UNIVERSIDADE TÉCNICA DE LISBOA

Faculdade de Medicina Veterinária

Molecular mechanisms of plant cell wall hydrolysis: developing novel biotechnological applications for carbohydrate-binding modules

Catarina Isabel Proença Duarte Guerreiro

Tese de Doutoramento em Ciências Veterinárias

CONSTITUIÇÃO DO JÚRI

PRESIDENTE

Reitor da Universidade Técnica de Lisboa

VOGAIS

Doutor Luís Manuel dos Anjos Ferreira

Doutora Isabel Maria Godinho de Sá Nogueira

Doutor Francisco Miguel Portela da Gama

Doutor Carlos Mendes Godinho de Andrade Fontes

Doutora Lucília Maria Alves Ribeiro Domingues

ORIENTADOR

Doutor Carlos Mendes Godinho de Andrade Fontes

CO-ORIENTADORES

Doutor Francisco Miguel Portela da Gama

Doutora Lucília Maria Alves Ribeiro Domingues

2008

LISBOA

Declaração

Nome: Catarina Isabel Proença Duarte Guerreiro
Endereço electrónico: catarina@fmv.utl.pt Telefone: +351 966050777
Número do Bilhete de Identidade: 10524074

Título: Molecular mechanisms of plant cell wall hydrolysis: developing novel biotechnological applications for carbohydrate-binding modules

Orientador: Doutor Carlos Mendes Godinho de Andrade Fontes
Co-orientadores: Doutor Francisco Miguel Portela da Gama e Doutora Lucília Maria Alves Ribeiro Domingues
Ano de conclusão: 2008
Designação do ramo de conhecimento do Doutoramento: Ciências Veterinárias

Faculdade de Medicina Veterinária da UTL, 9 de Setembro de 2008.

Assinatura: Catarina Isabel Proença Duarte Guerreiro

Aos meus Pais, Maria Helena e António

Ao meu irmão Sérgio, à Vanda e ao Francisco

Acknowledgements/Agradecimentos

Esta tese representa um culminar de muito trabalho e muito tempo, com altos e baixos, que só foi possível graças ao apoio de todos aqueles que vou referir e, sem os quais, tenho a certeza, não teria sido possível chegar aqui.
Assim, quero agradecer,

à Faculdade de Medicina Veterinária da Universidade Técnica de Lisboa, a minha escola-mãe, por me ter proporcionado meios humanos e materiais para a execução da maior parte do trabalho que originou esta tese;

ao Departamento de Engenharia Biológica da Universidade do Minho, por também me ter proporcionado meios humanos e materiais para a execução de uma parte do trabalho para esta tese e pela forma hospitaleira como me recebeu;

ao Professor Doutor Carlos Fontes, meu orientador científico, por toda a sua constante disponibilidade, por ter estado sempre presente, pelo seu apoio, por tudo o que me ensinou e, não menos importante, pela amizade demonstrada ao longo de todos estes anos; todas as palavras são poucas para expressar o meu profundo agradecimento;

aos Professores Doutores Miguel Gama e Lucília Domingues, meus co-orientadores científicos, pela forma hospitaleira como me receberam no departamento de Engenharia Biológica da Universidade do Minho, pela disponibilização de todos os meios para a execução da fusão dos péptidos antimicrobianos ao CBM3, pelas suas sugestões e pelo acompanhamento científico dessa parte desta tese;

ao Professor Doutor José António Mestre Prates pela disponibilização do laboratório de Bioquímica para as purificações e outras técnicas necessárias aos estudos estruturais das proteínas e pelos seus conselhos técnicos de elevado valor; ao Professor Doutor Luís Ferreira que teve, desde o início da minha carreira científica, um papel preponderante na tomada deste meu rumo profissional;

à Professora Doutora Maria João Romão, à Doutora Ana Luísa Carvalho e ao Doutor Shabir Najmudin, da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, por terem colaborado na resolução da estrutura do CBM30 e do CBM44;

aos Professores Doutores João Bengala Freire, Luísa Falcão e Madalena Lordelo, do Instituto Superior de Agronomia, por me terem disponibilizado recursos para a execução dos ensaios com os frangos de carne;

to Professor Harry Gilbert, from the University of Newcastle upon Tyne, UK, and his team; to Professor Gideon Davies, from the University of York, UK, and his team; to Professor Harry Brumer, from the Royal Institute of Technology, Sweden, and his team; to all the other co-authors of the papers in which this thesis is based;

to Professor Arun Goyal for his contribution to the *Ct*Lic26A study and, above all, for his wise technical advices and friendship during all the years he was working in Lisbon;

à Benedita, à Cristina, ao Fernando, à Helena, à Joana, à Márcia, à Marisa, à Patrícia, à Susana, à Paula, à Teresa, à Virgínia e à Zé, pela amizade e pelo bom ambiente de trabalho que me proporcionaram nos laboratórios de nutrição e de bioquímica da FMV; a vossa boa disposição contagiou-me muitas vezes; à Patrícia um agradecimento muito especial pela sua amizade, por ter sido tão boa companheira de escrita e pelo auxílio na análise estatística dos resultados dos ensaios com os frangos de carne; à Helena, um agradecimento também especial por toda a ajuda nos trabalhos de laboratório;

à Carla, à Catarina G., à Joana A., à Joana C., à Orquídea, ao Pedro, ao Reinaldo, ao Ricardo, à Susana M., à Susana A., à Vera e ao Wagner, pela amizade e pelo grande apoio que demonstraram nos laboratórios do departamento de Engenharia Biológica da UM, e sem o qual me teria sido difícil terminar algumas fases do meu trabalho da fusão dos péptidos antimicrobianos ao CBM3;

a todos os meus amigos extra-trabalho, aos meus primos, aos meus tios, avós e demais família; obrigada pelo vosso apoio nos momentos mais difíceis destes anos; uma referência especial à minha avó Lucília que eu gostava que ainda pudesse estar presente;

ao meu irmão Sérgio, pelo seu afecto, pela sua amizade e pelo seu apoio constante; obrigada irmão; à Vanda, pelo seu enorme carinho; ao Francisco que, com a sua inocência, me vai ensinando a ser melhor pessoa e a descobrir um mundo melhor;

por fim, aos meus Pais, pelo seu amor, por terem contribuído para eu ser o que sou hoje e por terem estado sempre ao meu lado nesta caminhada.

A todos Muito Obrigada por me terem ajudado nesta travessia, quer pessoal quer profissional. Foi graças a todos vós que conseguir chegar aqui.

This work was funded by Fundação para a Ciência e a Tecnologia,
grant SFRH/BD/16731/2004, and co-funded by POCI 2010 and FSE from Ministério da Ciência, Tecnologia e Ensino Superior

ABSTRACT

Molecular mechanisms of plant cell wall hydrolysis: developing novel biotechnological applications for carbohydrate-binding modules

Xyloglucan (XG) is the major plant cell wall hemicellulose. Biochemical and structural studies on a xyloglucanase, Xgh74A, from *Clostridium thermocellum* are presented here. In addition, the carbohydrate-binding modules (CBMs) from *C. thermocellum* Cel9D-Cel44A, *Ct*CBM30 and *Ct*CBM44, which recognize XG, are characterized in this work. *Ct*Cel9D-Cel44A's C-terminal module of unknown function is a CBM, constituting the founder member of family 44. Structural studies revealed that both *Ct*CBM30 and *Ct*CBM44 present Type B binding site topologies where tryptophans play an important role in ligand recognition. Cel44A is an endoglucanase domain displaying some xylanase activity, which action is potentiated by *Ct*CBM44 through a targeting effect. Biochemical and structural studies on *Ct*Lic26A-Cel5E are also described here. *Ct*Lic26A is a mixed β-1,4-β-1,3-glucanase and Cel5E an endo-β-1,4-glucanase. The three-dimensional structure of *Ct*Lic26A provides insights in the mechanism of lichenan recognition by family 26 glycoside hydrolases. Finally, novel biotechnological applications for CBMs were investigated. An experiment was conducted where a barley-based diet for broilers was supplemented with *Ct*Lic26A-Cel5 derivatives, with or without *Ct*CBM11, and with a commercial enzyme. The fusion of four antimicrobial peptides with CipA from *Ct*CBM3 originated recombinants with high affinity for crystalline cellulose, indicating CBMs could fix bioactive molecules to cellulose.

Key words: *Clostridium thermocellum*; glycoside hydrolases; carbohydrate-binding modules; poultry; antimicrobial peptides; biotechnology

RESUMO

Mecanismos moleculares de hidrólise da parede celular vegetal: desenvolvimento de novas aplicações biotecnológicas para os módulos de ligação a carbohidratos

O xiloglucano é a principal hemicelulose das paredes celulares vegetais. Neste trabalho apresentam-se estudos bioquímicos e estruturais sobre a xiloglucanase Xgh74A do *Clostridium thermocellum*. São também estudados módulos de ligação a carbohidratos (CBMs) com afinidade para o xiloglucano, o *Ct*CBM30 e o *Ct*CBM44 da *Ct*Cel9D-Cel44A. O módulo C-terminal revelou-se um CBM, sendo o membro fundador da família 44. As estruturas cristalográficas desses CBMs mostram locais de ligação aos polissacáridos com topologia do Tipo B, nos quais os triptofanos representam um papel importante no reconhecimento dos ligandos. A Cel44A é uma endoglucanase com actividade xilanásica, cuja acção é potenciada pelo *Ct*CBM44. Estudos bioquímicos e estruturais sobre a *Ct*Lic26A-Cel5E são também apresentados, sendo a *Ct*Lic26A uma β-1,4-β-1,3-glucanase e a Cel5E uma endo-β-1,4-glucanase. A estrutura tridimensional da *Ct*Lic26A revelou os mecanismos estruturais que modulam a especificidade da enzima. Neste trabalho são pesquisadas novas aplicações biotecnológicas para os CBMs. Efectuou-se um ensaio com frangos de carne, suplementando dietas à base de cevada com derivados recombinantes da *Ct*Lic26A-Cel5, com e sem *Ct*CBM11, e com uma enzima comercial. A fusão de quatro péptidos antimicrobianos com a CipA do *Ct*CBM3 originou recombinantes com elevada afinidade para celulose cristalina, indicando que os CBMs poderão fixar moléculas bioactivas a materiais celulósicos.

Palavras chave: *Clostridium thermocellum*; glicósido hidrolases; módulos de ligação a carbohidratos; frangos de carne; péptidos antimicrobianos; biotecnologia

PAPERS

This thesis is based on the following publications:

Taylor, E.J., Goyal, A., Guerreiro, C.I.P.D., Prates, J.A.M., Money, V.A., Ferry, N., Morland, C., Planas, A., Macdonald, J.A., Stick, R.V., Gilbert, H.J., Fontes, C.M.G.A., and Davies, G.J. (2005) How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides. Structure and activity of a *Clostridium thermocellum* lichenase *Ct*Lic26A. *J. Biol. Chem.* 280, 32761-32767.

Najmudin, S., Guerreiro, C.I.P.D., Ferreira, L.M.A., Romão, M.J., Fontes, C.M.G.A., and Prates, J.A.M. (2005) Overexpression, purification and crystallization of the two C-terminal domains of the bifunctional cellulase *Ct*Cel9D-Cel44A from *Clostridium thermocellum*. *Acta Cryst.* F61, 1043-1045.

Guerreiro, C.I.P.D., Martinez-Fleites, C., Baumann, M.J., Taylor, E.J., Prates, J.A.M., Ferreira, L.M.A., Fontes, C.M.C.A., Brumer, H., and Davies, G.J. (2006) Crystal structures of *Clostridium thermocellum* xyloglucanase, Xgh74A, reveal the structural basis for xyloglucan recognition and degradation. *J. Biol. Chem.* 281, 24922-24933.

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

Guerreiro, C.I.P.D., Ribeiro, T., Ponte, P.I.P., Lordelo, M.M.S., Falcão, L., Freire, J.P.B., Ferreira, L.M.A., Prates, J.A.M., and Fontes, C.M.G.A. (2008) Role of a family 11 Carbohydrate Binding Module (CBM) in the function of a recombinant cellulase used to supplement a barley-based diet for broiler chicken. *Br. Poult. Sci*. 49, 446-454.

