Rf 108	[(1-47)JUNK][(48-74)SIGN**][(75-150)UNK][(151-171)X159][(174-194)X159][(195-217)UNK][(241-261)X159][(264-284)X159][(287-307)X159][(308-333)UNK**][(334-352)LNK][(353-373)X159][(376-396)X159][(399-419)X159][(445-465)X159][(468-488)X159][(491-511)X159][(514-534)X159][(569-646)DOC1][(649-710)LNK][(711-764)UNK]	GIDITDGRITDSAETDNNETSGMCGNQLRWNFDSEGTLTIS GEGEMYPFGTDLAPWANLEVKKVVLESGVTYIATNAFYDC YMVTSIDIPDTVTEIDMYAFRGCTSLSEINFPEHLDYICWNA FEDTPWYKNNADDMIIVDNILMRYKGTATDVIIPDTVDEIAD YAFKDCTSLSSISIPDSVVEIGRSAFQKCTSLKSISVPEKIVSI GDYAFEDCSSLSEISIPDGILKIGIDAFSGTPWYTKCDDMIIIG GIFYQY
Rf 109	[(1-27)SIGN][**(28-75)UNK][(76-96)X159][(97-133)UNK][(134-154)X159][(157-177)X159][(179-201)X159][(204-224)X159][(227-247)X159][(248-384)UNK**][(385-445)DOC1]	ETIIVDNGLEYTAVSDSVLMLTGVRDKTATSVTIPASVDGKN VIVENGIFSDCPNLKSISTDENSSDIKSIDGVLFDEGGSRLLA FPRGLKGEYTIPEGTVGIAENAFENSAGLTMINIPDSLTTIGS YAFKNCTTLTGFSKPIPLTLTGEALYGCTALKSIELARSSELK YIGAFKFENCPNLETLIIPNNYILTSSFNINNCPKLKNVVLPD RSNDLLLTVTNCAQLDSLILPTSGDNGKGFYSHATISKCPSI TELNISNAQKVSVKDMDSLEVIKLTISPYGSIDEINNNIDYAT CPKLKDIYIYNADIQPNAKEIELMAANDITLHCRKTEKWSSY LDSHKVKYVFIDDEII
Rf 110	[(1-21)SIGN][**(22-95)UNK][(96-116)X159][(117-222)UNK][(223-243)X159][(246-266)X159][(269-289)X159][(292-312)X159][(313-384)UNK][(385-405)X159][(408-428)X159][(431-451)X159][(452-476)UNK][(477-497)X159][(500-520)X159][(523-543)X159][(546-566)X159][(569-589)X159][(592-612)X159][(613-658)UNK][(659-679)X159][(682-701)X159][(704-724)X159][(727-747)X159][(750-770)X159][(771-826)UNK**][(827-918)DOC1]	SMPVNGNNGYVTACADGETAPVSGELKTEYGIICYTIENGE VKIDNYYFDKSDIPEVSLPTEIEGYPVTTVAKDAFFNETNVT GIILPESIKRVEDYAFSSDSSSSLRWVRVENAELEIPQDSRP FSGDITLYGKKDSTAELYSYRENLSFIDYETKVKYGDITGEIV NGEIAIISCDKAAAEVTIPEEINGIPVTSINYLAFGGCQDLKSV YIPDSVKSIGESAFINCISLTDIRLSETITEIPAYCFSRCKVLM DITIPESVVSIGDSAFESCGYMERITIPESVTKIGSYAFRNCL CLDEVELLNHNISYGLAPFENTDFIEKFDNSDGIVVIDHCLV DGRNYKGDSFTVPCFVFEISPRAFEGNTNLTRVIIDPNVKKI GDSCFSGCTSLELAKISGDVQEIGSNCFSGCTALEWVKLP DTVKEIRDGTFDGCISLSEIFLPAKLEKIGKQAFGSCGKLEK MELPETVTEINSEAFKGCAGLKEITIPSAVTELSPYCFNGCS GLTEVTIPANITEISSNSFSGCTSLEKVTLHDNIKKIGFFAFS NCQSLKEIDIPDSVKEIEMAAFSNCKSLESIKIPEGCKLGIDV FMNCTSLAEVKLPENIDISNAMFKGTPWLDSIRKGSELIIFN NKVFDGTQCKGEVVIPEGVTEICGHAFDGSEITSVKFPDSL KTIGNYAFSNCDKLEEFTIPDGIGKISGGMFCGCENLKKVNI PDSVTVIESGAFEFCTGLTEFTVPASVKSVGMVFEYADRLK

		TITFLNPECFIAPDGENFLT
Rf 111	[(1-13)SIGN][**(14-148)X140][(156-222)X135][(223-361)UNK**][(362-703)pfam08757][(704-792)UNK][(793-860)DOC1]	VLMASSMHCGVLPAVAAQSADGLCINEVCTQNKNCYADSL GRASDWIELYNGGDSDISLDGFGLSDDAASPMKFSFPSGT VIRKGEYLLVIANKDTAAADELNTGFSLSKSGETVVLTSPEG ETIQSLNVPAMGEDETFGRTRDGGYAVMRPTPSEANSESA AEPVFSLESGFYSVNDVKELTISSSDTVYYTLDGSDPTTSD TAQVYSGAIPMYDRSIDDNVYSKYQHEENSPYSITLQSRFS ANPAKFDKATVVRAASRSTDGTFSKVATKTYFVMSDDKLA YYKNIPVVSLVTDPDNLFDKDKGIYVVGQQYVDWKKSPQY NPRKSEWDTDNVANFFSKGKDWER
Rf 112	[(1-13)SIGN][(14-148)X140][(156-222)X135][(223-361)UNK][**(362-703)pfam08757**][(704-792)UNK][(793-860)DOC1]	EASITYFDGGEMGFSQNMGIRIKGASTRNSQTKSFNVYAR SQYGDSKLDYKLIDDNYSAVDGDKIKRYDSFSLRSVSWVD RLRERVVHSSLRDLPALATYDSDRCMLFIDGELWGMYEIIE KSSDYFIQSNYGVPSENVAMVKNGELETGAEADYDDLEAL CDYCLAHDLSVQSNYDYVASKVDVESLIDCYCAGLYLGTW DWPNHNYLMWRNNGAAIEGNPYSDGKWRFGSFDFDYSV GLTYQSFGGVEGYQYDSFRKMDGEKEDMPTAIFTGMLKN EGFRQQFADKFYSYAYSVFEPSKMTAELDDEENRYMDYLT MTGWRWNNGQPNSDFNTYCSQQKA
Rf 113	[(1-24)SIGN][**(25-72)UNK][(73-93)X159][(96-116)X159][(119-172)UNK][(173-193)X159][(196-216)X159][(219-239)X159][(242-262)X159][(265-285)X159][(286-327)UNK**][(328-401)DOC1]	ADGTNNGYELIFTIDSGSVTITGVSGSGSTLEIPGTIAGLNVT SIADNFFTGSNELRTVILPDTLRSIGTRAFSACQKLNSVYIGS DTSYIGDYAFTACPSLSNISVSTANTVYRSENSSLYKGDAL VLYAGSDDAVISSSTRVIGKRAFFGRTALTSVDIPDSVEVIG DYAFSGCLALNDITIPDSVTSIGNYCFFSCSGLESAKLSNSL KAIPESCFSGCSALRDINIPVSVSYIGANAFYSCESLKSIYIP PTVETIGTNALGRTYSLRSGSEDNISDFRILGSPGSTAEKYA SALSMVF
Rf 114	[(1-24)SIGN][**(25-253)UNK**][(254-324)DOC1]	DDNYGTEVPELLELYESYCDTDYEALYFEEATSEYSSYCAI KNLADMGYVQFYGEADDEVDLNKLRQICGYSENELQLRCS YSKRAGNWSLNMTFLTFDHDANYEAALKITDAIAESYKIKSA YIEVDGKIIIRNNRTCWDRIRRIDEFGIESPLTEELTSKQIADL NSDIAGNGYKASIDENTGSIIFAEGVTEKEKLEFAIWFKGNY GFYAFTFSAALMGEEIPYER
Rf 115	[(1-21)SIGN][(22-207)UNK][**(208-228)X159][(229-307)UNK][(308-328)X159][(329-411)UNK][(412-432)X159][(433-556)UNK][(557-577)X159][(578-747)UNK][(748-768)X159][(769-838)UNK][(839-858)X159][(859-973)UNK**][(974-1020)LNK][(1021-1105)DOC1]	TGFDFIGDSALKDCKYITEVTIPKEIKYIGTGAFENSGLKKLY VNNEMPVVPEKFCSGTNLTEIKFAHPDYIRTIGKEAFKSTPL SDAVVTNADYKGDANYEFVEIGDSAFENCANIKTLAFPANVI TLGKSTFKNNTALTTLKFGENLIAADNSCFYGCTSLNSITWN NVLETLAGSCFTGCTSLKTVSGMPTTLMDWVPEDETLGW GVGDGVFSGCTSLEHVILPSSLKRIPGSMFEGDVKLKSVQF GDAKKEIDSKLTVASGDNIVKIKKNAFANCELLQKVDFGKTT YIEEKAFLNCKGMTSFNVGECSVVGTSALEGCSALKEITLL SDQYGGTAPNVKSPKDPNKTNSSEGYVFKGCTSAKKITIKT DDKVKLSSGMFQNCTSLTDIGGDLSKIEIIGKDCFSGCTALT HLNMPILKIIESSAFANCTALKQITDDPTLPIKATDYGDKAFQ NCSELKFTLTGDISTIGASAFQNSGVTSLSLNGMVGGTVVI GNSAFADCANLKSAVINSQDAKKFSVGTGVFANCPILEKAS YDGPLVTASMFKNCPKLNDIKLLADVINANAFEGCSSIQSLI NKNDGTLLIAKEIDGAAFKGCSALINSSADANTIFKGSQQYS GCAALKSANVSTLTSGMFENCTSLNSVKTNNVTSVPAGAF QNCTSLDTFVFTNIESIGANAFANTGFKTIELPANVASIGNN AFINCAKATSATIANKNAAITGRAFGYNANKQIPNFIISGIAG STAETYAKNNKLNFMDE
Rf 116	[(1-21)SIGN][**(22-86)UNK][(87-107)X159][(110-129)X159][(132-152)X159][(155-175)X159][(178-198)X159][(201-221)X159][(227-247)X159][(250-269)X159][(270-294)UNK][(295-315)X159][(318-338)X159][(341-360)X159][(361-432)UNK**][(433-465)LNK][(466-537)DOC1][(538-550)UNK][(551-620)DOC1][(621-666)UNK]	FISDIAPDTVITASAADYEKVTEGDFTFNVYADHAELIEYSNN AETDVTIPSKVNGQAVTVIGDRVFNEKKEISSIAIPDSVTQIG NSAFYGTGIRKIVLPSKLKKIGGFAFQSNKKLASVSIPDGVE EIAKNAFAYCEVITSLSLPNSLKVLGENAFSGCVALEEVTYP SSIEEAGDNIFFNCIAIKKVTFEKGVTEIPNGIFYMDNSNSVL DEITIPDGVKKIGDKAFYGTAIKKINIPSSVETIGDGAFWSCL KLGAVKIPAGISRIGKNTFRNCQVLVSVDLPSSIKYIDDSAFY VCNALEEITLPEGIVEINDSAFNGTGLKNVVLPKSLELLGKA FGGCKQLTGITILNPDCSITQDENTICNTTWAPLIGRKYEGVI YGYENSSAQKYAERAGYDFNLIGSAPE
Rf 117	[(1-24)SIGN][(25-348)UNK][**(349-369)X159][(372-392)X159][(395-414)X159][(417-436)X159][(439-459)X159][(461-481)X159][(482-547)UNK**][(548-614)DOC1]	GIPESVSVIKALYNMNFDEVNELTIPKNIKIFGAYREPKGVIV AEWGKIPASVPLKDNFTIKGYKGTEAEIYAKELNIPFIALDDL ETPLSGNYSENIKWTLDADGVLTLSGEGEIPDLTESAPWSS RRADIQTIIVEDGITSIGKDVFIGLEDLTSVSLPKGIKSISDNA FYKCINLKTINFPDGLERIGSDAFFGCQ
Rf 118	[(1-26)SIGN][**(27-40)UNK][(41-61)X159][(64-83)X159][(86-105)X159][(108-127)X159][(130-149)X159][(152-171)X159][(172-198)UNK][(199-219)X159][(220-242)X159**][(243-514)UNK][(515-578)DOC1]	VDTDAPEFDIENGVLVRYNGADSDVEIPDGVTAIGEYAFYD KKVQRVSLPDSVTKIGDSAFSRSTLTDINFPEGLQSIGNGAF QNTWITEANIPYGVTEIPSAVFFNSRLTKVTIPDTVTQIGESA FGNTPLQNVDIPDSLTYAASSSFSGTPFLRSLIETNGGWLIL SNGLLVVYAGDDINVVIPDSVKRIGTKSFTLRSRMQSLTIPD TVKSIE

Rf 119	[(1-30)SIGN][**(31-78)UNK][(79-99)X159][(102-122)X159][(123-244)UNK][(245-264)X159][(267-286)X159][(287-357)UNK**][(358-431)DOC1]	DVQLQDSPLSYENNGYEIKITACDTSYSGEVKVPSEIDGLP VTVIGEGAFKNCFRVKKLELPETLVRAEHEAFFNMIGLDELT IPKNVSTLGNYAIGDNSDFPLKVTFESADTNFSTACFDCVK ESGNEKRKELTLIGTEDSYAIKFAKGWKMKYKVVGEETASN GKKEVTADGMAYNVYSDHAELISSDKDITGKVVIPSEVEGV PVRKIGENAFTGRHIDEVIIPVSVREIGQAAFASTDLKKAVVP YSVTKIAKEAFISDQLESVVILNERCEIADDDSTIANKYPLSE GRYVYGGVITAPENSTAETYAKAWGYEFKELGLVE
Rf 120	[(1-28)SIGN][**(28-99)UNK][(100-120)X159][(121-195)UNK][(196-216)X159][(219-239)X159][(242-262)X159][(265-284)X159**][(286-501)UNK][(502-583)DOC1]	TEFTYSADYIAYKDETLPDLEADHREFVKDGLVFNIYDDFAY LVSCEDTDITDAVIPEEADGVPVVGLTDTPFGYCRSLKSVTL PDTMKYIDWLDLAASSGKVSTDKTDGEILPTLEKVTVSENN PYFTSENGIIYSKDMKELIGCPPAMEMKELKISEKAEAIKDF AFAACYKLEKAVIPENIKHIHNSAFVACKNLKSVEIPSGVTTI SGDAFFGCSSLSEVKINSKLEKIGFGAFSGCTALKEFNIPET VSVIG
Rf 121	[(1-23)SIGN**][(24-198)UNK][(199-311)X142][(313-422)X142**][(423-879)UNK][(880-952)DOC1]	VTGSAQITTDMNITQTAVKDESEDSDLSFMEDDESDGYYIAI GVKYAETDVPAQFEDFNVRKLCLDHVYVSDDCPIGSVGED FFTLNIPDGVEVVKKNWKCAITGIPCIKLVYASGETEVIKADD YDEMAEQTTEDLKNDGFDITEDLLFNAMVFSLRRNPQHQI MAYPIKRYNATKYYTYTGEDSLTYLRVNVSREDKNLLISIPD EYDGKKIDRIKLGDVVCQDNIDSGGFDSVELYISEGVELDP DVKWNAYETGFHTIIVYNSDAERKAYYAMVTEMGDDTDTY DTNYYAYTGEDSLTYLGVNVGMADINQSVNIPDEYDGKKID RIKLSDVVCPESIYAERLDAIELYLPEGVELDPDVKWNAEGS GLHTIVVYDSDLERKEYHVWPSEMADLY
Rf 122	[(1-24)SIGN][(25-348)UNK][**(349-369)X159][(372-392)X159][(395-414)X159][(417-436)X159][(439-459)X159][(461-481)X159][(484-504)X159][(505-547)UNK**][(548-615)DOC1]	VFLGLESVTSASLPDGLKSIDDSAFEECVNLKNVNIPEGVES IGNESFYKCALESLELPDSIKTIGEKAFVSGNFESIELPSGIT RLENGALGSCEYLNEVTIPDTVTYIGGTFGYCKSLTSITIPES VEEISYRAFSYCENLETIIFKNPDTVIIDEEGDVHDSFFYAW GGDDFKGVVKGYKGSKAEEYAIKYNTGFE
Rf 123	[(1-28)SIGN][**(29-92)UNK][(93-113)X159][(116-136)X159][(139-159)X159][(160-207)UNK][(208-228)X159][(231-250)X159][(253-273)X159][(274-324)UNK**][(325-382)DOC1]	SNVMPITAHALYSPTDTNYTTGQSGGLQYAKYSDHIEILGC DSNTTTIDIPETIDGLPVTAIARYGFECSSLTSVTLPESIKTIG YWAFAMCSDLTTVKLPDSLEVIEMHAFELCPKLDTIEFPDH MVEIHARVFDETPWLEAQRKIDPLVVVNGALIDGRTATGDV VVPSGVKYVSASTFWWNTKVTSVVFPSSVTTLIDNTFFQC EGLTSIELKGVTEIESMAFCGCTKLNDLKLSGKLTKIADDAF ADTSSSSTITFYGSRDTWERVEKPNDSAFLQRATMVFDES GGPADEV
Rf 124	[(1-29)SIGN][(30-79)LNK][**(80-149)UNK][(150-170)X159][(173-193)X159][(196-215)X159][(218-238)X159][(247-267)X159][(270-290)X159][(293-312)X159][(315-334)X159][(335-390)UNK**][(391-451)LNK][(452-532)DOC1]	AVAEEEEKPVVAETDEPATEEDKKTDEKEEYKLPEKGISGD TSHGDKHPEYRLTKDGLMSFDGEGQLDKWEPAYCYHVKD IRDKAVVWSSNITLGDGAFSWAPYTTVDLMLTTIDTVPMDT FYNDINLKEVMLPDTVENIGEYAFYNCQALETLEWPRSIKRI ENVAFENCGLKVLELPKTVEYVGDNAFRNCDNLTEVYINSE FKYSDEEKKDIGYIFTLCDNLEKVELSDDVETIWSSEFCQCS SLKDVKLSKNLKVIKDHAFQSTPLEKIDLPDGLEIIEESVFINT KLKEVIIPKSVTFIGNSAFSSPTLKKITILNPECDLNISSIYGSA DTVIYGYKGSTAEKYAAKNSIKFVAL
Rf 125	[(1-28)SIGN][**(29-128)UNK][(129-149)X159][(152-171)X159][(172-192)X159][(193-212)X159][(213-233)X159][(236-256)X159][(259-279)X159][(282-302)X159][(305-325)X159][(328-348)X159][(349-364)UNK**][(365-433)DOC1][(434-452)LNK][(453-526)UNK]	ADDITTTPAVTVYETTTEIVTTHPPYYTEIGPDVRKGTCSSG ENIVWELSDDGTLTISGKGVLCDSPWQNHLKDDIKKVVIGE GITKVSLSSFALPSQVSMFRGCSNLESVVLPKSLVNFDGQA FADCEKLSSIDFPEGLEYLTLMSCPAIKSVTIPDTVTYLHISC CDSITELDIPDSVTTINSLAGNTSLKTVSLPSELTELNTNLFD GDISLENIILPEKLTVLPKCTFYNCTSLTSIKLPENLTKIGDNA FTSCSSLRSVVIPDNVTEIGREAFKNCEQLETITIPKSVRTIG MYAFENCPSLKAIRGYRGSAGFIKANSLDVKFIDLEKIN
Rf 126	[(1-24)SIGN][**(25-199)UNK**][(200-265)DOC1]	EDNLLNYTELVKLTDEEMAEKYDLLPYHYGDDSVSDDIKFK EYLNQPIDGFIGDNYESVSYSAYLKFINGQNIPYIAFSVDRY TKLDTSLTAEDFGYPKEWKITAYDGVFNTDTGYPRQFHEY RIEIPVDIIADFEEYVRLEKSFIFNEYNYQEDNPYAIKTFFDID HQFVSYGG
Rf 127	[(1-25)SIGN][**(26-268)UNK**][(269-335)DOC1]	QTVCAYDQLIASVTKVSIAKDVLAIGESAKLELTWSSGKQN EVTFSSSDESVATVDKDGVVTGIADGKATITVSYLEGKGVK TIDISVSHEAVKSEVYNTSEVSLGDKFRKYDTLHYDGKSKG SCANVVNTKGNYDLVYINIDDYVLPFDAELVGIDGLVIYIAPD IEGITYLDGRKLNAGDVIDRNTHLLCYDYKIKSADKSSRMIF PVFLPEYYGEYIGDGEIKVKSIDHDAKVIELESVD
Rf 128	[(1-24)SIGN][**(25-509)X160**][(510-582)DOC1]	EGNPDLQALASSLGADTDHFSFKNYRGASISDEMLSAFKQ KITPTEAMLNRPENAVVSSKGGECNAMAILEILVHNGIISAS DIQEGADFLCDITFDDKINDILTYYQMTQVFQKQYLAIRNYW CNHDVSDAISDLVKYGNRAVEEGKYFYISFSWDKGAHAVV GIGETDGSWEFNGRKYNKCILTLDSSFDSFTEKACIYVNTE ENTFYFPAYDYTENDAQITMITGDETLLNYKGLLSPSYSSDT NTDDLTEIEIGNYKFVNGKSVGSDIEFYITDDNGTRTYTSKG FEPLDITKNYFINMIETNKQYTFPAAKNTKYKLKINETSDDPE VDSVEADLFQITENCFRYCYSIGFDKNTEIEFGEHYMSRYV