Guerreiro, C.I.P.D., Fontes, C.M.G.A., Gama, M., and Domingues, L. (2008) *Escherichia coli* expression and purification of four antimicrobial peptides fused to a family 3 Carbohydrate-Binding Module (CBM) from *Clostridium thermocellum. Protein Expr. Purif.* 59, 161-168.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS AND SYMBOLS

A	Alanine
α	Alpha
Å	Ångstrom
AGE	Affinity gel electrophoresis
Ala	Alanine
AMPs	Antimicrobial peptides
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid/Aspartate
β	Beta
BSA	Bovine serum albumine
***Bsp*CBM28**	CBM28 from *Bacillus* spp
BW	Body weight
C	Carbon
ºC	Degree Celsius
C*α*	Alpha carbon
CAZy	Carbohydrate-active enzyme
CAPS	3-(Cyclohexylamino)propanesulfonic acid
CBD	Cellulose-binding domain
CBM	Carbohydrate-binding module
CBM30	Carbohydrate-binding module family 30
CBM44	Carbohydrate-binding module family 44
***Cc*CBM17**	*Clostridium cellulovorans* CBM17
CCP4	Collaborative computational project number 4
***Cf*CBM4-1**	*Cellulomonas fimi* CBM4-1
cm	Centimetre
CMC	Carboxymethylcellulose
Co	Cobalt
cps	Centipoise
Ct	*Clostridium thermocellum*
***Ct*CBM11**	*Clostridium thermocellum* CBM11
***Ct*Lic26A**	*Clostridium thermocellum* Lichenase 26A
Cu	Copper
D	Aspartic acid
Da	Dalton
DM	Dry matter
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E	Glutamic acid/Glutamate
EC	Enzyme comission
EGII	Endoglucanase II
ESI	Electrospray ionization
ESRF	European synchrotron radiation facility
Fe	Iron
Fuc	Fucose
γ	Gama

g	Gram
G	Unsubstituted glucose
Gal	Galactose
GH	Glycoside hydrolase
GH5	Glycoside hydrolase family 5
GH9	Glycoside hydrolase family 9
GH26	Glycoside hydrolase family 26
GH44	Glycoside hydrolase family 44
GH74	Glycoside hydrolase family 74
GI	Gastro-intestinal
Glc	β-1,4 D-Glucan
Glc-isoF	Glc-β-1,3-isofagomine
Gln	Glutamine
Glu	Glutamic acid/Glutamate
Gly	Glycine
h	Hour(s)
H	Hydrogen
HEC	Hydroxyethylcellulose
HEPES	4-(2-Hydroxyethyl)-1-piperazine-ethanesulfonic acid
His	Histidine
HPAEC-PAD	High performance anion-exchange chromatography with pulsed amperometric detection
HPLC	High-performance liquid chromatography
I	Iodine
Ile	Isoleucine
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl 1-thio-β-D-galactopyranoside
ITC	Isothermal titration calorimetry
IU	International units
IUB	International union of biochemistry
K_a	Equilibrium affinity constant
k_{cat}	Catalytic efficiency
kDa	KiloDalton
kg	Kilogram
K_m	Michaelis constant
kW	Kilowatt
L	Litre
LB	Luria-Bertani broth medium
Leu	Leucine
LK	Linker
Lys	Lysine
M	Molar (mol L^{-1})
MAD	Multi-wavelengh anomalous dispersion
µg	Microgram
µL	Microlitre
µm	Micrometre
µM	Micromolar
mA	Miliampere
ME	Metabolizable energy
MES	4-Morpholine-ethanesulfonic acid

Met	Methionine
mg	Miligram
min	Minute(s)
mL	Mililitre
mM	Milimolar
Mn	Manganese
MS	Mass spectrometry
MU	Methylumbelliferyl
nL	Nanolitre
nM	Nanomolar
NRC	National research council
NSPs	Non-soluble polysaccharides
º	Degree
O	Oxygen
OD	Optical density
OREX	Oligoxyloglucan reducing end-specific xyloglucanobiohydrolase
OXG-RCBH	Reducing end-specific cellobiohydrolase
PCR	Polymerase chain reaction
PDB	Protein data bank
***Pe*CBM29-2**	*Piromyces equii* CBM29-2
PEG	Polyethyleneglycol
pH	Potential of hydrogen
Phe	Phenylalanine
PKD	Polycystic kidney disease
Pro	Proline
RMSD	Root mean square deviation
Rov	Rovazime
rpm	Rotation per minute
s	Second(s)
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Se	Selenium
SEM	Standard error mean
SeMet	Seleno-methionine
Ser	Serine
SLH	S-layer homology module
sp	Species
Thr	Threonine
Tris	2-Amino-2-hydroxymethyl-1,3-propanediol
Trp	Tryptophan
Tyr	Tyrosine
U	Units
V	Volt
Val	Valine
v/v	Volume/volume
X	Glucose decorated α-1,6 with xylose
Xgh74A	*Clostridium thermocellum* xyloglucanase 74A
XG	Xyloglucan
XGOs	Xylogluco-oligosaccharides
X-gal	5-Bromo-4-chloro-3-indolyl β-D-galactoside
Xyl	α-1,6 D-Xylosyl

g	Terrestrial gravity constant (9,80665 m s^{-2})
Zn	Zinc
%	Percent

CHAPTER 1 General Introduction and Objectives

1.1. INTRODUCTION

Plant cell wall degradation is one of the most complex and important processes on earth. This biological mechanism involves the cooperative action of a large range of hydrolases, the majority being secreted by microorganisms like bacteria and fungi. Although the mechanisms by which these enzymes act have been extensively studied over the last decades, much is still unknown. Whether it is for use in animal nutrition or to develop and/or optimize novel biotechnological applications, it is crucial that we improve our knowledge on the processes involved in plant cell wall hydrolysis. This work aims to make a contribution in this field.

This general introduction begins with a summary review on plant cell wall composition, with particular focus in cellulose and xyloglucan. Subsequently, attention is driven into the cellulosome, an enzymatic machinery that is usually present in anaerobic bacteria. *Clostridium thermocellum*'s cellulosome is described and the complexity and functionality of this multi-enzyme complex is highlighted. In the following subchapters hydrolases and carbohydrate-binding modules are explored, since both play a crucial role in plant cell wall degradation. Finally, an approach to the development of biotechnological applications for CBMs is presented, where current applications are described and two potential novel ones introduced. The chapter finishes with the clear discrimination of what are the objectives for this work.

1.2. THE PLANT CELL WALL

The plant cell wall is a complex structure composed of cellulose microfibrils embedded in a matrix of diverse molecules of which the most important are hemicelluloses, pectin and lignin (Cosgrove, 2001). Cellulose is a polysaccharide of 1,4-β-D-glucose. Hemicelluloses are branched polysaccharides with a backbone of 1,4-

β-D-hexosyl or β-D-pentosyl residues. The predominant hemicellulose in dicots is xyloglucan while in monocots, such as cereals, xylan predominates. Hemicelluloses are sometimes called pentosans, since the majority of these structural carbohydrates are pentose-containing polysaccharides.

Plants present two types of cell walls that differ in function and composition: the primary and secondary cell walls. Primary walls surround growing and dividing plant cells, providing mechanical strength but allowing the cells to expand. In contrast, secondary walls are much thicker and stronger and are deposited only when cells have ceased to grow. In some upper plants, the secondary walls are strengthened by the incorporation of lignin, a phenolic polymer, which cements and anchors the cellulose microfibrils and other matrix polysaccharides and thus stiffens the walls, preventing biochemical degradation and physical damage (Knudsen, 1997).

Several models have been proposed to explain the organization of plant cell walls, which reflect the interactions established between its components. In the seventies, Keegstra *et al.* (1973) proposed that polymers from the matrix (xylan, xyloglucan, pectic polysaccharides and structural proteins) were covalently linked and formed a very large macromolecular network. In this model, cellulose fibres are connected through hydrogen-bonding to xyloglucan chains (Cosgrove, 2001). An alternative model was proposed by Hayashi (1989) and Fry (1989) that suggest that single xyloglucan chains fill the gap between cellulose microfibrils and tether them together; the pectin polysaccharides and the structural proteins occupy the space between xyloglucan chains. Although this is presently the most popular model, two other models have, more recently, been proposed: the multicoat model by Talbott and Ray (1992), in which each cellulose microfibril is coated with successively looser layers of matrix polysaccharides, and the stratified wall model by Ha *et al.* (1997), in which cellulose-xyloglucan layers are separated by pectic polysaccharides. All these models have in common the concept that cellulose fibrils are coated with xyloglucan (Cosgrove, 2001).

1.2.1. Cellulose

Cellulose is the major structural constituent of plant cell walls. Cellulose molecules are linear polymers of of D-glucose residues linked by β-1,4-glycosidic bonds, the structural repeat being the disaccharide cellobiose due to the 180º rotation of the glucose moieties inside the carbohydrate (Figure 1.1). Cellulose differs from other glucose containing polysaccharides due to its insoluble and rigid structure, which

originates a natural resistance to biological degradation. This insoluble and rigid structure results from the propensity of cellulose chains to pack together to form long crystals that are stabilized by several intermolecular forces (Linder and Teeri, 1997), mainly hydrogen bonds and van der Waals forces. Therefore, due to this mechanism, several cellulose chains aggregate to form microfibrils that can also aggregate to form macrofibrils. In higher plant cell walls, cellulose fibres are encapsulated by hemicellulose and lignin.

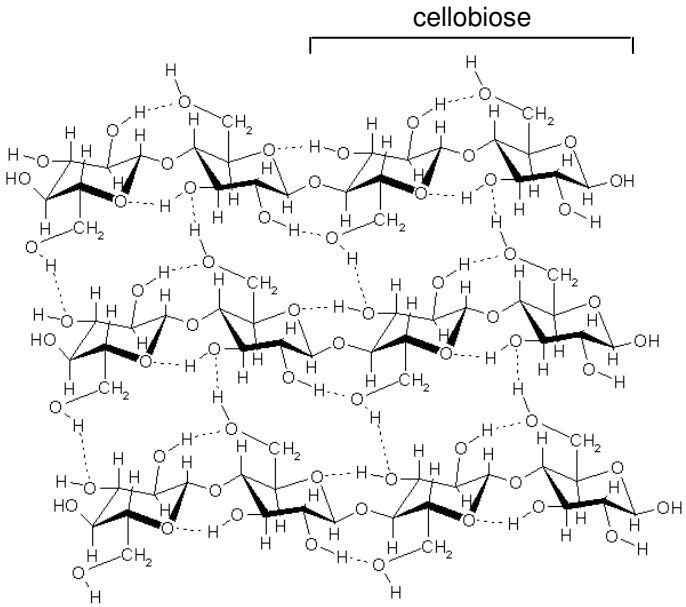

Figure 1.1. Structure of cellulose, highlighting the disaccharide cellobiose.
Adapted from http://www.doitpoms.ac.uk/tlplib/wood/structure_wood_pt1.php (April 2008).

In nature, cellulose chains have a degree of polymerization of approximately 10,000 glucopyranose units, in wood cellulose, and of 15,000 units in native cotton cellulose, while there is some evidence for a lower degree of polymerization in primary than in secondary cell walls (O´Sullivan, 1997). Natural crystalline cellulose is named cellulose I and comprises the forms Iα and Iβ, in which the chains lie parallel (Jamal *et al.*, 2004) although this cellulose structure does not seem to be the most stable form of the carbohydrate (O´Sullivan, 1997). Celluloses produced by primitive organisms, like the bacterium *Acetobacter xylinum*, present mostly the Iα component while those produced by higher plants have the Iβ form dominant (O´Sullivan, 1997). Some non-natural forms of cellulose and crystalline arrays of cello-oligosaccharides form cellulose

II, in which the chains lie in the anti-parallel orientation (Jamal *et al.*, 2004), that seems to be thermodynamically more favourable (O´Sullivan, 1997). In addition to crystalline regions, cellulose can also present areas that are not well structured, which constitute the amorphous regions of the cellulose molecule (the term amorphous not meaning without any structure). Cellulose can also present four other forms: III_I and III_{II} (formed from celluloses I and II, respectively) and IV_I and IV_{II} (from III_I and III_{II}, respectively).

1.2.2. Xyloglucan

Amongst hemicelluloses, xyloglucan is the predominant polysaccharide in dicots. It consists of a linear chain of *β*-1,4 D-glucan (Glc) regularly substituted with *α*-1,6 D-xylosyl (Xyl) units which, in a species-dependent manner, is further derivatized with *α*-L-arabinose or *β*-D-galactose (Vincken *et al.*, 1997; Carpita and McCann, 2000). Usually up to 75% of the residues are substituted with mono-, di- or triglycosyl side chains. As displayed in Figure 1.2, a single letter nomenclature was created to simplify the definition of the xyloglucan side-chain structures. For instance, a G represents an unbranched glucose residue and an X a glucose residue with a xylose 1,6-linked (Figure 1.2).

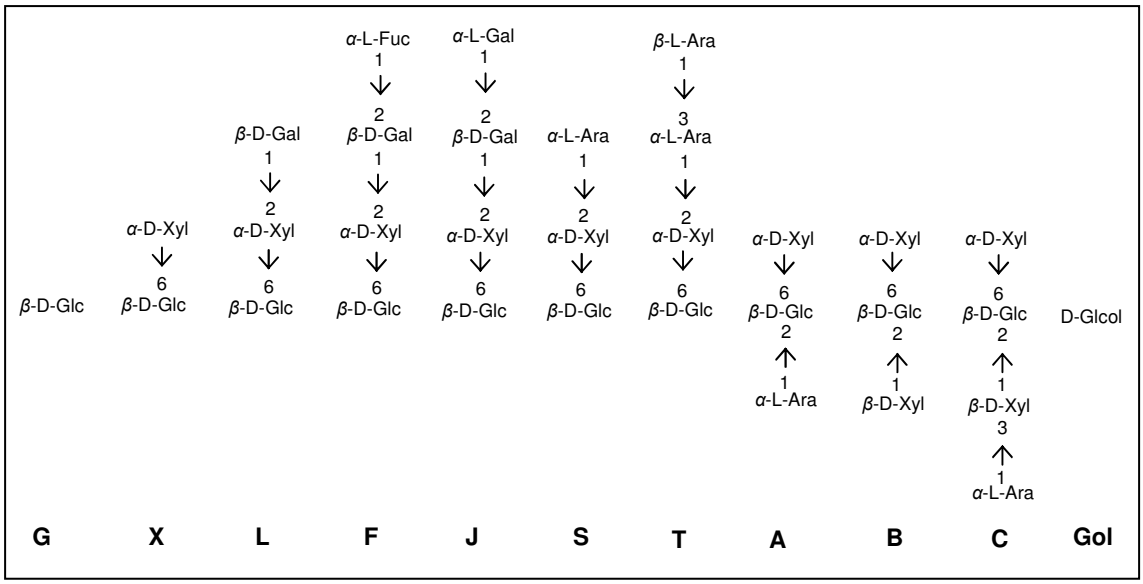

Figure 1.2. The one letter code to define xyloglucan structures.
Adapted from Fry *et al.* (1993).

Considering the type of decorations, xyloglucans are classified as XXXG-type or XXGG-type. Xyloglucans of the XXXG-type have three consecutive backbone Glc residues substituted with Xyl and a forth unbranched backbone sugar, Figure 1.3. In contrast, XXGG-type xyloglucans have two consecutive branched backbone residues followed by two unbranched backbone residues. It is well established that xyloglucan structures vary between plant species. However, primary walls of several dicotyledons, non-graminaceous monocotyledons and gymnosperms contain fucosylated xyloglucans with a XXXG-type structure. The glycosidic bond of the unbranched *β*-1,4-D-glucan residues in XXXG and XXGG xyloglucans are cleaved by *β*-1,4-xyloglucanases. Chapter 2 presents biochemical and structural studies on a typical xyloglucanase from the *Clostridium thermocellum* cellulosome, Xgh74A. In addition, in Chapter 3 the structure and function of two CBMs that bind xyloglucan is described.