		PNYSGFSFHLLSEDTPYKNDEFNMYMVLGNNIGTIAMHDR NDGILLTTDQKFDAFVSFEGLIKNDDGSIEAFKKSADQHIIYC TLSSVNNIMLRYNETDDRIAIFIDNNNDD
Rf 129	[(1-23)SIGN][**(24-311)X128**][(312-340)LNK][(341-405)DOC1][(406-819)UNK]	QSAVHFHMNNTITASAESISDMPADFQYAADWIWQNRIDR EKSTVRRNTIFDQIVAGKGTINYVVKWQSYRTVTYEQRQKF EKMLSDCINAWNDWLAGYENWPYDHIDVKIVGWAVINKNC LLDLHDDEVVYTDTKYYDAQYDTSNGRDTIPDKEPFAPSEL SRFDHFTEKNYQYPGGLDSRFDMYMWATQGFPDIGGCG GDWGQRLSDTAYLGMIQHGSLHVLEHEIGHGFGMTDFYG GEGESDGFPPGGFPRGENSIMMAGSSAKITDFDGWMLRY MWTKIKDDKD
Rf 130	[(1-39)SIGN][**(40-88)UNK][(89-108)X159][(111-131)X159][(134-154)X159][(155-202)UNK][(203-223)X159][(226-245)X159][(248-268)X159][(269-317)UNK**][(318-381)DOC1]	QAEEEYKDGNLTFMLYDDHAEVINFDFTATTAEIPATVKGL PVTSIGIYAFNGSSVTSVTIPDSVTYIGQWAFAMCGSLKEVT IPDSMEHIDINAFQLCSSLSEVSFPDKFVKISGGAFDSTPWL DAERKKDPLVIVNGALVDGRTCKGDVEIPSTVKYIASGAFQ RNSDLTSVVVPSSVKEINDSTFFYCDNLVSATLPNVELIDSM AFDGCTKLSEVKLSGKLKSIASYAFDDISASGTITFYGSKET WDKVEKPDDCEYLNKAKYIFDENAQPPE
Rf 131	[(1-40)SIGN][**(41-89)UNK][(90-109)X159][(112-132)X159][(135-155)X159][(158-178)X159][(181-201)X159][(204-224)X159][(227-247)X159][(250-270)X159][(271-440)UNK**][(441-485)LNK][(486-580)UNK][(581-599)LNK][(600-669)DOC1][(670-873)UNK]	EETYGDYSYTINEDSEGEYVTITGYSGNDTTVVIPSKIKGLP VKDIGGSSFKNTRISSIKIPEGVTSIGGWAFLGCWNLESVSI PDTVKNIGSLCFYNCSKLKEIVIPGSVNTITEQSFYGCTSLK SIEFKSGVKEIGKDVFDGCTVLTDVSIPDTVTKIDEKAFNNC TSLESITIPDSVEEIGASIFYGCTSLAKVKLSNNINTIPVNAFF NCVSLKEITIPYYVESIGGSAFIYKNKSGYYYSIDIEKINIGAK LKSLKNLPVNSETLQEINVDTQNGFFSSEDGVLYNKDKSQL IKYPSAKADTEFTVPDTVEQISENAFANNTVLKTVYISENTK AIEPKAFYNCTHLDEVYFYNKDCEIYMSKDTINSSAVINGYK DSTAEEYANTYGFAFNEIQ
Rf 132	[(1-20)SIGN][(**21-49)UNK][(50-140)X142][(141-149)UNK][(150-240)X142][(241-313)UNK][(314-400)X142][(405-500)X142**][(501-881)UNK][(882-956)DOC1]	VPFVISYAEDQYTEESWNNEDHGVLKWNYDYGFEYINLKID PNSNEVVIPSEVNGEEITGLSLDCFQCPDSYCMNNTIKLIVP ECVEDIWTNWTYEKTHIKKMTLEYSSGESETLMAADLNNS TFSYLYDEEDDDYYLGINLAGEFDALPTEYKGKKVDKLDLE HIHYTDNFDPYTGYELEIPEGIRIVNKHWKGAVTGIPEITLKY PSGETEVLRSDDYDELVEYTKNILRNNEFDENYIEENLQNE MSFLLMSHPDHRCKDYISEDIEAPVNNEEIPQKDEDGIIRWF YNSDNEYIDLKIDKESYEVILPTEANGKKIEKLCLKDLDFSNN TILNKAPVIIIPEGMKVVDKNWNNNDTSIRAMYLSYPSGEKE ALLASDIHNTPFSYLYDPEDDDYYLSVNISCMSGDVLPTEY NGKKVSKIDLEHVYYSEAYENADISLNIPDGISIANKHWKGA VTGIAEITLKYPSGETEVLRSD
Rf 133	[(1-26)SIGN][**(27-288)UNK**][(289-355)DOC1]	EDNADTAAAAATEAAETVEPVTADDIKGTWSGTYTGSTGS TTIEREISLNIDECDADGKFSGVATITSSENQSYCFSGTCDL ATGEIRFKGDEWIKNANKWSFLDFKGNLVSGKITGLTDNKK DRPFSLEKKSDSFVRYSVDPAAVKREWYGEYDGHSGSVV VRRNIKFSITDISDDGKITGSAVFSPSNKAEAKYALDGSYYF TGTLDERYGRLRIQGNEWIEHPAMENFTFIEFIGSVQGDIID GTTENGIWKMEATSIL
Rf 134	[(1-25)SIGN][**(26-257)UNK**][(258-334)DOC1][(335-454)UNK]	GGDMTKTPIDLQDYYDSVENFSKEEIPKYKCLTVDDVLDDI NGGEYVSLSTGLGSTMDEKGKLHKIDKRFSAAVAKIKAGTE SKPAFEELAKSPLVVIDGQYIIIRHGAAENIDKYVDTLKNDPD VLSIDLDYGIYEDTANYVSIEGFFYDGEKLTDEFLAEFPALAL KHSDNWSTEAHVYFVVFGGRDSDPVDIYRDIVKFKEKVPSI DVCAISTCLAMMKPEMYFCHESV
Rf 135	[(1-12)SIGN][**(13-49)UNK][(50-70)X159][(74-94)X159][(96-116)X159][(117-226)UNK**][(227-297)DOC1][(298-400)UNK]	TISMTSFAFLPPAAAAAPQTDLSDVQVAPEFDFTGQKIAELP AVDKLTVAANLKFENNQDLEKVVIPENYVLSGSLKFSKCPN LKEIVMPELATELWVTVTECDKLASFTMPVSPQTDLTQFSY VTVSNCKSMTELEVSNARRMSVKDMPALETLKLTASPHAT SDEEYYNIDYASCPKLKDIYLYNADVQPTPKEAALMAENGIT IHCPAADG
Rf 136	[(1-29)SIGN][(30-110)DOC1][(111-549)UNK**][(550-880)pfam08757**][(881-952)UNK]	FDNVAIRTKGNSSRQFVSQAGKDKFSFRIKLNKYDKLQNYH GLTDICLNNMYSDPSCMRDILCYNACYEVGSYAPLCSYTD MKLNGQLYSFYFMAEQPAETLAERLAVTDDSVFYKAADKM LAGSSYDCSFKPSMALENFEVKFGEDEQLQHIAEVKDAINK VSSSNYKFIEDIIDVPSFLQGFAVNAVMCNYDSYNGMMPH NYYLEYTNGKMYYVSWDFNLSLGNFMDNGASVNSDIKTAT YQTTVADRPLLKLLEVPEYYDMYVGYVKQIVNMYSQPEQT VDGIATLIRSHVKADPRFFFTGDQFETNIAKSANGLQVNDG GGWNMWGN
Rf 137	[(1-26)SIGN][(27-38)UNK][(39-181)CBM13][(183-249)DOC1][**(251-564)UNK**][(565-620)UNK][(621-945)GH43_G][(946-964)UNK]	**LSKFLLGEKINADSYCEISCEKPDVAYLFAYFLGNAPEQER LSYAVSTDGYHFKALNGGKAVWNSSVGTKCLRDPYIFKG EDGLYHLLATDMKSSLGWNSNRDILSAKSTDLVHWFDET SIPIANKYPNMMSADRAWAPQAIYDPEKDSYMIYFAARVP DIDNRTIMYYAYSKDLKKLDTAPQLLFAPKNGNDAIDSDIIF ENGRYYMYYKNETNKRIYLAESDHASGPYSEIKQVSEGSL GVEGPNIYKKIGSDKWLMMSDAYGDGYYVMQETSDLENF TSVSRNSYSFDFTPRHGYVIPINADQYTALVNAY**
Rf 138	[(1-30)SIGN][**(31-93)UNK][(94-**	EEEREASYTQQDMRQFIRIGPMFFRCDLEHAVLVNFDKTY