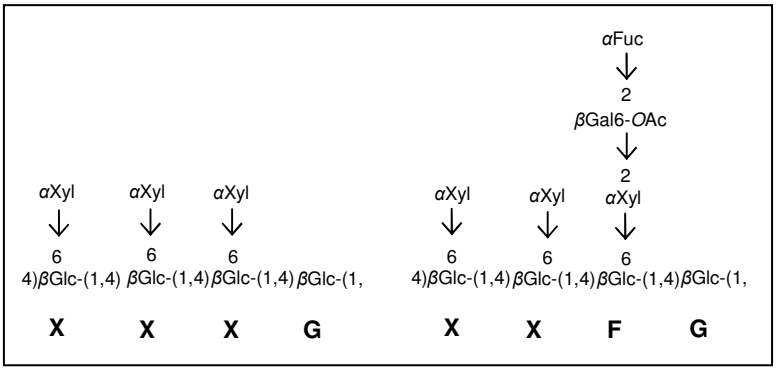

Figure 1.3. Schematic representation of XXXG-type xyloglucans.
Adapted from http://www.ccrc.uga.edu/~mao/xyloglc/Xtext.htm (April 2008).

1.3. THE CELLULOSOME

The cellulosome is a multi-enzyme extra-cellular complex of cellulases and hemicellulases produced by several anaerobic bacteria, such as *Clostridium, Acetivibrio, Bacteroides* and *Ruminococcus*, which is highly efficient for the degradation of cellulose, hemicelluloses and pectin. These microorganisms are located in soil, wood chip piles, sewage and animal rumen and their optimal growth temperatures can vary from mesophilic to thermophilic (Table 1.1). Cellulosomes may be the largest extracellular enzyme complexes found in nature, since their molecular masses range from 650,000 Da to 2.5 MDa while polycellulosomes have been reported to be as large as 100 MDa (Doi *et al.*, 2003).

Generally, enzymatic complexes from anaerobic microorganisms are much more elaborated and complex when compared with their aerobic counterparts where enzymes act individually during cell wall hydrolysis. It is possible that the anaerobic environment presents a greater selective pressure for the evolution of a highly efficient machinery (Bayer *et al.*, 2004). Aerobic cellulolytic fungi, for example, are able to produce and secrete copious amounts of free cellulases and hemicellulases which are directed to the surface of the plant cell wall in the environment in close proximity of the microorganism.

Table 1.1. Description of the anaerobic bacteria producing cellulosomes, their usual environmental niches and optimal growth temperatures.
Adapted from Doi *et al.* (2003).

Bacteria species	Source	Optimal growth temperatures	References
Acetivibrio cellulolyticus	sewage	mesophilic	Ding *et al.* (1999)
Bacteroides cellulosolvens	sewage	mesophilic	Ding *et al.* (2000)
Butyrivibrio fibrisolvens	rumen	mesophilic	Hespell and O'Bryan (1992)
Clostridium acetobutylicum	soil	mesophilic	Weyer and Rettger (1927)
Clostridium cellulovorans	wood fermenter	mesophilic	Sleat *et al.* (1984)
Clostridium cellobioparum	rumen	mesophilic	Lamed *et al.* (1987)
Clostridium cellulolyticum	compost	mesophilic	Pagès *et al.* (1997)
Clostridium josui	compost	mesophilic	Kakiuchi *et al.* (1998)
Clostridium papyrosolvens	paper mill	mesophilic	Pohlschröder *et al.* (1995)
Clostridium thermocellum	sewage soil	thermophilic	Bayer *et al.* (1985)
Ruminococcus albus	rumen	mesophilic	Ohara *et al.* (2000)
Ruminococcus flavefaciens	rumen	mesophilic	Ding *et al.* (2001)
Ruminococcus succinogenes	rumen	mesophilic	Miron *et al.* (2001)

When the first attempts to study bacterial anaerobic cellulases and hemicellulases were made, scientists found rather odd that those enzymes could not be found in the free state. Only later it was recognized that this phenomenon resulted from the association of the hydrolytic enzymes in cellulosomes that are usually attached to the bacterial surface. Studies indicating that *C. thermocellum* enzymes were probably associated in a multi-enzyme complex date back from the eighty decade of the last century (Garcia-Martinez *et al.*, 1980; Lamed and Zeikus, 1980). In these initial studies, Bayer *et al.* (1985) and Bayer and Lamed (1986) presented electron microscopy images showing the *C. thermocellum* cellulosome (Figure 1.4), suggesting that this multi-enzyme complex appears to be centralized on protuberant structures primarily located on the bacterial surface.

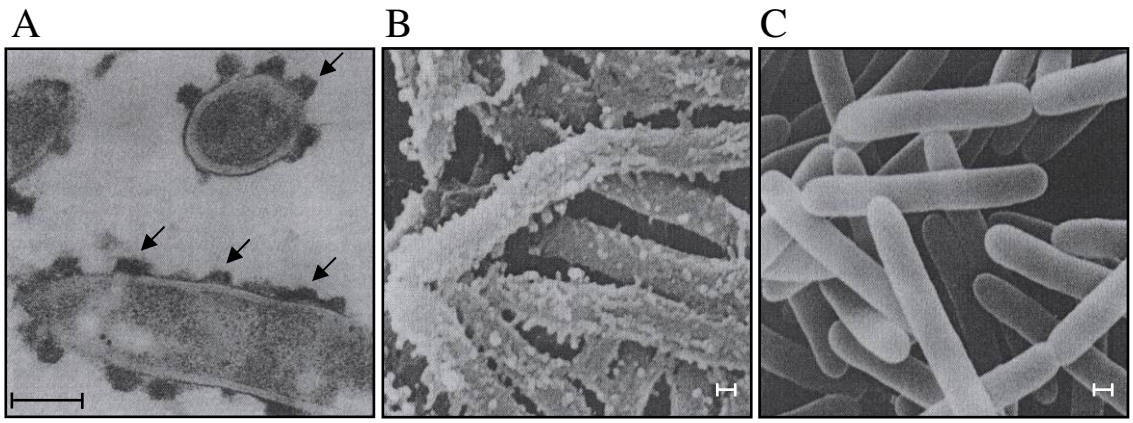

Figure 1.4. Images showing *Clostridium thermocellum*.
(A) Transmission electron microscopy image showing *C. thermocellum* ultrastructure; cellulosomes are the protuberant structures at the bacteria surfaces (arrows); bar, 250 nm; adapted from Bayer *et al.* (1985). (B) and (C) Scanning electron microscopy images depicting *C. thermocellum* wild-type strain YS (B) and *C. thermocellum* mutant AD2 with inability to adhere to cellulose (C); the protuberances at cell surface in the wild type are cellulosomes and do not appear in the mutants; bar, 200 nm; adapted from Bayer and Lamed (1986).

The initial definition of the cellulosome concept was based on studies on the cellulase system of *Clostridium thermocellum* (Bayer *et al.*, 1983), where adherence of the microorganism to the cellulose was proven. It was initially thought that the system included only cellulases, but presently the cellulosome concept has been broadened (Morag *et al.*, 1990; Mohand-Oussaid *et al.*, 1999). It is now well established that many of the known cellulosomes also include in their constitution different types of hemicellulases, like xylanases, mannanases, arabinofuranosidases, lichenases and pectin lyases (Bayer *et al.*, 2004). In Figure 1.5 it is presented a schematic representation of the *Clostridium thermocellum* cellulosome. There is also a simple schematic image of this cellulosome in Chapter 2 (Figure 2.2). The principal component of *C. thermocellum* cellulosome is a scaffoldin subunit, CipA, a large enzyme-integrating protein that contains Type I cohesin modules for incorporating the different cellulosomal catalytic sub-units. Enzymes contain a complementary module, named Type I dockerin, which binds to CipA Type I cohesin modules. These cohesin-dockerin interactions are named of Type I and mediate the integration of *Clostridium* enzymes into the complex, thus having a very important role in the selectivity and the stability of the cellulosome (Bayer *et al.*, 2004; Carvalho *et al*., 2003).

The gene encoding the scaffoldin from *Clostridium cellulovorans* cellulosome was the first to be cloned and sequenced (Shoseyov *et al.*, 1992). Scaffoldins from *C. thermocellum*, *C. cellulolyticum* and *C. josui* were subsequently identified. *C.*

thermocellum scaffoldin is unique amongst clostridia since it presents a C-terminal dockerin that mediates attachment of the multi-enzyme complex to the cell surface through a Type II cohesin-dockerin interaction. CipA C-terminal Type II dockerin binds to Type II cohesins from proteins that are attached to the cell surface by an S-layer homology module (SLH). Other bacteria, like *Acetivibrio cellulolyticus*, *Bacteroides cellulosolvens* and *Ruminococcus flavefaciens* also seem to have dockerin-containing scaffoldins and SLH-bearing anchoring proteins (Ding *et al.*, 1999; Ding *et al.*, 2000; Ding *et al.*, 2001; Rincon *et al.*, 2003; Xu *et al.*, 2003; Xu *et al.*, 2004).

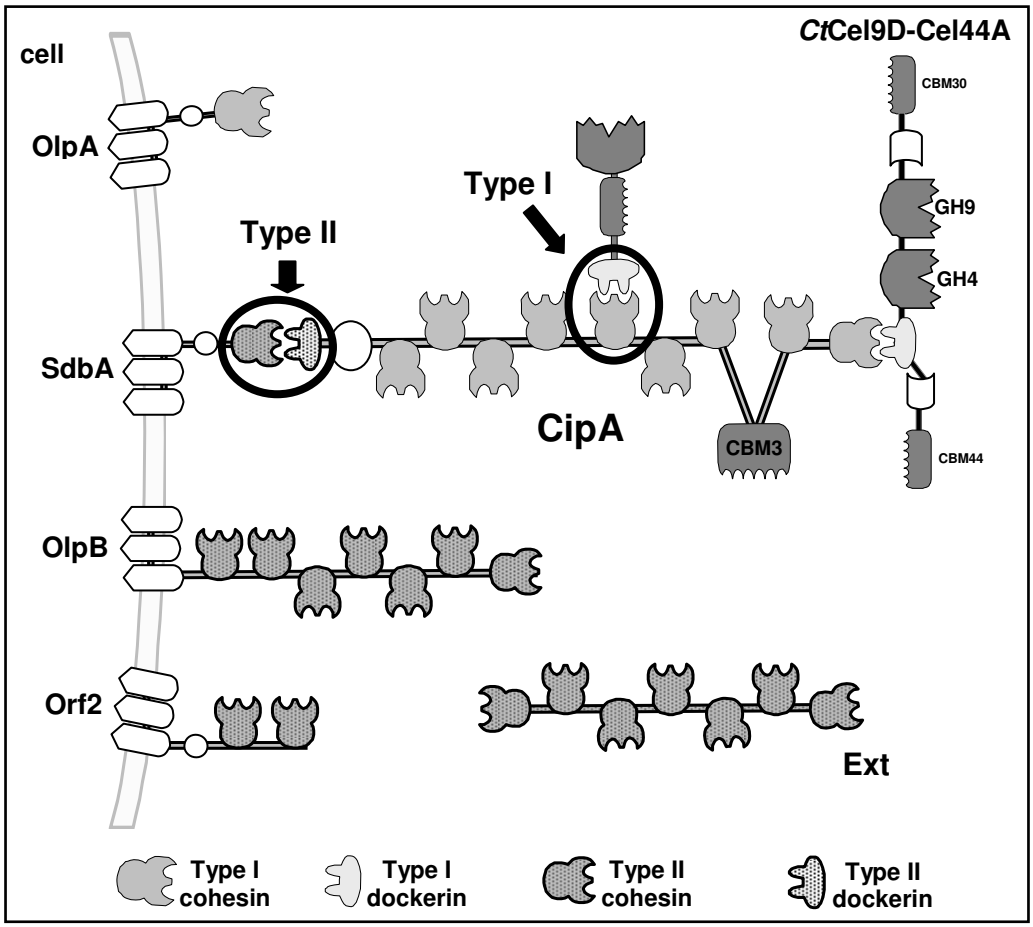

Figure 1.5. Schematic representation of *Clostridium thermocellum* cellulosome, showing the Type I and Type II cohesin-dockerin interactions.
The main component of the cellulosome is the scaffoldin subunit CipA that possesses a family 3 CBM and incorporates the enzymatic subunits through Type I cohesin-dockerin interactions. *C. thermocellum* cellulosome also possesses anchoring scaffoldins with SLH modules for binding to bacteria surface (SdbA, OlpA, OlpB and Orf2) and free scaffoldins, which do not bind cell surface (Ext). Type I interaction between a CipA cohesin and dockerin from *Ct*Cel9D-Cel44A (Chapter 3) is shown on the right.

During cell growth there is a tight association between bacteria and the substrate, which facilitates degradation and the ready assimilation of the resulting products (Doi *et al.*, 2003). Bayer *et al.* (2004) refer that any cohesin-containing protein can be considered a scaffoldin and suggest that bacterial cellulosomal systems can be classified into two major types: (i) those with multiple types of interacting scaffoldins, such as in *C. thermocellum*, and (ii) those that contain a single scaffoldin, such as in *C. cellulovorans*. In *C. thermocellum*, CipA includes an internal family 3 Carbohydrate-Binding Module (CBM), used for binding and fixing the cellulosome into the plant cell wall. The function of this CBM3 is well characterized and studied (Tormo *et al.*, 1996; Tomme *et al.*, 2003) and its three-dimensional structure has been elucidated (Tormo *et al.*, 1996). The data revealed that CipA CBM3 is a Type A CBM, which means that it contains a planar ligand interaction surface with a very high affinity for crystalline cellulose (see 1.5.1.1.); its equilibrium affinity constant (K_a) for Avicel was reported to be 7.7×10^6 M^{-1} (Tomme *et al.*, 2003). This CBM was fused with several antimicrobial peptides in Chapter 6.

Cellulosomal enzymes contain, at minimum, a catalytic domain and a dockerin module, as for instance Xgh74 (Chapter 2). Therefore, the presence of a dockerin module in an enzyme usually indicates that it is a cellulosomal component (Doi *et al.*, 2003). Cellulosomes are heterogeneous in enzymatic composition and this allows anaerobic bacteria to more efficiently degrade complex substrates such as the plant cell wall.

There has been a recent interest in constructing *in vitro* mini-cellulosomes for purposes such as studying cohesin functions, analyzing synergy between various cellulosomal enzymes and improving the efficiency of cellulose degradation (Doi *et al.*, 2003). There has also been an increasing interest in developing *in vivo* systems in bacteria in order to use relatively inexpensive biomass or agricultural wastes as substrates to obtain valuable products (Doi *et al.*, 2003). Cellulosomes have promising biotechnological applications since they degrade crystalline forms of cellulose, which are recalcitrant to many other enzymes that degrade soluble or amorphous forms of cellulose. Furthermore, cellulosomes also degrade hemicelluloses and pectin, two other major components of plant cell wall materials. Today, recombinant DNA technology allows the construction of engineered cellulosomes that might include enzymes that would suit a specific purpose. Improved or alternative multiple enzyme systems, novel affinity tags and self-assembling chimeric protein constituents can now be employed for

biotechnological and nanotechnological applications (Bayer *et al.*, 2004). It will be possible to insert cellulosomal genes into hosts that already produce valuable products so that those hosts could produce their own cellulosomes and grow on inexpensive plant cell wall materials.

1.4. GLYCOSIDE HYDROLASES

The extensive variety of carbohydrates is only paralleled by the large multiplicity of enzymes involved in their metabolism (Henrissat, 1991). Carbohydrate-active enzymes include glycoside hydrolases (GHs), glycosyltransferases (GTs), carbohydrate esterases (CE) and polysaccharide lyases (PL). Here attention will be focused on GHs, which are key enzymes for carbohydrate metabolism and can be found in archaebacteria, eubacteria and eukaryotes (Henrissat, 1991). GHs hydrolyse the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate and many of these enzymes are produced by microorganisms that are involved in the degradation of plant cell wall polysaccharides (Henrissat *et al.*, 1998).

1.4.1. GHs nomenclature

The IUB Enzyme Nomenclature (1984) is based on the type of reaction that the enzymes catalyse and on their substrate-specificity. For GHs (EC 3.2.1.*x*), the first three digits indicate enzymes hydrolysing *O*-glycosyl linkages whereas the last number indicates the substrate and sometimes reflects the molecular mechanism (Henrissat, 1991). This classification avoids ambiguities and the proliferation of trivial names but, at least in the case of GHs, it does not reflect the enzymes' structural features (Henrissat, 1991). Also it is not appropriate for enzymes that show broad specificity, *i.e.* that act on different substrates. Henrissat (1991) compared the primary sequences of GHs and have organized this class of enzymes in families. Interestingly, this initial work revealed that families of homologous enzymes contained several EC entries, such that enzymes with similar substrate-specificities belonged to non-related families. Therefore, Henrissat (1991) proposed a new classification for GHs, that organized these enzymes in families based on the homology shared at the level of the primary sequence. A family is defined when at least two sequences display significant amino acid similarity and when no significant similarity is found with other known families (Henrissat, 1991).

Henrissat's classification does not substitute but rather complements the IUB nomenclature. An advantage of this new classification is that a protein, a gene translation, or a domain, can be classified even when its enzymatic activity remains unknown (Henrissat and Bairoch, 1993). Significant sequence similarity is a strong indication of folding similarities, so it was found that members of one family share the same folding characteristics. Therefore, if the three-dimensional structure of one family member is known, homology modelling can be applied to determine the other members' three-dimensional structures. In addition, if the catalytic residue of a family member is known, the presence of this residue in the other members can immediately be traced (Henrissat, 1991). According to the Henrissat classification, the catalytic modules of GHs are currently classified into 112 different families based on amino acid sequence similarities (http://afmb.cnrs-mrs.fr/CAZY; April 2008). These families are also grouped into 14 clans or superfamilies, using the criteria of conservation of the protein fold, catalytic machinery and mechanism of glycosidic bond cleavage (Boraston *et al.*, 2004), as shown in Table 1.2.

Table 1.2. Glycoside hydrolases fold superfamilies.
Adapted from http://afmb.cnrs-mrs.fr/CAZY (April 2008).

Superfamily name	Protein fold	GH families
GH-A	$(\beta/\alpha)_8$	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86
GH-B	β-jelly roll	7, 16
GH-C	β-jelly roll	11, 12
GH-D	$(\beta/\alpha)_8$	27, 31, 36
GH-E	6-fold β-propeller	33, 34, 83
GH-F	5-fold β-propeller	43, 62
GH-G		37, 63
GH-H	$(\beta/\alpha)_8$	13, 70, 77
GH-I	$\alpha+\beta$	24, 46, 80
GH-J	5-fold β-propeller	32, 68
GH-K	$(\beta/\alpha)_8$	18, 20
GH-L	$(\alpha/\alpha)_6$	15, 65
GH-M	$(\alpha/\alpha)_6$	8, 48
GH-N	β-helix	28, 49

Henrissat *et al.* (1998) proposed that enzymes could also be named according to their target substrate, following the three-letter standard used in bacterial genetics for genes, although this designation would comprehend the family to which the enzyme belongs. Thus, an enzyme from a family 5 GH will be Cel5 or Man5, depending on its

substrate, which could be cellulose or mannan, respectively. If an organism produces multiple enzymes from a family, then they would be designated Cel5A, Cel5B, and so on, where the letters after the family number correspond to the order in which the enzymes were first reported (Henrissat *et al.*, 1998). If an enzyme contains more than one catalytic domain, its designation would include all of them. For example, previously named CelJ from *C. thermocellum* is composed of two cellulases, from families 9 and 44. According to this new nomenclature proposed by Henrissat *et al.* (1998), the enzyme is named *Ct*Cel9D-Cel44A, written from the amino- to carboxyl-terminus. The microorganism abbreviation is also included, before the enzyme name, to differentiate similar enzymes of different origins.

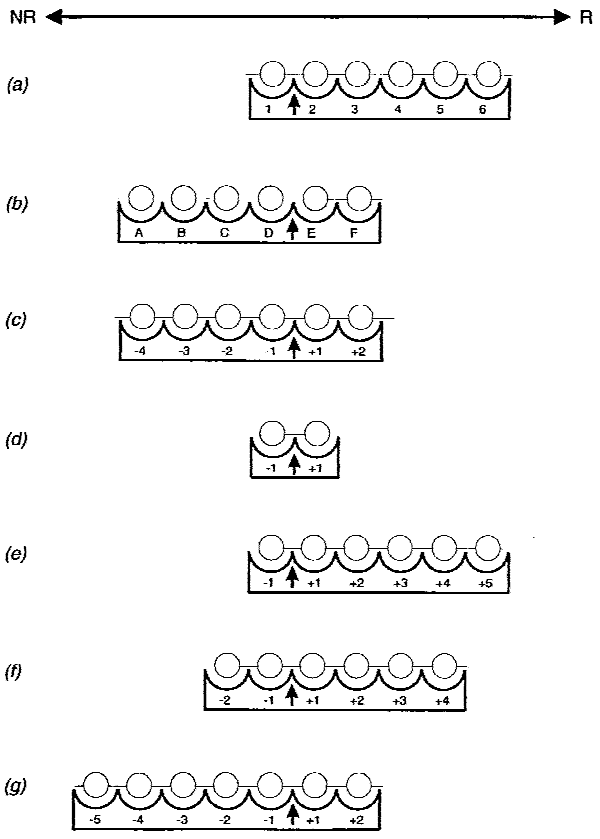

Figure 1.6. Schematic drawing of the sugar-binding subsites in several glycoside hydrolases.
By convention, the non-reducing end of the substrates is drawn on the left and the reducing end on the right. The point of cleavage is indicated by an arrow. *(a)* early subsite labelling by numbers, applied to glucoamylase; *(b)* subsite labelling, applied to hen egg-white lysozyme; *(c)* the subsites of hen egg-white lysozyme labelled with the *–n* to *+n* scheme; *(d) –n*, *+n* scheme applied to non-specific monoglycosidases and disaccharidases; *(e) –n*, *+n* scheme applied to enzymes cleaving a monosaccharide from the non-reducing end of the substrate; *(f) –n*, *+n* scheme applied to enzymes cleaving disaccharide units; *(g) –n*, *+n* scheme applied to enzymes cleaving disaccharide units from the reducing end of the substrate. Adapted from Davies *et al.* (1997b).

As the three-dimensional structures of more and more GHs are being solved, it became imperative that there is a system to designate the sugar-binding subsites within the active sites of the enzymes (Henrissat *et al.*, 1998). Davies *et al.* (1997b) proposed a system in which subsites are labelled from $-n$ to $+n$, with $-n$ at the non-reducing end and $+n$ at the reducing end of the substrate. In this system, cleavage occurs between the -1 and +1 subsites, as it is shown in Figure 1.6.

1.4.2. GHs in cellulolytic microbes

Cellulolytic microbes occur and act in many different environments ranging from plant litter in the soil to the anaerobic ecosystem of the animal rumen. These organisms developed a great variety of enzymes to degrade a large diversity of complex and insoluble plant cell wall polysaccharides. In addition, the enzyme systems produced by microorganisms for the hydrolysis of structural carbohydrates are remarkably complex and usually comprise hydrolases from several families (Henrissat *et al.*, 1998). Interestingly, it has been revealed that multiple enzymes degrading the same polysaccharide are expressed by individual aerobic or anaerobic microbes.

While aerobic bacteria and filamentous fungi typically use batteries of secreted and synergistically-acting cellulases, anaerobic bacteria, like *Clostridium thermocellum*, use large multi-enzyme complexes (cellulosomes) which operate at the cell-surface (Linder and Teeri, 1997). Within each system, two major classes of cellulases were identified: endoglucanases, which cleave the cellulose chains internally, and cellobiohydrolases, which release cellobiose from the chains ends (Warren, 1996). Both type of enzymes cleave β-1,4-glycosidic bonds and their endo/exo specificity seems to be determined by active site topology: while endoglucanases have open active site clefts, exoglucanases have their active site located in tunnels formed by long loops at the surface of the protein structure (Spezio *et al.*, 1993; Meinke *et al.*, 1995).

In this work, two GHs from *Clostridium thermocellum* were identified and characterized. Chapter 2 presents studies on Xgh74A, a xyloglucanase from the *C. thermocellum* cellulosome. It is the first xyloglucanase identified in this microorganism and the first active xyloglucanase in a cellulosome and it seems to play a very important role in the degradation of dicotyledons cell wall polysaccharides. Chapter 4 is dedicated to *Ct*Lic26A-Cel5E, another cellulosomal enzymatic complex from *C. thermocellum*, which expresses cellulase activity and possesses a family 11 CBM.

1.5. CARBOHYDRATE-BINDING MODULES

In general, GHs that degrade structural polysaccharides are modular enzymes containing, in addition to the catalytic domains, non-catalytic Carbohydrate-Binding Modules (CBMs) that target the appended catalytic modules to their specific glycosidic bonds (Boraston *et al.*, 2004). CBMs are known to fulfil three main roles with respect to the function of their associated catalytic modules: (i) a proximity effect, (ii) a targeting function and (iii) a disruptive function (Boraston *et al.*, 2004). Through their carbohydrate-binding activity, CBMs promote the association of the enzyme with its target substrate. It is believed that maintaining the enzyme in the proximity of its substrate leads to more rapid degradation of the polysaccharide (Bolam *et al.*, 1998), therefore resulting in improved enzyme efficiency. There are numerous examples where proteolytic excision or genetic truncation of CBMs from the catalytic modules significantly decreases the activity of the enzyme against insoluble substrates (Gilkes *et al.*, 1988; Bolam *et al.*, 1998; Tomme *et al.*, 1998; Charnock *et al.*, 2000; Ali *et al.*, 2001; Zverlov *et al.*, 2001; Boraston *et al.*, 2003a). In addition, there are CBMs that have become components of the substrate-binding sites of GHs, and are pivotal to the substrate specificity and mode of action of the enzymes. Therefore, the CBM22 from *Clostridium stercorarium* was shown to change the specificity of a family 10 xylanase to a β-1,4-β-1,3-glucanase (Araki *et al.*, 2004; Boraston *et al.*, 2004). Targeting and disrupting activities of CBMs will be focused later in this chapter (1.5.2.1 and 1.5.2.3, respectively).

Over the last years, the importance of CBMs has increased and many studies have revealed the structures and functions of these important non-catalytic modules. Structural information provided by NMR spectroscopic and X-ray crystallographic studies have contributed to improve our understanding of the biological functions of CBMs (Boraston *et al.*, 2004). In addition, the recent elucidation of several CBM structures in complex with oligosaccharide ligands provided valuable information on how these proteins recognize their ligands. In Chapter 3 two novel CBM structures from the *Clostridium thermocellum* modular enzyme *Ct*Cel9D-Cel44A are described.

The non-catalytic modules from modular enzymes were initially named cellulose-binding domains (CBDs) because the first to be discovered bound to crystalline cellulose, as it was described by Van Tilbeurgh *et al.* (1986), Gilkes *et al.* (1988) and Tomme *et al.* (1988). Later, other modules that bound to a range of different polysaccharides were found, so the wider name carbohydrate-binding module was

adopted. Although the term CBM is now in current use, the term CBD remains to describe the CBMs that bind specifically to cellulose (Boraston *et al.*, 2004).

CBMs exist in enzymes other than cellulases and xylanases. For instance, in *Trichoderma reesei* CBMs were identified in a hemicellulase, an endomannanase and an acetylxylanesterase (Stålbrand *et al.*, 1995; Margolles-Clark *et al.*, 1996; Shoseyov *et al.*, 2006). Other examples of enzymes with a CBM in their structure are an esterase from *Penicillium funiculosum* (Kroon *et al.*, 2000), an isomaltodextranase from *Arthrobacter globiformis* (Hatada *et al.*, 2004), arabinofuranosidases from *Aspergillus kawachii* and from *Cellvibrio japonicus* (Bolam *et al.*, 2004; Miyanaga *et al.*, 2004), a pectate lyase from *Pseudomonas cellulosa* (Brown *et al.*, 2001), a *β*-glucosidase from *Phanerochaete chrysosporium* (Lymar *et al.*, 1995) and a dextranase from *Paenibacillus* sp (Finnegan *et al.*, 2005).