	112)X159][(115-135)X159][(136-263)UNK][(264-284)X159][(288-308)X159][(309-384)UNK][(385-457)DOC1]	EGEVEIPAYVQGVPVTALGERCLAFADGITKLTIPETVRTISA AALCDLYNIKELIIPASVETISGNAMYGMSGLEKVTFRCPYP YIIGENINFMCHLGCEGPQVPNDMKICGYKGSDVEKIAKELF CDFCDETTEEGSDVIAADGTWKDGFCFEINSEKNEAKLIGC DIVKANKSGVIPSEVEGVPVTEIAEYALTGAHFQKIVIPDTVR YIGDCGLGNYMYGNPEIVIPESVEYVGRDFIVAAYSSIETITF MSPDTKIDSCNFRVFDFDPTGGTYFRGPDAFGTECTIRGF EGSDAEKVAKEYGLKFEALT
Rf 139	[(22-47)SIGN][**(48-341)UNK**][(342-425)DOC1]	SFAGIMPPVSSVDEAAAYMRQCMKERQAEFTVTVPYNGYI DEDDAVKKMLAEAMKETSSGTEGDYLRFGMKGYKYGTAV RGGNITIIYQMFYYTTAEQEKAAEIKINAMVDHFSIYGRTPYE TIRAVYDYIAENVTYAEVDPNEEEHDDLSIFSAYGAAVNGV AVCQGYSQLCYRLLKDAGISCRIISGTSRGVRHTWNIVELD GKYYYLDPTWDTELGGSDGAFFMKGTSDFDEFSSKITHIPT YDYEIIFPDYESAEFKSAYPIANSKYIPPRYNKGDVDANGIID GRDATA
Rf 140	[(1-34)SIGN][(35-113)UNK][**(114-134)X159][(137-157)X159][(160-179)X159][(181-201)X159][(202-216)UNK][(217-237)X159][(240-260)X159][(263-283)X159][(286-305)X159][(307-327)X159][(328-338)UNK][(339-359)X159][(362-382)X159][(385-405)X159][(408-427)X159][(428-464)UNK][(465-485)X159][(488-508)X159][(511-531)X159][(534-553)X159][(555-575)X159][(576-585)UNK][(586-606)X159][(609-629)X159][(632-652)X159][(655-675)X159][(678-697)X159][(699-719)X159][(720-737)UNK][(738-758)X159][(761-781)X159][(784-804)X159][(807-826)X159][(828-848)X159]**[(849-866)LNK][(867-887)X159][(888-912)UNK][(913-933)X159][(934-960)UNK][(961-981)X159][(982-1714)UNK][(1715-1791)LNK][(1792-1878)DOC1]	AFKGNTEITSVKFPETIIQIRDAAFRGASSLASIDLPEGLQTL SPYCFAETALTYVKLPTTLTNASCPFSKCLSLEKVEIAEGAT AVPHGSMNLDTYGDTSYGCFENCTALKEVILPEGIVQIGSY AFYRCSSLEDIKLPSTLKTIGEKAFADCSSMKNIDMPEGVET LNFGVFSGTAIESVTIPSGLKKASRPFAGCETLKKVTFGPD MVKIPSGTTHLLEDIGIFEGCENLEEAVLPEGIEEIGYGAFSH CPSLKKINFPSTLKSIDALAFEDTPSLTSVELNEGLKSLNGS CFKNSGLTYVKIPSTVTYARRPFTQSQNLKKVEFAEGMVTT PSEHDIKSFAYTSEHGIFEDCPVLEEIVFPSTLEYIGHYAFAD CPSLKEVKIPDTVKGIGSYAFNNDLGITSIEIPENLESLYIGSF GNTGIESVTIPSTLKFAAAPFTHCEKLKKVVFEGDVTEVISG TKGSDTFGIMQSCYGLEEVVIPEGVEKIGTYAFYECSSLRK VTLPDSVKTISQDGFSSCPMLEEINIPSSVTEIGNRAFANDK ALKNIALPEGLETMGTEVFDGAGIRTITIPSTVKKADRPFSG CTALEKAVFADGTEVIPEGTVLSGKGFSAYDVPVGILENCA SLKEVVLPEGVTEIGRCAFSSCPKLNKVNFPSTLKKIGECAF KGDSRLTSVELPEGLEDIGVRSFAESGLTSVKIPSTVTSGS RSFEKCPGLKEAEIADGMTSIP
Rf 141	[(1-20)SIGN][**(21-94)UNK][(95-115)X159][(118-138)X159][(139-206)UNK][(207-227)X159][(230-250)X159][(253-273)X159][(276-296)X159][(299-319)X159][(322-342)X159][(343-389)UNK][(411-420)UNK][(421-441)X159][(444-464)X159][(467-487)X159][(511-613)UNK][(637-657)X159][(660-680)X159][(683-703)X159][(704-798)UNK][(822-842)X159][(845-865)X159][(868-888)X159]**[(889-1000)UNK][(1001-1069)DOC1]	TTATVNDKDRLTAVAAEADEIIEPDEDGIVTKDGVKYEIYNG DTWVVGYTKDLKGDLVIPEKINGVPVTIIDDEVFANCAEITS VTIPASILSIGEYAFKNCEKLTSVDIRSNCTGNGAFTFCPNIE KLTLSSEDNDFEYALFHTNKLSAADYDADKYIYDEARSFLIP KSLKTVEITEGTSIGTDEFKGMSSLETVILPDSIEDIGWSAFE GCTALKSINLPKNLTVIGSYAFSGCESLEEIVLPDSLEEICEG AFNGCTSLKTIDLPKNLKTIGKEAFENCAGITSVTIPASVKTV GQFAFKNCEKLKSADIKNPDIFNASISVSDWSGLFVFCPNL EKVTYAADGNSNFKATLFYTGKLRAGDYDADKYTYESFYG MLIPKSLKTIEITEGTTINEEEFKDIASLETVRLPKNLETIGEE AFSGCENIKELVLPDSLKTIEDFAFYGLTNIKSVTVPASVEHI GQGAFACCDSLTSADFAGDIIQEKSDGIYTVVIGFNENGAT SVSYGSGYGIFCFCHKLEKVSFNNADVESLASFLFETTESA VKDNNMEEFVVTAVDESYAGKTHLFAIPKSFKTIEMREGKT VKEKAFKDLSTVENIIIPDTIEVIDAYAFSGCSSIKSFDMPKDL KIIGDYAFSGMKGITSVKVPASVVYVGEGSFSECDNLTSAE FEGSYFNDSYESLYKSELHLSFYDWKGIFLYCPKLERVSCP AENDVGMARYFFQTDRDDIKERKLDDYVLTSSDDYYKEPE YPYVIPKSFKTVTITKGEILAEASMANLKTLETVTLPDSLKEI GVSSFSGCSGLKSIDIPENVTLISDWAFAGCNNIEELVLPEK LETIGNEAFGDMDSLTSAVLPASVKKVG
Rf 142	[**(1-240)UNK][(241-261)X159][(264-284)X159][(285-337)UNK**][(338-426)DOC1]	MKFSSWQDAYWYFLNISNEGKEPDEDRNLGPMYETRDIT GDGIPELFISESAYPLSKVYVYTYHNGTAEYLMSGGGNGIIG YNTDKSHEGIYLITSFMNQGITTYKVDIYNNKTVSPLAVFTS EDMYADEKNPAKYTFNDKEVTKEEFDKEFSKYKDLEISYVG RKDYFDERYPIIDDVVYYYYFDHFEVAGADNDIKNMKIADE VRGIKVTGIHEYAAEKHRKLESVTFGKNIEFIGRNSFEGCSA LKSVIIPENVKDIGEQAFLDCGSLEDVTILNPECSIWDFWEG GQPLTFCNTVDKDGNTVFNGVIKGQKNSTAEKYAGIFKLDF EEIESKS
Ct 143	[(1-38)SIGN][**(39-757)X141**][(758-901)CBM6_2][(902-965)DOC1]	AAEPITYYVSPTGSDSNTGTIDAPFKTIAKARDVVRTVNGN MKSDIYVYLRGGTYNITETITFGPQDSGTNGYRIYYMAYPG ETPVLSGATKVTGWTRHNGNIYKAKLNRSTKLRNLYVNDQ RASMTSKRVTARGGHGTYTVTAGQAPWAWTSGSKSDGV RYDMSEVPEITRNKDDLEIVNGTTWNENIVCTRDVITANGY RVLLLQQPYGAIAQTPGWGAAFTTSGTHTIYNAFEFLNSPG