CBMs were shown to interact with a range of ligands such as crystalline cellulose, non-crystalline cellulose, chitin, *β*-1,3-glucans and *β*-1,3-1,4-mixed linkage glucans, xylan, mannan, galactan and starch, while other CBMs display "lectin-like" specificity and bind to a variety of cell-surface glycans (Boraston *et al.*, 2004). A xyloglucan-binding CBM, CBM44, from *C. thermocellum Ct*Cel9D-Cel44A is identified and characterized in Chapter 3. It is the first description of a xyloglucan-binding CBM and is also the founder member of family 44.

CBMs are also part of proteins with no hydrolytic activity, comprising part of a scaffoldin subunit that organizes the catalytic subunits in the cellulosome, like family 3 CBM from *Clostridium thermocellum* scaffoldin CipA (see 1.3). It is clearly established that removal of CBMs from enzymes, or from scaffoldins of cellulosome, dramatically reduces the enzymatic activity of the associated catalytic modules (Van Tilbeurgh *et al.*, 1986; Tomme *et al.*, 1988; Hall *et al.*, 1995; Bolam *et al.*, 1998; Zverlov *et al.*, 2001; Boraston *et al.*, 2003a; Ali *et al.*, 2005).

Expansins and lectins are molecules homologous to CBMs in the sense that they also bind carbohydrates. Expansins seem to have a role in non-hydrolytic cell wall expansion and possess cellulose-binding capacity *in vitro* (Shoseyov *et al.*, 2006). Lectins are involved in the recognition of glycoproteins at cell surfaces (McGreal *et al.*, 2004). Therefore, like CBMs, lectins are proteins that recognize carbohydrates which fulfil an important a role in the immune system that has been extensively studied over the last years.

1.5.1. CBMs nomenclature and types

Like GHs, CBMs are divided into families according to the homology shared at the level of their primary amino acid sequences. So far, CBMs have been grouped into 51 families (Coutinho and Henrissat, 1999; http://afmb.cnrs-mrs.fr/CAZY; April 2008), in a systematic nomenclature similar to the one adopted for GHs. A CBM is named by its family; for instance the family 3 CBM from *Clostridium thermocellum* scaffoldin CipA would be called CBM3. The organism name can also be included, or even the enzyme from where it derives, so in the previous example the CBM would be named *Ct*CBM3 or *Ct*CipACBM3. If GHs contain tandem CBMs belonging to the same family, a number corresponding to the position of the CBM in the enzyme relative to the N-terminus is also included (Boraston *et al.*, 2004). For example, *Clostridium thermocellum* Xyn10B contains two family 22 CBMs, so the first is referred to as *Ct*CBM22-1 and the second as *Ct*CBM22-2.

Boraston *et al.* (2004) proposed the classification of CBM families by their structural fold (Table 1.3), similarly to what was described to the GHs superfamily classification. The β-sandwich fold, family fold 1 (Table 1.3), is the dominant fold among CBMs and it comprises two β-sheets, each one consisting of three to six antiparallel β-strands (Boraston *et al.*, 2004). CBMs share this fold with plant legume lectins and animal galectins, pentraxins, spermadhesins and calnexins, although no CBM has a significant residue similarity with these other proteins (Boraston *et al.*, 2004). All β-sandwich-fold-containing CBMs have, at least, one bound metal atom, with exception of *Cf*CBM2a from *Cellulomonas fimi* xylanase 10A. In most cases, the metal atom seems to have a structural role, but in *Pp*CBM36 from *Paenibacillus polymyxa* Xyn43A (Figure 1.7) the ligand interaction is mediated by a calcium atom (Jamal-Talabani *et al.*, 2004). In the beginning of 2004, when Boraston *et al.* (2004) classified CBMs into fold families, the majority of β-sandwich fold CBMs had a β-jelly roll fold (Hashimoto, 2006). However, in the last years, several CBMs with β-sandwich fold were shown to have an immunoglobulin fold instead, like CBMs from families 9, 20, 25, 26, 31, 33 and 34 (Hashimoto, 2006). Thus, the β-sandwich CBM fold family is currently divided into two fold sub-families, β-jelly roll fold and immunoglobulin fold (Table 1.3). More information about β-sandwich superfamily can be found in Chapter 3, where the three-dimensional structures of *Ct*CBM30 and *Ct*CBM44 are described. Only *Ct*CBM44 was shown to contain structural calcium atoms.

Table 1.3. CBM fold families' classification.
Adapted from Boraston *et al.* (2004) and Hashimoto (2006).

Fold family	Type of fold		CBM families
1	*β*-sandwich		
		β-jelly roll	2, 3, 4, 6, 11, 15, 17, 22, 27, 28, 29, 30, 32, 35, 36, 44
		immunoglobulin	9, 20, 25, 26, 31, 33, 34
2	*β*-trefoil		13, 42
3	cysteine knot		1
4	unique		5, 12
5	OB (oligonucleotide/oligosaccharide binding fold)		10
6	hevein fold		18
7	unique; contains hevein-like fold		14

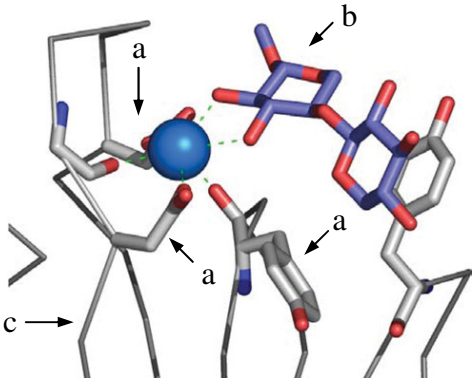

Figure 1.7. The role of calcium in xylan recognition by family 36 CBM, *Pp*CBM36, from *Paenibacillus polymyxa* Xyn43A.
The calcium atom located in the bindind site is shown as a blue sphere; a, residues involved in coordinating the calcium atom and binding to the sugar are presented in grey; b, bound xylo-oligosaccharide is shown in blue; c, C*α* backbone is represented as grey cylinders. Adapted from Boraston *et al.* (2004).

The second CBM-fold in frequency is the *β*-trefoil, fold family 2 (Table 1.3). Until now only members of families 13 and 42 belong to this superfamily. This CBM fold contains 12 *β*-strands, forming six hairpin turns. A *β*-barrel structure is formed by six of the strands, attendant with three hairpin turns and the other three hairpin turns form a triangular cap on one end of the *β*-barrel, named the hairpin triplet (Boraston *et al.*, 2004). The subunit of this fold, which can be called a trefoil domain, is a contiguous residue sequence with a four *β*-strand, two-hairpin structure having a trefoil shape; each trefoil domain contributes one hairpin (two *β*-strands) to the *β*-barrel and one hairpin to the hairpin triplet (Boraston *et al.*, 2004). An example of a family 13 CBM, *Sl*CBM13, from *Streptomyces lividans* Xyn10A, is shown later in this chapter (Figure 1.10, in

complex with xylopentaose, and Figure 1.12, in complex with three lactose molecules that occupy the three CBM binding sites).

Members of fold families 3, 4 and 5 (Table 1.3) are small polypeptides with 30-60 amino acids, that contain only β-sheets and coils. These CBMs show less diversity in respect to their ligand specificities since these folds appear more specialized in the recognition of cellulose and/or chitin (Boraston *et al.*, 2004). The majority of these CBMs have a planar carbohydrate-binding site comprising usually three aromatic residues, with the exception of family CBM12 (Boraston *et al.*, 2004). Hevein domains, fold families 6 and 7 (Table 1.3), are very small CBMs with approximately 40 amino acids, originally identified as chitin-binding proteins in plants (Boraston *et al.*, 2004). The fold presents predominantly a coil but also have two small β-sheets and a small α-helix. Two examples of these fold family are presented in Figure 1.10c and Figure 1.10e. The minimal hevein fold is found in family CBM18, which is thus classified as family fold 6 (Table 1.3). Family 14 CBMs incorporate aspects of hevein domain but also have a fusion with a small β-sheet structure, which leads to its classification in another family fold, family fold 7 (Table 1.3) (Boraston *et al.*, 2004).

Although CBM families can be grouped into fold families based on protein fold conservation, these groups do not predict CBM function. Thus, CBMs have also been classified into three types, Type A, Type B and Type C, regarding the relationship established between the topology of the binding site and CBM function (Boraston *et al.*, 2004), as depicted in Table 1.4.

Table 1.4. CBM Types classification.
Boraston *et al.* (2004) and http://afmb.cnrs-mrs.fr/CAZY (April 2008).

CBM Type	Fold family	CBM families
A	1, 3, 4, 5	1, 2a, 3, 5, 10
B	1	2b, 4, 6, 11, 15, 17, 20, 22, 25, 26, 27, 28, 29, 30, 31, 33, 34, 35, 36, 41, 44, 47
C	1, 2, 6, 7	9, 12, 13, 14, 18, 32, 40, 42, 43, 50

1.5.1.1. Type A - surface-binding CBMs

Type A CBMs bind strongly to insoluble polysaccharides surfaces. Type A CBMs are found in families 1, 2a, 3, 5 and 10 (Figure 1.8), that bind to insoluble highly crystalline cellulose, chitin and mannan and show little or no affinity for soluble

carbohydrates. These CBMs present a flat or platform-like binding surface, which is complementar to the flat surfaces presented by cellulose or chitin crystals (Tormo *et al.*, 1996). Interaction of Type A CBMs with crystalline cellulose is associated with a positive entropy, which demonstrates that the thermodynamic forces driving the binding of CBMs to crystalline ligands are relatively unique among carbohydrate-binding proteins, being different from the protein interactions to soluble carbohydrates (Creagh *et al.*, 1996). The molecular basis for the thermodynamic forces that drive protein-carbohydrate interactions remains, presently, a controversial theme, particularly in what concerns the role of water molecules and the loss of entropy through conformational restriction (Boraston *et al.*, 2004).

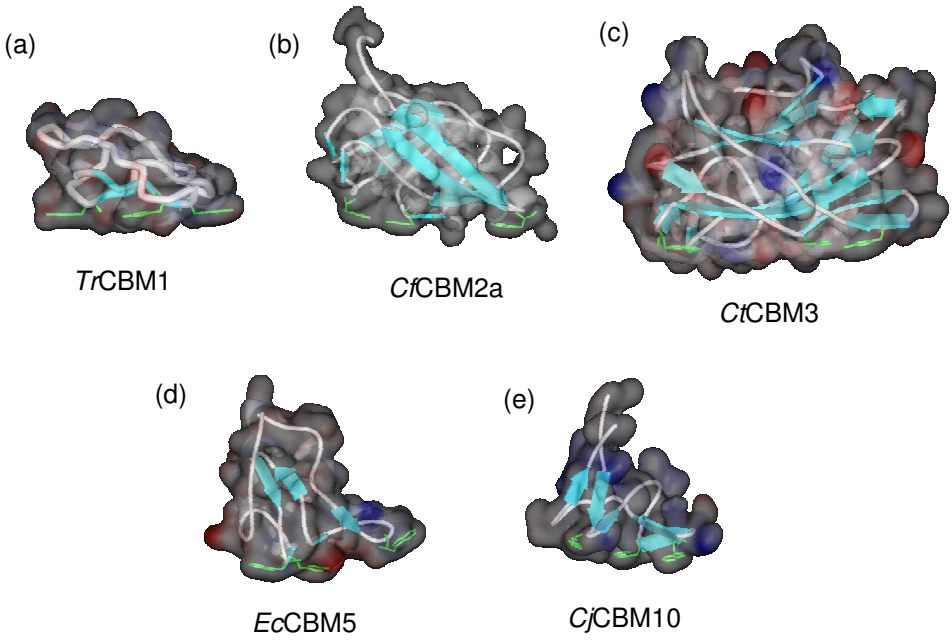

Figure 1.8. Examples of CBMs belonging to the functional Type A.
(a) Family 1 CBM, *Tr*CBM1, from *Trichoderma reesei* cellobiohydrolase I (PDB code 1CBH; Kraulis *et al.*, 1989); (b) family 2a CBM, *Cf*CBM2a, from *Cellulomonas fimi* Xyn10A (PDB code 1EXG; Xu *et al.*, 1995); (c) family 3 CBM, *Ct*CBM3, from *Clostridium thermocellum* CipA (PDB code 1NBC; Tormo *et al.*, 1996); (d) family 5 CBM, *Ec*CBM5, from *Erwinia chrysanthemi* Cel5A (PDB code 1AIW; Brun *et al.*, 1997); (e) family 10 CBM, *Cj*CBM10, from *Cellvibrio japonicus* Xyn10A (PDB code 1E8R; Raghothama *et al.*, 2000). Aromatic amino acids involved in ligand recognition and binding are shown in green.

1.5.1.2. Type B - glycan-chain-binding CBMs

Type B CBMs bind to soluble polysaccharide chains. The carbohydrate-binding sites of Type B CBMs are extended and are usually described as grooves or clefts

comprising several subsites that are able to accommodate the individual sugar units of the ligand (Boraston *et al.*, 2004). Type B CBMs include members from families 2, 4, 6, 15, 22 and 29 (Figure 1.9). In general, the binding competence of Type B CBMs is determined by the degree of polymerization of the carbohydrate ligand: biochemical affinity studies have demonstrated increased binding for hexasaccharides and negligible interactions in the case of smaller oligosaccharides (Boraston *et al.*, 2004). Therefore, these CBMs are usually described as "chain binders". The depth of the binding clefts varies from very shallow to being able to accommodate an entire pyranose ring. Chapter 3 presents the biochemical and structural analysis of two Type B CBMs from *Ct*Cel9D-Cel44A, *Ct*CBM30 and *Ct*CBM44.

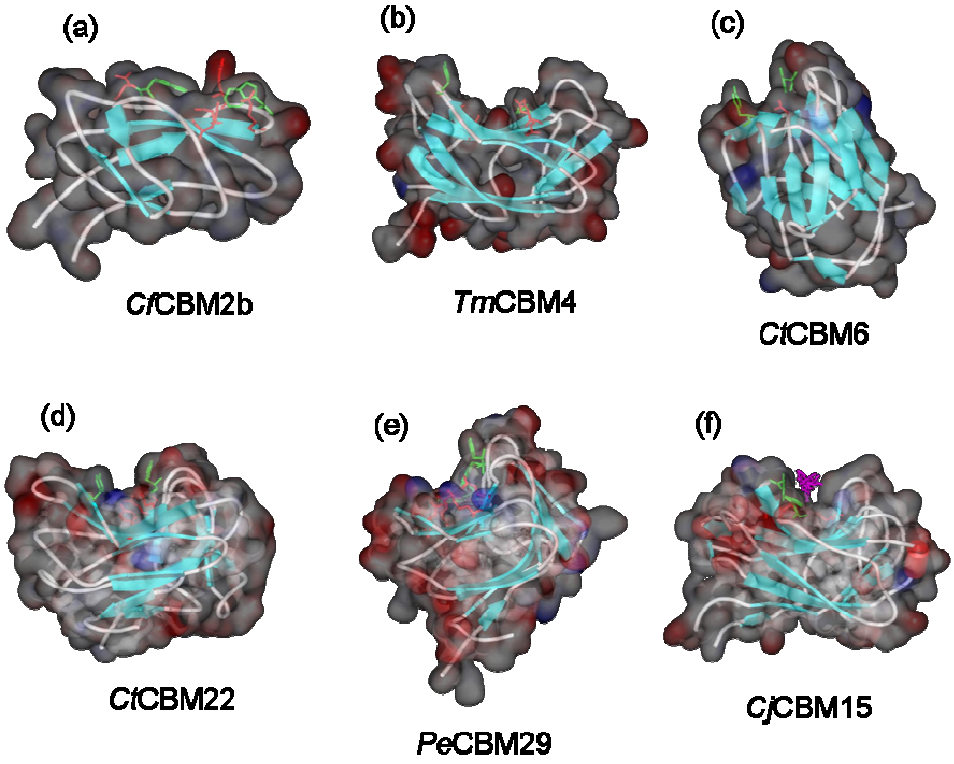

Figure 1.9. Examples of CBMs belonging to the functional Type B.
(a) family 2b CBM, *Cf*CBM2b, from *Cellulomonas fimi* Xyn11A (PDB code 2XBD; Simpson *et al.*, 1999); (b) family 4 CBM, *Tm*CBM4, from *Thermotoga maritima* Lam16 (PDB code 1GUI; Boraston *et al.*, 2002); (c) family 6 CBM, *Ct*CBM6, from *Clostridium thermocellum* Xyn11A (PDB code 1GMM; Czjzek *et al.*, 2001); (d) family 22 CBM, *Ct*CBM22, from *Clostridium thermocellum* Xyn10B (PDB code 1H6Y; Xie *et al.*, 2001b) (e) family 29 CBM, *Pe*CBM29, from *Piromyces equi* NCP1 (PDB code 1GWK; Charnock *et al.*, 2002); (f) family 15 CBM, *Cj*CBM15 from *Cellvibrio japonicus* Xyn10C (PDB code 1GNY; Szabo *et al.*, 2001). Aromatic amino acids involved in ligand recognition and binding are shown in green.

1.5.1.3. Type C - small sugar-binding CBMs

This class of CBMs has the lectin-like property of binding exclusively to mono-, di- or tri- saccharides, thus lacking the extended binding-site clefts of Type B CBMs (Boraston *et al.*, 2004). However, the distinction between Type B and Type C CBMs can, sometimes, be very subtle. Type C CBMs include members from families 9, 13, 14, 18, 32, 40, 42 and 43 (Table 1.4 and Figure 1.10). Members of families 13, 14 and 18, which are located in a variety of GHs, were first discovered as lectins and only subsequently have been named as CBMs, due to the discovery of their exclusive specificity for carbohydrates (Boraston *et al.*, 2004).

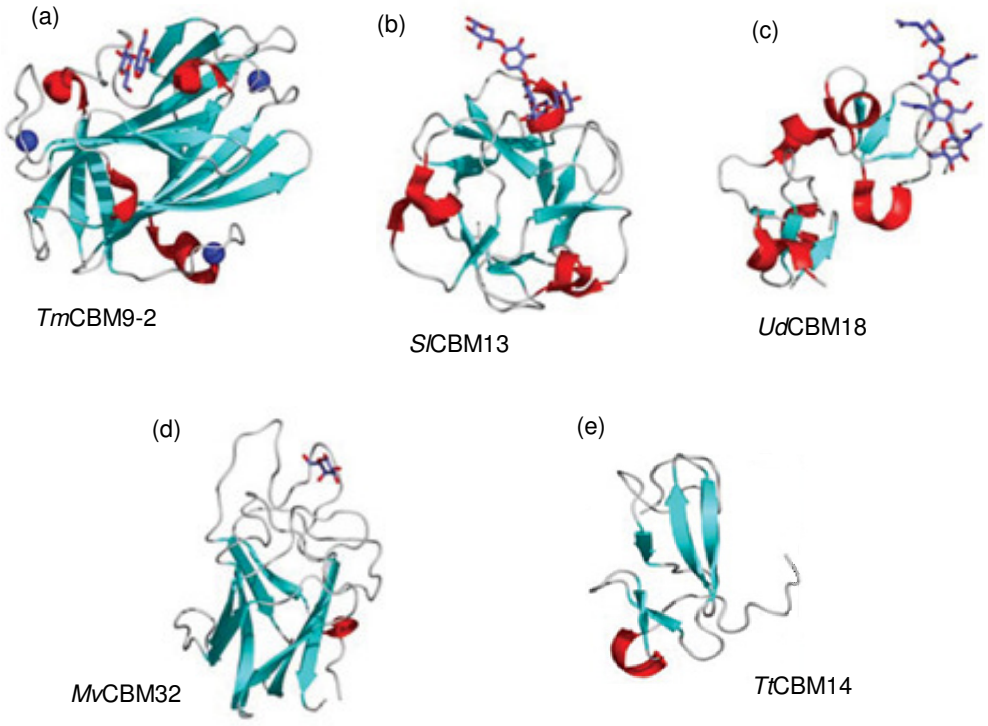

Figure 1.10. Examples of CBMs belonging to the functional Type C.
(a) Family 9 CBM, *Tm*CBM9-2, from *Thermotoga maritima* Xyn10A in complex with cellobiose (PDB code 1I82; Notenboom *et al.*, 2001b); (b) family 13 CBM, *Sl*CBM13, from *Streptomyces lividans* Xyn10A in complex with xylopentaose (PDB code 1MC9; Notenboom *et al.*, 2002); (c) family 18 CBM, *Ud*CBM18, from *Urtica dioca* agglutinin in complex with chitotriose (PDB code 1EN2; Saul *et al.*, 2000); (d) family 32 CBM, *Mv*CBM32, from a *Micromonospora viridifaciens* sialidase in complex with galactose (PDB code 1EUU; Gaskell *et al.*, 1995); (e) family 14 CBM, *Tt*CBM14, from *Tachypleus tridentatus* tachycitin (PDB code 1DQC; Suetake *et al.*, 2000). Adapted from Boraston *et al.* (2004).

1.5.2. The mode of action of CBMs

1.5.2.1. Enzyme targeting

It has been shown that Type A CBMs can be appended to GHs expressing a variety of substrate specificities while Type B CBMs, in contrast, bind to polysaccharides that are the substrates for the adjacent catalytic modules (Boraston *et al.*, 2004). For example, cellulases, xylanases, galactanases or mannanases contain Type B CBMs that bind to cellulose, xylan, galactan or mannan, respectively. Thus, Type B CBMs maintain the proximity of the catalytic module of the enzyme to the target substrate. The targeting function is particular important in the context of the plant cell wall, a complex macromolecule composed of a variety of structural polysaccharides. CBM are the biochemical tools that direct cellulases and hemicellulases to their target substrates as it will be further explored in Chapter 3.

Interaction of CBMs with cellulose is occasionally irreversible (Carrard and Linder, 1999; Pinto *et al.*, 2004), although the contact with the cellulose surface is a dynamic process, as demonstrated by Jervis *et al.* (1997). Using CBMs labelled with a fluorescent tags, these authors showed that CBMs from *Cellulomonas fimi* are mobile on the surface of crystalline cellulose. At the diffusion rates reported, CBM transverse several hundreds of lattice units on the cellulose crystal in 1 min. This dynamic binding behaviour of CBMs is understandable as often they are part of processive acting cellulases (Lehtiö *et al.*, 2003). Although CBMs have a high affinity for carbohydrates, it is obvious that an irreversible binding would be fatal for the catalytic efficiency of the appended hydrolytic domains.

1.5.2.2. The role of the aromatic residues in ligand binding and specificity

All CBMs use aromatic and polar residues to establish the interaction with their target ligands (Lehtiö, 2001). However, it has been demonstrated that binding-site topography is a key determinant of binding specificity (Boraston *et al.*, 2004). The two major factors involved in CBMs binding-site topography seem to be the location and orientation of the aromatic amino acid side-chains at the binding site and the loop architecture at the binding sites, which usually shape ligand conformation (Boraston *et al.*, 2004). The side-chains of tryptophan, tyrosine and, less commonly, phenylalanine, form the hydrophobic platforms in most CBMs binding-sites, which can be planar, twisted or form a sandwich, as presented in Figure 1.11 (Boraston *et al.*, 2004). As previously referred, planar platforms (Figure 1.11A) are characteristic of Type A

CBMs. The apolar platform can be twisted, like in families 2b, 15, 17, 27, 29, 34 and 36, due to the rotation of two or three amino acid chains relative to each other (Figure 1.11B). In Type B CBMs, the binding cleft can also present a sandwich form, where the aromatic amino acid chains sandwich a ligand sugar unit by stacking the β and α faces of the pyranose ring (Figure 1.11C). This latest case was observed in CBMs of families 4, 6, 9 and 22. The sandwich and twisted platforms may be used concurrently in the same CBM and can both accommodate the conformations of soluble oligosaccharide ligands (Boraston *et al.*, 2004). CBMs appear to have preformed carbohydrate-recognition sites which mirror the solution conformations of their target ligands, therefore minimizing the energy required for binding (Boraston *et al.*, 2004).

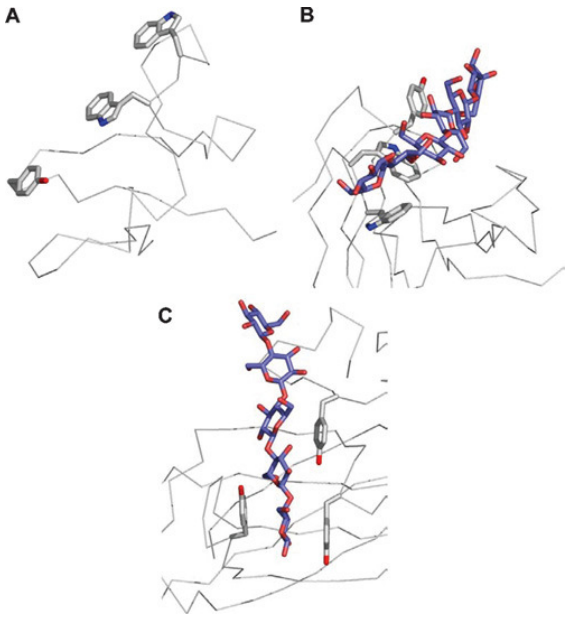

Figure 1.11. The three types of binding-site platforms formed by aromatic acid residues.
(A) the planar platform of family 10 Type A CBM, *Cj*CBM10, from *Cellvibrio japonicus*; (B) the twisted platform of family 29 Type B CBM, *Pe*CBM29-2, from *Piromyces equi*; (C) the sandwich platform of family 4 Type B CBM, *Cf*CBM4-2, from *Cellulomonas fimi*. Cα backbone is represented as grey cylinders, the aromatic amino acid side-chains forming the binding sites are presented in grey and bound oligosaccharides are shown in blue. Adapted from Boraston *et al.* (2004).

In Type A CBMs binding to crystalline cellulose the flat platform contains aromatic residues, most often tyrosines and tryptophans, separated by a distance corresponding to the length of the repeating unit in cellulose, 10.3 Å. In this case, the flat aromatic rings interact with the pyranose rings of the polysaccharides (Tomme *et al.*, 1995). This driving interaction may be supplemented by few hydrogen bonds

mediated by polar residues located at the binding surface (Tormo *et al.*, 1996). In contrast, Type B CBMs binding sites are equipped to interact with individual glycan chains rather than crystalline surfaces (Boraston *et al.*, 2004). As with Type A, in Type B CBMs the aromatic residues play a pivotal role in ligand binding and the orientation of these amino acids are key determinants of the CBM binding specificity (Simpson *et al.*, 2000). For example, CBM family 2 contains Type A and Type B members, named CBM2a and CBM2b, which bind crystalline cellulose or xylan, respectively. The explanation for this contrasting ligand specificity results from the rotation by 90º of the side-chain of one surface tryptophan (Trp^{259}) involved in the protein-carbohydrate interaction in CBM2b compared with its position in CBM2a (Simpson *et al.*, 2000). Therefore, in CBM2b, the surface tryptophan matches the helical secondary structure of xylan. To prove this possibility, Simpson *et al.* (2000) showed that a mutation of a single residue (Arg^{262} to Gly) in CBM2b changed the specificity of the CBM from xylan to cellulose due to an alteration of the orientation of the aromatic side-chain. In fact, the structural effect of the mutation was revealed by NMR spectroscopy and confirmed that Trp^{259} rotated 90º to lie flat against the protein surface, therefore explaining the mutant affinity for cellulose. Taken together, the data suggest that in CBMs belonging to family 2, ligand specificity is primarily determined by the accommodation of the three-dimensional shape of the ligand, rather than by the specific generation of hydrogen bonding patterns, as it is typically seen in other proteins that recognize monosaccharides (Simpson *et al.*, 2000).