		QFYFDKTEQMLYYYLRPGENIETIDVQAPMVEKLIEIAGTST SNRVKNITFQGITFAYTDYNLVEVGGSRGKSTCQAAQGFIA FFNDNWHYTKYDLVDTLPGMINLRNCDSIDFIENVIKHSGA DGISMVNDVINCKIIGNYITDITSSGITVGHPQHVYIGDGGSR AKFPSGVEGVCKNNTISNNVLYDISMVPGFGGCAGITAYFV EGLEITHNHVQKTAYNGIHLGWGWCNFKDSTTCKNNTISY NRVVDTLSRLHDSGAIYTIGQMPGTNINENYVKGIPPATYG PTYGLHNDEGTAYIIENDNVLNIDPGVKYTINCEDFGEKHDL TILRTYATVSKMGKNPPNSRIDPPVAVPDNVWPLRQYNVC LNSGIQDEYRKIMPESLLSTPDYVFPASCAAEAASIINIRSSG DPSNTVWFAPPGTTTFVEGATMTKAAGDATSIIAPYTAGTY KLYIVNSQGVKIGESESILRVS
Ct 144	**[(1-256)UNK]**[(257-760)GH10]	IVGKVLDMDEKTAIIMTDDFAFLNVVRTSEMAVGKKVKVLD SDIIKPKNSLRRYLPVAAVAACFVIVLSFVLMFINGNTARKNI YAYVGIDINPSIELWINYNNKIAEAKALNGDAETVLEGLELKE KTVAEAVNEIVQKSMELGFISREKENIILISTACDLKAGEGSE NKDVQNKIGQLFDDVNKAVSDLKNSGITTRILNLTLEERESS KEENISMGRYAVYLKAKEQNVNLTIDEIKDADLLELIAKVGID NENVPEDIVTEDKDNLDAINTGPAESAVPEVTETLPATSTP GRTEGNTATGSVDSTPALSKNETPGKTETPGRTFNTPAKS SLGQSSTPKPVSPVQTATATKGIGTLTPRN
Ct 145	[(1-29)SIGN][(30-103)DOC1]**[(104-469)UNK]**	ASNNPDAVIQFESGFAHSVLLKKDGTVWVLGNNGKGQLGL PEVSAVNEPVMINGLSGIKSVAAGREHTLALQEDGTLWAW GNNYSLQLIEYMERDPDTKERFTSIPIKVETHSDIKYVAAKF SRTLIVKNDGTVWLYSLPPINTSSDAEYMPWEIKGFGDIKM ADIGTGHIVALREDGTVWTWGENVWGQLGNGWQQHHNIH TYIYFEPNQAKNLSDIVSIAAGDAHSVALKSDGTVWTWGSN FNGELGNGTTTYILEPKKVEGLEDIVAIDAGIGHTVALKADG TVWVWGKNSYGQLGNGTTMRSTVPIQVEGLEGIVAIQAGM ECTIAYKNDGTVWAWGKNDFGQLGDGTFENILRPVKVFER K
Ct 146	[(1-23)SIGN]**[(24-477)UNK]**[(478-512)LNK][(513-580)DOC1]	AYNAEINGEVIVWNPGIKGGIPTKPVVANVKDFGAKGDGLT DDSNAFKKAVESVKDGGAVLIPSGEYLIKSKITLDKPVVLRG EGPGKTILLIDHSSDAFEVITYKRGNWVSLVGGYTRGSTEL VVSDPTGFEAGKYVEIQQDNDPDVMYTLPEWNQGWAAGS VGQITKVVSIWGNKITIEEPLRITYRSELNPVIRTQGFAEYIG FEDFTVKRIDTSDTNMFFFKNAANCWIKNIHSIKPAKAHVSV TTGYRIEVRDSFFDDATNWGGGGHGYGVELGFHVSDCLIE NNIFKHLRHSMMVHLGANGNVFGYNYSTQPYQSEGGNWT PADISVHGHYAYSNLFEGNIVQEITVSDYWGPSGPYNTFLR NRIESESVCLEDSSNYQNFIGNEIVNGNILWDTDNRYPHKID PSTLFLHGNLINGSIQWNQQTQDRTIPNSYYLDSKPAFFGG INW
Ct 147	[(1-26)SIGN][(27-97)DOC1]**[(98-537)UNK]**	NQYPTTPEPSPTPTPAVDEEAWKNNTGTIELGDTIKVSGEG ISVNGSVVTITAGGDHLVTGTLNNGMIFVNTTERVKLRLSG VNIKNPNGPAIYFYNVDKGFITIEKGTVNYLSDGSTYTDQDA KAALFSNDDLELKGKGTLYVTGNYKHGIASDDDLIIENGDIY VTAVTDGLHANSGIEIKGGNITVTAKSDAIESEKDFEMTGGT LNLTADDDAIHSEKDLVIDDGEINILKCYEGIESKTTITINGGK ININSNEDGLNAASGLYINGGELYITSGYDGIDSNGPIYINGG YIFSFGGNIPEGGIDCDWNPLIINGGTLIAAGGSNSTPSTSS TQCSVLLGSGTANSVISIQRNGSEIISFTAPKNYQNMVFSSP DLVLNATYVVYRNGVQSVTFTTNSIVTNAGGSSGGWFPGG GFPGGGFPGGGGGWFPGGPGW
Ct 148	[(1-28)SIGN][**(29-79)UNK][(80-204)cd05379][(205-328)UNK**][(329-411)DOC1]	ESVLQDRTIDDIVKRYQNNPFRINVSVSDIYEIEPKAEPPYV AGKLKSDYLKEALNCVNFMRYLVGLPNDLVLDDNYNNYAQ HGTVLLARLRGIAHYPQKPGDMPDEFYNLAYKGTSSSSIAY GFSSLMDSIMAFMKDNNSELNLSTVGHRRWLLNPGMEKT GFGQCGRYYCTYILDSVMGASVKFDFIAWPARNYMPVEYF NDASVPWSVNLGSDYFSPSLNEVEVTLKRRSDNKVWIFNK DNIEEYGLFNVNNDYYGMTKCIIFRPKGIGSYNKNDVFDVNI KGIRLSTDGPTEINYT
Ct 149	[(1-25)SIGN][(26-94)DOC1]**[(95-710)UNK]**	IEPTPTLEVSPTPTETSEEVFAFKIKLFSDGDTYRFPIQEISE NNNIVVDWGDGTTSTITDYSTLRHKYEKAGVYTIKVLWFDH IPIRFTGDKYVIEILTPLPDIGLTDFSSFFKNCSNLERIPDRLF SNNINATDFNFCFSGCTSLTEIPESLFAGNVNATTFVRCFY RCSNLIKVPEGLFENNVNATNFLGCFDECSSLKEIPEGLFS NNVNAANFSWCFSECVSLAKIPEGLFRNNTNATDFSYCFY GCTSITKIPGGLFENNINAEDFGNCFSGCSSITEIPGGLFEN NINAANFGSCFSGCSSITEIPEGLFENNINAEDFRGCFSGCS SIMEIPEGLFKNNINAEDFRGCFSGCSSITEIPGGLFENNINA EDFGGCFSGCSSITEIPGGLFENNINASDFSSCFSGCSSITE IPGGLFRNNINTTRFMECFKGCSSVTEIPEELFANNVDTAIFI GCFSECISLRKIPEGLFKNNINVISFMECFKGCSNLTEIPEGL FVNNTNATDFQGCFYGCSSLTEIPARLFTNNVNVTNFRECF RDCTSLIEIPESLFDSNVNVTNFYRCFYGCKNLTGVAPALW LRTNVKEFSGCFGSCTKLSNYNDIPKGWK