In contrast with Type A CBMs, direct hydrogen bonds may play a key role in defining the affinity and ligand specificity of Type B CBMs, usually described as chain binders (Notenboom *et al.*, 2001; Xie *et al.*, 2001; Pell *et al.*, 2003). There is no evidence, however, indicating that water-mediated hydrogen bonds are critical in the interaction of CBMs with their target ligands (Pell *et al.*, 2003). Although the orientation and positioning of the aromatic residues in the binding sites is the primary driver of specificity and affinity for carbohydrates, direct hydrogen bonds and calcium mediated coordination (see 1.5.1 and Figure 1.7) may additionally play important roles in CBMs ligand recognition (Boraston *et al.*, 2004).

1.5.2.3. Non-hydrolytic substrate disruption and surface/interfacial modifications

Specific CBMs, mainly Type A, were described as being able to disrupt the structure of cellulose fibers in a non-hydrolytic fashion (Shoseyov *et al.*, 2006). This

function was first described for the N-terminal CBM2a of Cel6A from *Cellulomonas fimi* (Din *et al.*, 1994), which appears to mediate the non-catalytic disruption of the crystalline structure of cellulose. In addition, this effect was shown to enhance the catalytic capacity of the appended catalytic module (Boraston *et al.*, 2004).

It has also been demonstrated that CBMs prevent the flocculation of crystalline cellulose (Pagès *et al.*, 1997; Xiao *et al.*, 2001; Levy *et al.*, 2002), while a CBM from *Trichoderma reesei* caused peeling and smoothed surface on cotton fibers (Lee *et al.*, 2000). Other studies indicated that treatment of cellulose fibres with CBMs alters their interfacial properties (Suurnäkki *et al.*, 2000; Pala *et al.*, 2001). This non-disruptive ability gives CBMs the potential for use in biotechnological applications in the textile and paper industries (see 1.6).

1.5.3. CBMs multivalency

The weak interactions between a carbohydrate-binding protein and its ligand are often compensated in nature by multivalent interactions, where multiple clustered carbohydrate-binding sites interact simultaneously with carbohydrate ligands, which present multiple recognition elements (Boraston *et al.*, 2004). This multivalency results in an increased association constant in relation to any one of the isolated carbohydrate-protein interactions (Boraston *et al.*, 2004). It is clear that clustered carbohydrate-binding sites can result from (i) a single protein having multiple binding sites, (ii) the association of two or more univalent carbohydrate-binding proteins into quaternary structures, which is the most common case in cellulases and hemicellulases, or (iii) clustering of receptors, for example, in cell membranes.

To date, no CBM has been found to form quaternary structures in its natural state but multiple CBMs are often found in GHs and that is the form of multivalency used by CBMs to overcome relatively weak binding (Boraston *et al.*, 2004). The presence of multiple CBMs in GHs seems to occur most frequently in thermophilic or hyperthermophilic enzymes, most possibly as a response for the loss of affinity that accompanies the molecular interactions at elevated temperatures (Boraston *et al.*, 2003a).

The first tandem CBMs to be investigated were the family 2b CBMs of *Cellulomonas fimi* xylanase 11A (Bolam *et al.*, 2001). The individual association constants for xylan are approximately 10^4 M^{-1} but the in tandem association, as it occurs in nature, has an affinity constant of approximately 10^6 M^{-1}. The same was observed for

the three family 6 CBMs from the *Clostridium stercorarium* putative xylanase (Boraston *et al.*, 2002b) and the tandem CBM17 and CBM28 from *Bacillus* sp. 1139 Cel5 (Boraston *et al.*, 2003a).

Individual CBMs containing multiple carbohydrate-binding sites occur in a variety of CBM families, as shown in Figure 1.12. The family 13 Type C CBM from *Streptomyces lividans* xylanase 10A has three separate binding sites, one in each trefoil domain, Figure 1.12a (Boraston *et al.*, 2000b; Notenboom *et al.*, 2002; Schärpf *et al.*, 2002). In addition, the family 6 CBM from *Cellvibrio mixtus* endoglucanase 5A was shown to contain two binding sites (Figure 1.12b) that were named as cleft A, which accommodates the chain ends of *β*-1,4-glucans, *β*-1,3-glucans and xylans, and cleft B, which binds to internal regions of *β*-1,4-glucans and mixed *β*-(1,4)(1,3)-glucans (Henshaw *et al.*, 2004).

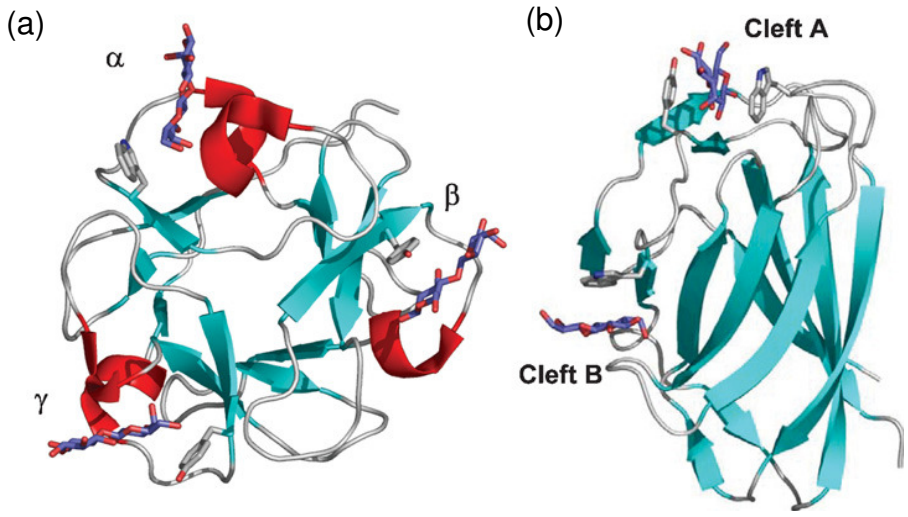

Figure 1.12. Multivalent CBMs that display different interacting binding sites.
(a) The family 13 CBM, *Sl*CBM13, from *Streptomyces lividans* Xyn10A, in complex with three lactose molecules that occupy its three binding sites; the three *β*-trefoil subdomains of *Sl*CBM13 are labelled as *α*, *β* and *γ*; (b) the family 6 CBM, *Cm*CBM6, from *Cellvibrio mixtus* endoglucanase 5A, in complex with two cellobiose molecules that occupy its two binding sites. The aromatic amino acid side-chains forming the binding sites are presented in grey and bound oligosaccharides in blue. Adapted from Boraston *et al.* (2004).

1.6. BIOTECHNOLOGICAL APPLICATIONS FOR CBMS

CBMs are presently used in a variety of industrial applications. Since one of the objectives of this work is to explore novel biotechnological applications for these non-catalytic anchoring domains, a review on the most significant achievements in the field is required. Levy and Shoseyov (2002) and Shoseyov *et al.* (2006) suggested that three basic properties have contributed to consider CBMs as the perfect candidates for a range of biotechnological applications and those are: (i) CBMs are usually independent and, therefore, can work in an autonomous manner; (ii) the attachment matrices for CBMs are abundant and inexpensive and also have excellent chemical and physical properties; and (iii) the binding capacities of CBMs can be controlled and adapted (Shoseyov *et al.*, 2006).

CBMs are considered interesting affinity tags for the expression and purification of proteins particularly for large-scale production. When compared with other immobilization systems, cellulose is a very economical support-matrix for the industry (Levy and Shoseyov, 2002), while CBM tags allow the development of secure and quick purification protocols. Several studies established that CBMs can be employed as high-capacity purification tags for the isolation of biologically active target peptides (Levy and Shoseyov, 2002), maintaining those peptides functional. Tomme *et al.* (1998) have also suggested that CBMs can be attached to a target protein without interfering with that protein biological activity. Protein engineering using CBMs is also an emerging field and high-level expression vectors have been designed for the production of CBM-fused proteins (Levy and Shoseyov, 2002). Fusing a biological active protein with a CBM seems to increase its level of expression in *E. coli* (works by Kim *et al.*, 1998 and Otomo *et al.*, 1999). In addition, several studies have demonstrated the potential use of CBMs for improving enzyme activity (Levy and Shoseyov, 2002).

The textile industry has also been extensively exploring the possibility of using CBMs to change the properties of specific fabrics, or to target a variety of enzymes to the cellulose-based fabrics. In addition, CBMs can be associated with enzymes in laundry powders, so that affinity for the cellulose substrate is improved and enzyme performance increased. Additional substances can also be targeted to the cellulosic fabrics through the use of CBMs, as for instance fragrance-bearing particles in the case of laundry powders. Many patents have been produced exploring these uses for CBMs. It was also discovered that CBM treatment of cellulose fibres originates less peeling and more smoothing of the fibre surface, what is an enormous advantage for the textile

industry and paper production (Lee *et al.*, 2000; Suurnäkki *et al.*, 2000). Applying CBMs to cellulose fibres also has a potential use in paper recycling (Pala *et al.*, 2001; Levy and Shoseyov, 2002).

Environmental technology has been also exploring the function of CBMs. Several authors fused CBMs with decontamination enzymes and the fusion molecules, attached to cellulose supports, successfully degraded toxic compounds (Richins *et al.*, 2000; Wang *et al.*, 2002; Xu *et al.*, 2002). Additional biotechnological uses for CBMs are in the production of oral care products (since CBMs may contribute to disperse the polysaccharides present in the dental plaque), in the baking industry (as CBMs can retard staling and aging of baked bread), in diagnose test kits (Levy and Shoseyov, 2002) and as part of a system for parenteral vaccination of fish (Maurice *et al.*, 2003). Since CBMs specificity for carbohydrate ligands is very high, these modules have been used as molecular probes for the analysis of plant cell wall polysaccharides (McCartney *et al.*, 2004). Furthermore, Shoseyov *et al.* (2001) also showed that CBMs could modulate plant growth on transgenic plants. The introduction of a family 3 CBM gene from *Clostridium cellulovorans* in poplars (*Populus tremula*) made the transgenic plants to grow faster than the wild-type, showing an increase in fibre cell length and in the average degree of polymerization of cellulose (Levy and Shoseyov, 2002).

Together with structural and biochemical studies about CBMs and GHs (Chapters 2, 3 and 4), the project presented in this thesis aims to explore some novel biotechnological applications for CBMs. One is to apply CBMs in animal nutrition, to improve the efficacy of cellulases used to supplement barley-based diets for poultry (Chapter 5). The other is to analyse whether CBMs could be used as partner molecules for fixing antimicrobial peptides into cellulose supports (Chapter 6).

1.6.1. Using CBMs to improve the efficacy of enzymes used in poultry nutrition

Poultry diets are based on cereals and, amongst them, corn has always been the elected basis ingredient. Nevertheless, alternative feedstuffs have been explored that could decrease the cost of poultry diets. Animal feeding is and will always be the higher cost in animal production and cheaper raw materials will certainly bring benefits and profits. Barley has been accepted as an alternative cereal for poultry feeding although its anti-nutritional effects are one of the inconvenients precluding its extensive use.

The monosaccharides most commonly present in cereals are hexoses (D-glucose, D-galactose and D-manose), pentoses (L-arabinose and D-xylose) and acidic sugars (D-

galacturonic acid, D-glucuronic acid and its 4-O-methyl ethers). The levels of soluble non-starch polysaccharides (NSPs) in cereals vary between plant species. Barley, for instance, contains high levels of β-glucans. In this cereal, β-glucans contain approximately 70% β-1,4 linkages and 30% β-1,3 linkages, in which segments of two or three β-1,4 linkages are separated by single β-1,3 linkages, although up to five contiguous β-1,3 linkages have also been reported (Choct, 1997).

The role of fibre in monogastric diets has attracted much attention in recent years as soluble NSPs elicit anti-nutritive effects precluding their extensive use in poultry diets (Choct, 1997). Anti-nutritive effects of NSPs in monogastric animals are diverse, but they are mostly associated with their inherent viscosity, their physiological and morphological effects on the digestive tract and their interaction with the intestinal microflora (Choct, 1997). Digesta viscosity alters intestinal transit times and, consequently, the intestinal mucosa surface is affected, while influencing the rate of nutrient absorption. NSPs viscosity depends on solubility and polysaccharide molecular weight and, generally, high gut viscosity decreases the rate of substrates and digestive enzymes diffusion and absorption. Therefore, increased viscosity caused by soluble NSPs is associated with low apparent metabolisable energy (Austin *et al.*, 1999), poor nutrient absorption and high incidence of wet and sticky droppings (Smits and Annison, 1996).

The development of strategies for improving the nutritive value of diets rich in soluble NSPs by monogastrics is, therefore, of great interest. Supplementing poultry diets with exogenous cellulases and hemicellulases, which cause a partial depolymerisation of NSPs to smaller polymers, results in improved performance and contributes to a higher incorporation of NSP rich compounds in monogastric diets. Enzyme supplementation results in a dramatic reduction in digesta viscosity reducing the anti-nutritive effect of water-soluble NSPs (Józefiak *et al.*, 2006). However, enzymes used for feed supplementation usually consist of complex mixtures of biocatalysts, being difficult to recognize the precise biochemical role of each individual enzyme during the hydrolysis of feed polysaccharides (Fontes *et al.*, 2004). Genetic engineering is a valuable tool for producing high concentrations of stable enzymes and the possibility of using individual enzymes in feed supplementation is very useful for the identification of the most efficient molecular architectures and enzyme combinations.