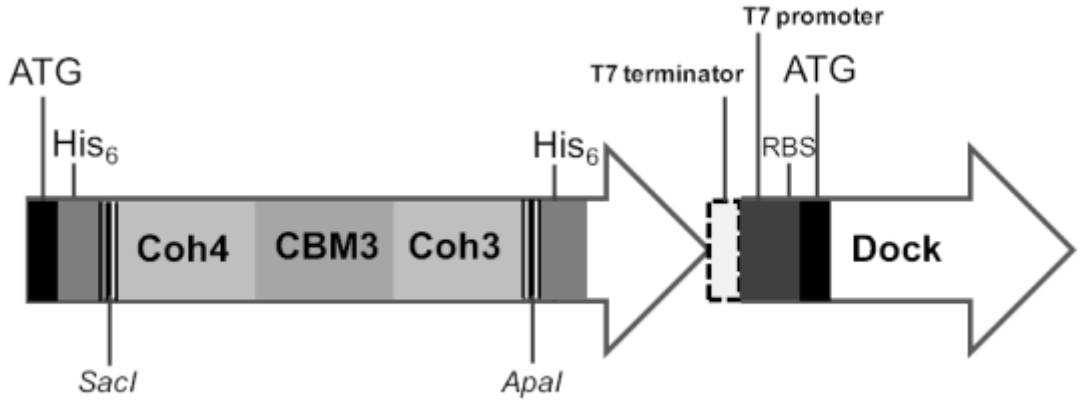
Ct 150	[(1-30)SIGN][**31-176)X140][(185-330)X140][331-490UNK][491-973)X139**][974-1051)DOC1]	SQTLFINEVMSSNVFTIRDGDVTDPKHGSKGGAYSDWIEIY NAGPYDVDLTGYILADSSAEWVFPQGIVPAGGYLLVWASD KNMVAQDGQLHTNFKLSASGENITLKKPDGTIVDSVDIIGLG DDQSYGRKSDGASEFVVFINPTPGAANVYNAPGTSTLFINE VMASNTRTIRDGDVDDPKDGSKGGAYSDWIEIYNAGPYDV DLTGYILADSSAEWVFPQGIVPAGGYLLVWASDKNKVATD GGLHTNFKISSSGEALTLKDPDGNVIDILVTINLLDDQAYGR KTDGSSELVLLKPTPGTANIYDPSLIPVSEPVFSHQGGFYT GAFKLELTTNEPGVKIYYTTDGSDPVPGKSGTIEYTSGINIK SRKGEANVLSMIQDISNDQWNRWRAPNGEVFKCTTIKAVAI RDDGARSKVVTHSYFVDPQMNTRYTLPVISIVTDYDNFFDK STGIYVNGNYENRGKEWERPVHIEYFETDGKLGFSMDMGL RIHGGYTRKYPQKSFRLYADHNNDIGEIKYEIFPGLRGTGT GKKIKSFERLILRNAGNDWTGALFRDEMMQSLVSHLKIDTQ AFRPCIVFLNGEYWGIYHIRERYDDKYLKSHYGLDDDKVAIL DVYQTPEVQEGDSSDVLAYTNDVINYLKTHSITEKSTYDYIK TKIDIENYIDYYVAQIFFGNTDWPGNNVSIWRYKTDDGQYH PEAPYGQDGRWRWMLKDTDFGFGLYGKSPSHNTLAFAA GDIREGQANEEWAVFLFKTLLKNEEFRNEFINRFADQLNTS FVPSRVISIIDDIVATLEPEMKEHTDRWPFIKLTATSPWDTT WSQEVNRIRNYANSRPSYVRQHILSKFRNNGVTGTALVTL NTDSTRGHIRINSIDIVSDTPAVTNPNRWSGTYFKGVPITLK AIPKEGYVFDHWEGINGSVEASSDTITVNLSNDLNVTAVFR P
Rf 151	[(1-24)SIGN][(25-49)UNK][(50-416)GH5_4][**(417-607)UNK**][(608-681)DOC1]	GSGSAPVINTTVPTSPTTTTTLAPLADGEKLYGKKGSEGTV TFTKAIGDNAFVEIKTGADTGFMNGCLGFSESIDGKNYWVA YVWQTKKSDTISIDMSSPVQIAEIIGTETQEVTDADTIKKLTD KIKTEKSALLQVWYASDKTGKQIDPADSASESIEVYIPSASA DEATSSIETTTTTSTETITTTEEK
Rf 152	[(1-26)SIGN][(27-33)UNK][(34-354)GH5_4][**(355-536)UNK**][(537-832)GH5_7][(833-864)LNK][(865-947)DOC1]	ECHGYLNRSNLTWYTESEPVVNKMMEVLGVSSSNPPTTT ASTPSGNDTTTTTAEEEDTAILYPFTISGNDRNGNFTINFKG TPNSTNNGCIGYSYNGDWEKIEWEGSCDGNGNLVVEVPM SKIPAGVTSGEIQIWWHSGDLKMTDYKAGSGSSQTNTTPQ QTTNNNNTTVTTAKNDQPQTAG
Rf 153	[(1-25)SIGN][(26-36)UNK][(37-455)GH5_1][**(456-634)UNK**][(635-708)DOC1]	MGRAGNGISLGEFYSSGAKSNMADVGGSSGPKTTTPVVT TASNVTTTTSAPKTTQESTTTTAEVTTPDVSTHQTYKFSKV ETYPTKTEYLLEEDFDPSGLKFSATPDKEYAGPFKASDFVIT DYKDEPVSASEFSKLPEGVYTVKMNETYGSYYTYDLCSGI DISFKVTVGSDDPGTKP
Rf 154	[(1-27)SIGN][(32-469)GH5_1][**(470-670)UNK**][(671-744)DOC1]	EGSNLDGGKGSTQGTTKPPVTAAPTSSVTTGPTQTTSRPT TTTTSDWVTTTTTSDWVTTTTTSDSSTSIATTTVTEPKATTT TAPVTTYPVFGTDHDFLPYSVKIITEPTKTKYQIGEKLDLSGI KVEYNDYVIGADSLKLLSEYKPGSGYPALFVNDYEFDSSKE GEYRIFISFSGRPESYPSDAFTVTVGEKKIDTPTT
Rf 155	[(1-22)SIGN][(23-40)UNK][(41-449)GH9][**(473-640)UNK**][(641-808)UNK][(809-843)LNK][(844-925)DOC1]	QPTTQPTTQTPTVTTQTPSSSSDGYTIKPNKKVTYSALGED ERMIGFSYKDFGISSSEKITEVQVNISANKNIGKYVGQFGTS TTDSANGYWAMGDEITQSISGNSGTITWKVPSDISSIIQTQY GGEIKFGVWWIDCDEFTIDSVVLKTSGGSSSNITTQRTTQT TT
Rf 156	[(1-33)SIGN][(34-46)UNK][(47-194)CBM4][**(195-359)UNK**][(360-917)GH9][(918-964)LNK][(965-1045)DOC1]	KLRDNTEGAICETVEPDLGIVRPKSNVRINQHGYSSKLTKK ASYCTDNKEPCQFELRDSSGKAVYTGTASAVVEDRSAGNT ATEKTPFGQKLKDSGKYVQILDFSDFTEAGEYTIFVKDTVG VSDTAYFGHEGFYDTSLDGDKLMWNEGRNSYCMNESAAF TIKE
Rf 157	[(1-24)SIGN][(25-45)UNK][(46-193)CBM4][**(194-361)UNK**][(362-924)GH9_1][(925-949)LNK][(950-1021)DOC1]	LSLVDNTSDEGDFDTTNEFGVVRPRSNVRINQVGYYPNLE KRASYCTDNSSPCEFEVRDSSGKAVYTGTASKVVEDPDS GNGETTETKYGKKIKDSGKYVQILDFSKLTTPGEYTIFVKDS VGVSDTACFNLKGAWDTKASGDKLEWTNWKTGKVYTMN ESMKFRIDE
Rf 158	[(1-19)SIGN][(20-30)UNK][(31-530)GH44][**(531-702)UNK**][(703-726)LNK][(727-800)DOC1]	DAAAFDGEKIYEEKKVTEKTEEFKDPSSMINKNGYVEIPITD PEHLSKIVINGDVTSSAGSGWATAGCAVCINAKDKAGKDF WTYKSYSLPLGKGQSAIVKFDGTLTKTTGEGEDKVSEDLE AYVADGKVELQKWWDASEKGDPDDATKDKIEVEYTSIQVV YEYAEGDEPQ
Rf 159	[(1-24)SIGN][(25-59)UNK][(60-295)GH124][**(296-441)UNK**][(442-516)DOC1]	GDTEPSMAPIEEKPQTTEPVTTTAETTTSTTSGVASTTTAA SSETTTSATTYIPTYKLTVETPPDTTVYLIGQKLDLTGMTMK FSTDTGSSMSTDIILESEAVNDNNAFQHEYMVHDYGINEAG EYTFIVSMISDESYTTSFTVKY
Rf 160	[(1-27)SIGN][**(28-206)UNK**][(207-261)DOC1_dist]	EDTTAPETLVYRTSVNTTTQTGWDFSAAMKCYDNGHIHMD IQCEGYTSWSSNIKKIGSVIINDSMISEIDTSGYAGTSSFDRN GFSYSGYTLNYNEDTLVYDLMYASRDSDLLGKTLMKLDFY VKEDYLKTNQIITVFDKEIEIPFGEPYANPYDTINSLNLEINEL TRQIAELKQNNG
Rf 161	[(1-21)SIGN][(23-618)X134][**(619-826)UNK**][(827-890)DOC1]	TDGMISGYGFDSECSIWSQDRHAATDAILCGNAQWTKERT SAKGVVSLDGIGSYIELDNSVVKTDDLQISLGILWKGGSKV QDVFFAGDEKAYMRLTPSNENGVAEFTITDGKTTQSLTAD KALEKGTWSQVSVRIIDGKGELIINNKSAASADVSLTPLNVL

		SASENDNAYIGKSSIASDFNGAVDYCNFWFKPADEPDMKY SGKEE
Ct 162	[(1-25)SIGN][(26-145)UNK][(146-215)DOC1]**[(216-600)SERPIN**]	EKNEGVSENFIRGNSNFAFNIFKEINKDEQGKNVFISPFGIS TALSMVYQGAKSDTREEMAKVLGYEGLDIEEVNKSYKYLL QYFNGLDNDTKIKSSNSIWMNSLHGNAIKEDFISTNKDVFD ALAETRDFSDKGVVDEINDWISKATEGQIDKMLSEIDMDML AYIISALYFKGTWTEEFDIEKTVSVPFASEDGGADHVMMMR KELCTIEFGEGDGYKAVRLPYGDGEMAMYCILPDEDTSIND FIQKLDLSMWEKIKNSITKRENGTIYLPRFKMEYAKGESGSI MESLKALGMKKAFEEDADLSGMTEADAFISDVLHKAVVEV NEKGTEASGVVVIPIAPTSIAPGPKFIANRPFAFVIADEKYDT ILFMGKLCDGGLIN
163 Ct	[(1-19)SIGN][**(20-125)UNK**][(126-197)DOC1][(198-599)SERPIN]	ANWYTYYELYPDPVVTDTSAVIRFKLTTEKIDDYITMPGYIY EFRYWKVDEPSKVKSVSIGLYVNEVQTVELTDLEPNTEYEC KVWGQIYTSNKEGTPKTITFKTL
Ct 164	[(1-19)SIGN][(20-125)UNK][(126-197)DOC1][**(198-599)SERPIN**]	TDEGINDGFNDETDEDINDSFIEANSKFAFDIFKQISKDEQG KNVFISPFGISMALSMVYQGAESDTREEMAKVLGYEGLDIE EVNKSYKLLLKYFNELIGNVKLKNSNSIWKNSLKGDVIKEDF ISVNKDVFNALVETRDFSDESVVDKINNWISDATEGKVKKA LNAVNPDELLYIISALYFNGAWKEEFEFDINDTTMSTFKSED GSTDYVMMMRKSYNNWVGVMEFGKGDGYSAIRLPYGNG EMAMYCILPDEDISINDFIQNLDVSLWNEIKNSIRKTLQGLIC LPRFKIEYFKDGNGSIKESLKALGLEKVFSLAEADLTGMSET NAYVSDVLHKAVVEVNEKGTEASSSVVVIPVPGFGTRSEFI ADRPFVFIIADEKYDTILFMGKLAKGELIN
Ct 165	[(1-30)SIGN][**(31-176)X140**][(185-330)X140][(331-490)UNK][(491-973)X139][(974-1051)DOC1]	SQTLFINEVMSSNVFTIRDGDVTDPKHGSKGGAYSDWIEIY NAGPYDVDLTGYILADSSAEWVFPQGIVPAGGYLLVWASD KNMVAQDGQLHTNFKLSASGENITLKKPDGTIVDSVDIIGLG DDQSYGRKSDGASEFVVFINPTP
Ct 166	[(1-30)SIGN][(31-176)X140][(185-330)X140]**[(331-490)UNK**][(491-973)X139][(974-1051)DOC1]	PGTANIYDPSLIPVSEPVFSHQGGFYTGAFKLELTTNEPGV KIYYTTDGSDPVPGKSGTIEYTSGINIKSRKGEANVLSMIQDI SNDQWNRWRAPNGEVFKCTTIKAVAIRDDGARSKVVTHSY FVDPQMNTRYTLPVISIVTDYDNFFDKSTGIYVNGN
Rf 167	[(1-25)SIGN][**(26-164)UNK**][(165-456)PL1][(457-596)UNK][(597-1022)PL9_1][(1023-1062)LNK][(1063-1175)X215][(1176-1206)LNK][(1207-1278)DOC1]	GTLSGTFSFLTGGAPITAFAAESSVKFISTVGYGEGMYAMW SSVSGASGYNVYVDGTQIDSMLIRQYSGYMRADAVGLKAG SHTMKVVPVVGGKEDASKAAETKATAYAHDRSGFGFVNG DSSGAYNKDGTLKTGATVV
Rf 168	[(1-25)SIGN][(26-232)UNK][(233-253)X159][**(254-448)UNK**][(449-468)X159][(471-490)X159][(493-513)X159][(515-535)X159][(537-557)X159][(558-732)UNK][(733-803)DOC1]	ADALPDDPLFDYPKEKVVDNLSEVYVSDTEVYTFDPETEW CYKLFANNSKVEVTLVYAPDLQGPPPSEYMGYPLNIELYPS PPYGIPVITIPEGTEDLSGLDFFEPQKIRQIIVKSKHLYIFPQA FAHSVITDLNFPGTIKIGGSAFQFCNKLQRVNFNGTDPIIHIE FGAFWYDEALRDLLFPDSVAFLTIEND
Rf 169	[(1-24)SIGN][(25-229)UNK][(230-249)X159][(252-272)X159][(275-295)X159][**(296-495)UNK**][(496-516)X159][(519-539)X159][(542-561)X159][(564-583)X159][(586-606)X159][(608-628)X159][(629-700)UNK][(701-761)DOC1]	GNVFDSAKAETVFLNIKEPSSWLYGGFAEVKNFSFGDNTEI LTMDYMIDFSNLGIPESVSVIKALYNMNFDEVNELTIPKNIKI FGAYREPKGVIVAEWGKIPASVPLKDNFTIKGYKGTEAEIYA KELNIPFIALDDLETPLSGNYSENIKWTLDADGVLTLSGEGEI PDLTESAPWSSRRADIQTIIVEDGITSIGKD
Rf 170	[(1-34)SIGN][**(35-181)UNK**][(182-262)DOC1]	EPSQEDFKAMAEEMVILVNEAREAEGLKPLYMVPYLCDVA NVRCRESIFSFSHNRPDGSSFITVLDDSLVPYSKAAENLAA GSDTAEATFNQWKNSPNHWRSIMNPDYTHIGVAVSYDMN SEYKYYWEQFFVAVSGKLNDQTIPERY
Rf 171	[(1-24)SIGN][**(25-166)UNK**][(167-258)DOC1]	ADSELVYDDFNKNDVNGEVGFVLPKGATGTFTITFDSPEG KDIPYYTGSIESGKDYAFKLEGRDNTKDDFRTYTLSVEITGG DFGRTSAAFTDTINGTGDDTFIIHDPHDNPDTYQKCVYKFTI DDKDTGNPWDVTASDATS
Rf 172	[(1-20)SIGN]**[(21-216)UNK**][(217-297)DOC1][(298-727)UNK]	VPVNSFAASAEKTASEQTSEKGSVLQAYADEISDLMLRSRI AGYAFVRDDKVVITYSAYEDKIKTYIDEKGYDKSIFVYEAGK NVATEKESALKPVAEDINEYMRSNGISGFTYVQEVDGVEKI FISYDGFFDKIKAYVTGKGIDENIVVYQQSHDDVIVSAAVSG VINNDGTFNIADVDISGVMDLSEFELKETV
Rf 173	[(1-21)SIGN][(24-125)DOC1][**(126-310)UNK**]	LEQYLYEAAHEGAASPGPENPDQGEGDVSLNDSRYFFWL DLSKDKNFVFNDQFIKVTFKIKEDAPDNDYTIKLKPDLSDIG GVSISPDKVIDGTVRVGGESIDPADVSAEEGLTVYGDNVAC KQGDTIDYYINIKNNDGLAAALVWFYFDSNAMEFEFAEEAG EFADMITDGTVKVGARSGKSQK
Rf 174	[(1-21)SIGN][**(22-207)UNK**][(208-228)X159][(229-307)UNK][(308-328)X159][(329-411)UNK][(412-432)X159][(433-556)UNK][(557-577)X159][(578-747)UNK][(748-	CVAGGMRTAATTTNANTKNSGIVITASAADTEVIKEGKNAS KYTYKYADATIYSTSFQALTIPVEELNEKTTSSWEDKETSKI YKTDTKWTDSKTGNSYSIYKITIKENKGNKSTVSTEYRIGLA TLKPTTIDFKLTDDIAYPVDEKEVKAYLSEKNKDFTGISSKKA TIIGPSALASSYVKTVDF

	768)X159][(769-838)UNK][(839-858)X159][(859-973)UNK][(974-1020)LNK][(1021-1105)DOC1]	
Rf 175	[(1-30)SIGN][(31-553)UNK][(554-620)DOC1][**(621-803)UNK**]	SEAPKPVMVKISQNFAWFPYQVVDVYDQYGTVYVHSYSS NTEHSWDEIQEKTVNMRNDNWYKVIRNIMDTAVEDSEDQV KSQPEFCVKRRFNNDRLISDFAAFSKNTEKYSKTKMFTVNY ACDMGSTTIYSIGETADGKPVCAELATFGDAVGYINDTEVK VFIQKLVDNHMISPAIFNVVNQ
Rf 176	[(1-28)SIGN][(29-498)UNK][(499-562)DOC1][**(563-713)UNK**]	KNETPPPFNEAEATQTCNMLNIDHDESYDNITATALQPTYA TNTDEIICEVKNHNKGKGFFYFPIAFIERYDNEKWEEIYNAD PAIRYQYGDGYVLCGIADHEHDDMEFSTWIRVQTNNIFPSL KEGHYRFKIYTAKNILYAEFDVVAAEQ
Rf 177	[(1-22)SIGN][(23-399)UNK][(400-420)LNK][(421-489)DOC1][**(490-624)UNK**]	KCETIMDWYSLGHSISDFYNKGYKPAVITNPDELKEYIEQAI PEENISSYLEKYNDSFFKDNVMLIYALQQTMGMNPAYAFVD TEVLDDQIKIYYKSAIQWGHPHPMFESICFAQVIVPKTAYNG QTVNWEFVES
Rf 178	[(1-28)SIGN][**(29-227)UNK**][(228-292)DOC1]	SENRDKVYETSLVLPEDIAYPLEALKSFLWDSETQRPSMDA FTELDDNKKIIIVADNEDVLAIVKAFVKEKGLDENLIGYRIESF ENMVPSGIADDHVYLHWDEMTIDQQYFMAESGDPVMTYS AAEKEVPADDVEDYIGRFYMSGYDWYEAIYYHCETDAYKIK GNDNADKIAVRFDGDEKYYLYTLQTYNADEEQLT
Rf 179	[(1-25)SIGN][(26-100)UNK][(101-121)X159][(124-143)X159][(146-166)X159][**(167-343)UNK**][(344-389)LNK][(390-616)UNK][(617-686)DOC1][(687-701)UNK]	GGKTSVTSNATIYAPSGSAAQAYANETGHSFAVLSGSQPT QPTTTRTTTTSSSTTVTTTTVKVDLQLIAVTTKKVETGVTTE YVAAQGYIGYTNRVLNNVEFDVEVQGNIQVSNKKMMYDSP VALKGAYVFAFDLYVPENVDGSYPYTLRVTRAYDMSNRDV TSQLPWTEYKGTIRV
Rf 180	[(1-22)SIGN][(23-40)UNK][(41-449)GH9][(473-640)UNK][**(641-808)UNK**][(809-843)LNK][(844-925)DOC1]	QRTTQPTTQRTTQTTQTPSSSSDGFSIKPNQKVTYSALGE DERMIGFAYKDFGISSSDKITEVQVNISANKNIGKYVGQFGT STTDSANGYWAMGDEITQSISGNSGTITWKVPSDISSIIQTQ YGGEIKFGVWWIDCDEFTIDSVVLKTSGGSSSNITTQRTTQ TTT
Rf 181	[(1-22)SIGN][**(23-142)UNK**][(143-433)GH43][(434-587)CBM62][(588-796)UNK][(797-858)DOC1]	LPCITAHGRDAAAVISDEIGFVGNLFCISDNGSADHAELQW STTLSAKSYTLYRSTDKNSGYEPVYSGNGNSWQDNSMEM GKDYFYQLEVTTDKGTAYSEIRELTPCEVPGGLSKYDNQH
Rf 182	[(1-25)SIGN][(26-43)UNK][(44-438)GH43_B][**(439-569)UNK][(**570-718)CBM13][(719-856)CBM13][(857-900)LNK][(901-981)DOC1]	SDIVGDYEYINHGNSTDGKIIGYKKIKLNADGTISGDVSGTW TQDAASSAAVITIGGQKYSGYFMAAQNEKGTKVMSFTAVG SNNQTIWGAQNKAFTGKERDGGAADYTNSNDVLTFAPETI GDISSDLKI
Rf 183	[(1-38)SIGN][(39-318)UNK][(319-725)PL1_2][(726-773)X149][**(788-931)X157**][(932-995)DOC1]	GRLVSVDVLDTAYYSSWGIDNTLAVGDSLFGDRTAEKCAV SELPDKLSGAELVLTPCDAKASSKDQAELTIAQDCTVNVGL DSRVENVPAWMSDFTKTNSVIKTTNDVTFELYAKPVKAGE VVKLGSNEQSASCMNYIVIASEK
Rf 184	[(1-36)SIGN][(37-41)UNK][(42-406)CE8][(407-520)UNK][(521-797)PL1][(798-891)UNK][**(892-1037)X157**][(1038-1070)LNK][(1071-1154)DOC1]	LIASLDVKDTTYGSAWSLDKNTANGSKAFGDRDFTITALPG GLAGAEHIITACDSKKTDADLAVLTAAKDITVYVAMDQRNVT LPAWLGSFTKTGDILGITDAEGEKPFDVYAIDLAAGQSLTLG TNGMMGNVMGYTAFVKEKEIV

Supplemental information – Chapter 4

Figure S4.1| Schematic representation of the CEL synthesized gene.



The cohesins cassette was cloned between two His_6 tags. Dockerin was cloned downstream to a T7 promoter, after the cohesins cassette. The CEL gene was cloned downstream to the T7 promoter of the pHTP plasmid. Coh, cohesin; CBM, carbohydrate-binding module; Dock, dockerin.

Table S4.1| Molecular weights (kDa) of the His6-tagged proteins (A, B, C, D, E, F, G, and H) fused with Trx, GST, MBP, NusA, SUMO, DsbA, DsbC, Fh8, CEL, Rf1 and Rf47 system tags.

	Proteins in test							
Vectors	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
pHTP	30.25	63.73	25.35	12.37	18.32	70.10	28.51	38.02
pHTP-TRX	42.68	76.16	37.78	24.80	30.75	82.53	40.94	50.45
pHTP-GST	56.57	90.05	51.67	38.69	44.64	96.42	54.83	64.34
pHTP-MBP	73.11	106.59	68.21	55.23	61.18	112.96	71.37	80.88
pHTP-NusA	85.67	119.15	80.77	67.79	73.74	125.52	83.93	93.44
pHTP-SUMO	42.25	75.73	37.35	24.37	30.32	82.10	40.51	50.02
pHTP-DsbA	53.97	87.45	49.07	36.09	42.04	93.82	52.23	61.74
pHTP-DsbC	56.51	89.99	51.61	38.63	44.58	96.36	54.77	64.28
pHTP-Fh8	38.39	71.87	33.49	20.51	26.46	78.24	36.65	46.16
pHTP-CEL	38.92	72.40	34.02	21.04	26.99	78.77	37.18	46.69
pHTP-Rf1	60.44	93.92	55.54	42.56	48.51	100.29	58.70	68.21
pHTP-Rf47	58.74	92.22	53.84	40.86	46.81	98.59	57.00	66.51

Table S4.2| Molecular weight ratio (%) between recombinant proteins alone and total fusion proteins. Levels of relative recombinant protein were estimated by dividing the molecular mass of the recombinant protein by the molecular mass of the fusion, full-length protein.

	Proteins in test							
Vectors	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
pHTP:	94.2	97.3	93.1	85.8	90.4	97.5	93.9	95.4
pHTP-TRX	66.8	81.4	62.5	42.8	53.9	82.8	65.4	71.9
pHTP-GST	50.4	68.8	45.7	27.4	37.1	70.9	48.8	56.4
pHTP-MBP	39.0	58.1	34.6	19.2	27.1	60.5	37.5	44.8
pHTP-NusA	33.3	52.0	29.2	15.7	22.5	54.5	31.9	38.8
pHTP-SUMO	67.4	81.8	63.2	43.6	54.6	83.2	66.1	72.5
pHTP-DsbA	52.8	70.9	48.1	29.4	39.4	72.8	51.2	58.7
pHTP-DsbC	50.4	68.9	45.7	27.5	37.2	70.9	48.9	56.4
pHTP-Fh8	74.2	86.2	70.5	51.8	62.6	87.4	73.0	78.6
pHTP-CEL	73.2	85.6	69.4	50.5	61.4	86.8	72.0	77.7
pHTP-Rf1	47.2	66.0	42.5	25.0	34.2	68.2	45.6	53.2
pHTP-Rf47	48.5	67.2	43.8	26.0	35.4	69.3	46.9	54.5


FCT
Fundação para a Ciência e a Tecnologia


CIISA


nzytech
genes & enzymes


PoR
LISBOA


QREN
QUADRO
DE REFERÊNCIA
ESTRATÉGICO
NACIONAL


UNIÃO EUROPEIA
Fundo Europeu de
Desenvolvimento Regional