The importance of non-catalytic CBMs in the function of feed enzymes is an emerging theme in animal nutrition. It was recently showed that CBMs can be used to improve the efficiency of associated enzymes in poultry, in a study where wheat and rye-based diets were supplemented with recombinant derivatives of xylanase 11A from *Clostridium thermocellum*, containing or not a family 6 CBM (Fontes *et al.*, 2004). Results from these studies suggest that the modular xylanase containing the family 6 CBM yields better animal performances than its single domain counterpart consisting exclusively on the enzyme's catalytic module. One of the future challenges for the animal feed industry will be to produce even more effective enzymes which will lead to the utilization of both soluble and insoluble NSPs as an energy source for monogastric animals (Choct, 1997). For this particular application, CBMs could have a preponderant role in the development of such biocatalysts. Therefore, another biotechnological application for CBMs could be to improve the efficiency of exogenous enzymes used to supplement monogastric diets. In Chapter 5 an experiment is presented where broilers were fed on a barley-based diet not supplemented or supplemented with a commercial cellulase mixture or with truncated recombinant derivatives of C. *thermocellum* *Ct*Lic26A-Cel5E β-glucanase (from Chapter 4), containing or not a family 11 β-glucan specific CBM.

1.6.2. Using CBMs to fix antimicrobial peptides into cellulosic surfaces

Antimicrobial peptides (AMPs) are molecules with a wide range of antimicrobial effects in bacteria, fungi, parasites and viruses. They are produced by many tissues and cell types in a variety of single-cell organisms, invertebrate, animal and plant species (Brogden, 2005), including mammals (Table 1.5). Hundreds of such molecules have already been identified (http://www.bbcm.units.it/~tossi; April 2008) and over the last years many reviews have been written about AMPs, where lists of antiviral, antibacterial and antifungal peptides are displayed (Brogden, 2005; Brown and Hancock, 2006; Jenssen *et al.*, 2006).

AMPs are usually relatively short (from few to some hundred residues) and their amino acid composition, amphipathicity, cationic charge and size allow them to attach to and disrupt membrane bilayers (Brogden, 2005). AMPs have been classified according to their secondary structures, where four major classes can be identified (Hancock, 1997; Jenssen *et al.*, 2006): (i) amphiphilic peptides with two to four β-strands (*e.g.* human β-defensin-1), (ii) amphiphilic α-helices (*e.g.* magainin-2; PDB

code 2MAG), (iii) loop structures (*e.g.* bactenecin), and (iv) extended structures (*e.g.* indolicidin; PDB code 1G89).

Table 1.5. Properties of antimicrobial peptides isolated from different organisms.

Producing organisms	AMPs	References
Bacteria	bacteriocins: nisin (*Lactobacillus lactis*) and mersacidin (*Bacillus* spp)	Klaenhammer (1988); Chatterjee *et al.* (1992a); Chatterjee *et al.* (1992b); Riley (1998); Jenssen *et al.* (2006)
Invertebrates	cecropins (fly hemolymph); melittin (bee venom); tachyplesin and polyphemusin (horseshoe crab); defensins	Jenssen *et al.* (2006)
Amphibians	magainins; dermaseptins; buforin and buforin II	Zasloff (1987); Simmaco *et al.* (1998); Rinaldi (2002); Jenssen *et al.* (2006)
Mammals	cathelicidins; α-, β-, and θ-defensins; indolicidin; bactenecins	Selsted *et al.* (1992); Selsted *et al.* (1993); Zanetti *et al.* (1995); Falla *et al.* (1996); Travis *et al.* (2000); Lehrer and Ganz (2002); Jenssen *et al.* (2006)
Plants	thionins; defensins	Garcia-Olmedo *et al.* (1998); Jenssen *et al.* (2006)

AMPs have been recognized as an important component of the non-specific host defence system and innate immunity of all animal classes, including insects, amphibians and mammals (Kang *et al.*, 1999). Because of this intrinsic characteristic, some authors have named AMPs as cationic host defence peptides (Brown and Hancock, 2006). AMPs are able to enhance phagocytosis, stimulate prostaglandin release, neutralize the septic effects of lipopolysaccharides from Gram-negative bacteria, promote recruitment and accumulation of various immune cells at inflammatory sites (Yang *et al.*, 2002; Elsbach, 2003), promote angiogenesis (Koczulla *et al.*, 2003) and induce wound repair (Chan and Gallo, 1998). The interests on AMPs are obviously enormous and research is much based on whether they could, in the future, be used as novel drugs to fight infections, and to stimulate the patient's immune system. When disease and/or treatment lower the host stock of AMPs, replenishment with exogenous analogs may be an important adjunctive therapy (Elsbach, 2003). Therefore, AMPs represent very appealing novel molecules for fighting infections and one of their main advantages are that microorganisms do not easily create resistance to them, as it has been described with other types of antimicrobial molecules. This is probably due to AMPs' mode of

action, which is mainly directed to the cellular membrane, disrupting the microbial cellular structure. It is clear that mutations in the overall plasma membrane structure are less likely to occur than variations in the enzyme physiology within a host cell (Lee *et al.*, 2002).

Several authors were able to fuse CBMs with other active molecules using recombinant DNA techniques. Maurice *et al.* (2003) fused a CBM from *Clostridium cellulovorans* with a virulent protein to produce a system for fish parenteral vaccination. In this case the recombinant fusion protein was expressed in *Escherichia coli* BL21 (DE3)pLysS (Novagen). Richins *et al.* (2000) also fused a CBM from *C. cellulovorans* with an organophosphate hydrolase to create an efficient system for purification of organophosphate nerve agents with cellulose materials, expressing the recombinants in *E. coli* BL21 (DE3)pLysE (Novagen). In addition, Hwang *et al.* (2004) fused a CBM from *Trichoderma hazianum* with a lipase, expressing the recombinant proteins in *Saccharomyces cerevisiae* and the same expression agent was used by Lewis *et al.* (2006), who fused a CBM from *Trichoderma reesei* with an antibody. In some works, fusion proteins were produced so that CBMs could be used as purification tags (Doheny *et al.*, 1999; Rechter *et al.*, 1999; Boraston *et al.*, 2001; Kavoosi *et al.*, 2004). Expression of foreign proteins fused to CBMs often results in high level of expression of the fusion protein (Doheny *et al.*, 1999; Otomo *et al.*, 1999; Rechter *et al.*, 1999, Kauffmann *et al.*, 2000; Richins *et al.*, 2000; Boraston *et al.*, 2001).

In early studies involving AMPs, these peptides were obtained by isolation from their original host or by chemical synthesis. This, of course, presented many restrictions to AMPs research. In recent years, many small cationic AMPs have been successfully produced by recombinant DNA methods (Peng *et al.*, 2004; Yang *et al.*, 2004; Čipáková *et al.*, 2006; Xu *et al.*, 2006), which gave larger and more pure quantities of the protein. Many host cells have been used for expressing AMPs but presently *E. coli* has been more and more commonly used because of its fast growth rate and well-established expression systems (Xu *et al.*, 2006). Since AMPs are usually very small to be produced individually, many authors aimed to express them in fusion with other molecules (Peng *et al.*, 2004; Moon *et al.*, 2006; Xu *et al.*, 2006; Xu *et al.*, 2007). The problems of low expression levels, poor solubility, toxicity for the host cell, product proteolysis and poor recovery yields, all very common when it comes to AMP production in *E. coli*, can be overcome by fusion techniques (Wei *et al.*, 2005).

It has been previously demonstrated that CBMs may be attached to a non-related protein without interfering with the target protein biological activity (Tomme *et al.*, 1998). In addition, it has also been confirmed that AMPs produced in *E. coli* fused to a CBM maintained their properties after purification and separation from the fusion partner (Klocke *et al.*, 2005). Some authors fused a CBM with another molecule and aimed to use the fusion molecules in cellulose supports. For instance, a CBM was fused to an organophosphorus hydrolase, with the aim that the resulting fusion molecules could be fixed to protective clothing for people exposed to organophosphates or fixed to fabrics for the decontamination of spills, objects or surfaces exposed to these highly toxic nerve agents (Richins *et al.*, 2000). This immobilized fusion protein successfully degraded an organophosphate compound and its hydrolysis efficiency was maintained over a period of 45 days. Furthermore, the kinetic properties of the fusion recombinant when immobilized onto cellulose or in solution were shown to be similar to the wild type organophosphorus hydrolase. It was suggested that the presence of microcrystalline cellulose had a significant stabilizing effect on the enzyme, much like BSA and glycerol in stabilizing other enzymes (Richins *et al.*, 2000).

From the above discussion, it is clear that fusing AMPs to a CBM with high binding capacity for cellulose would allow fixing these peptides to cotton fabrics, like medical clothes and bandages. To this moment, no report has been made on fusing AMPs with CBMs for this purpose, so the idea expressed here and in Chapter 6 (*Escherichia coli* expression and purification of four antimicrobial peptides fused to a family 3 carbohydrate-binding module from *Clostridium thermocellum*), is medically very appealing.

1.7. OBJECTIVES FOR THIS WORK

Overall, the work presented here aims to elucidate several unresolved questions concerning the function and structure of GHs and CBMs. The specific aims of this project can, however, be resumed on:

1. To study two glycoside hydrolases from the *Clostridium thermocellum* cellulosome: xyloglucanase Xgh74A (Chapter 2) and lichenase *Ct*Lic26A-Cel5E (Chapter 4);

2. To study two novel CBMs from *Ct*Cel9D-Cel44A (formerly known as CelJ) of *Clostridium thermocellum* cellulosome, which were named *Ct*CBM30 and *Ct*CBM44 (Chapter 3);

3. To study the effect of supplementing barley-based diets for broiler chicken with recombinant enzymes derived from lichenase *Ct*Lic26A-Cel5E, referred in 1., including or not its family 11 CBM (Chapter 5);

4. To fuse four antimicrobial peptides to *Ct*CBM3 from *Clostridium thermocellum* CipA, and produce the recombinants in *Escherichia coli*, to exploit its future use as a fixing tag for cellulosic supports (Chapter 6).

CHAPTER 2 Crystal structures of *Clostridium thermocellum* xyloglucanase, Xgh74A, reveal the structural basis for xyloglucan recognition and degradation

Catarina I.P.D. Guerreiro[1*], Carlos Martinez-Fleites[2*], Martin J. Baumann[3], Edward J. Taylor[2], José A.M. Prates[1], Luís M.A. Ferreira[1], Carlos M.G.A. Fontes[1], Harry Brumer[3], and Gideon J. Davies[2]

[1]CIISA, Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; [2]York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, UK; [3]School of Biotechnology, Royal Institute of Technology (KTH), AlbaNova University Centre, 106 91 Stockholm, Sweden

Adapted from *J. Biol. Chem.* (2006) 281, 24922-24933

Abstract

The enzymatic degradation of the plant cell wall is central both to the natural carbon cycle and, increasingly, to environmentally-friendly routes to biomass conversion, including the production of biofuels. The plant cell wall is a complex composite of cellulose microfibrils embedded in diverse polysaccharides collectively termed hemicelluloses. Xyloglucan is one such polysaccharide whose hydrolysis is catalysed by diverse xyloglucanases. Here the structure of the *Clostridium thermocellum* xyloglucanase Xgh74A is presented, in both apo and ligand-complexed forms. The structures, in combination with mutagenesis data on the catalytic residues and the kinetics and specificity of xyloglucan hydrolysis, reveal a complex subsite specificity accommodating seventeen monosaccharide moieties of the multibranched substrate in an open substrate-binding terrain.

[*]Contributed equally to this work

2.1. INTRODUCTION

Plant cell wall polysaccharides are the most abundant carbohydrate polymers in nature and constitute an important renewable natural source of energy available for conversion to biofuels (Demain *et al*., 2005). The plant resource is, however, difficult to exploit primarily because its components are extremely resistant to degradation; plant cell wall polysaccharides are often present as insoluble, often cross-linked, structures in which cellulose is the most abundant component. In the cell wall of flowering plants, cellulose is cross-linked by two major types of glycan: xyloglucans and glucuronoarabinoxylans. The xyloglucans form a complex network of hydrogen-bonded interactions with cellulose microfibrils that confers rigidity and extensibility to the walls of all dicotyledons and about one-half of monocotyledons (Carpita and McCann, 2000). The xyloglucan polysaccharide consists of a linear chain of β-1,4 D-glucan regularly substituted with α-1,6 D-xylosyl units which is, in a species-dependent manner, further derivatized with α-L-arabinose or β-D-galactose (Vincken *et al.*, 1997; Carpita and McCann, 2000), Figure 2.1.

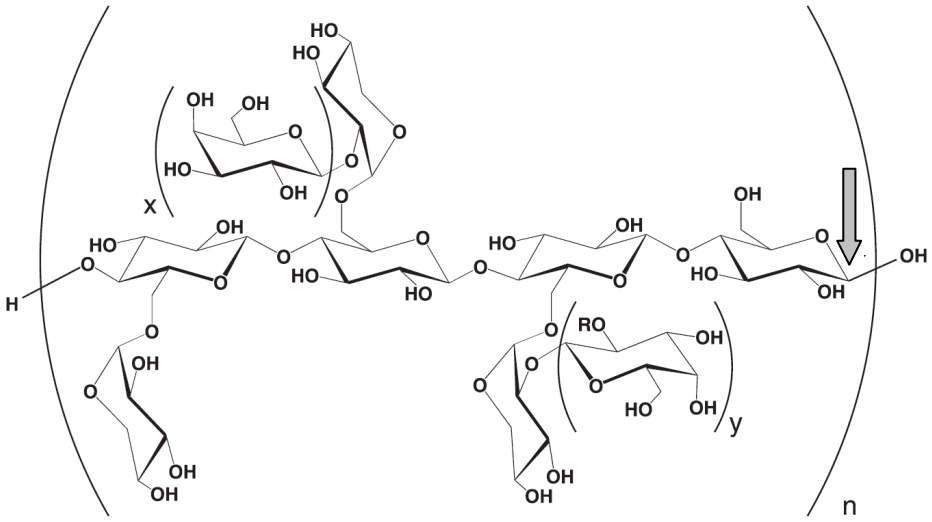

Figure 2.1. Structure of xyloglucan.
Endoglucanases, including Xgh74, typically cleave the glycosidic bond of the unbranched Glc residue (arrow) to yield xylogluco-oligosaccharides. Common structures in seed xyloglucans (*e.g*., *Tamarindus indica*) include XXXG (x=0, y=0, R=H), XLXG (x=1, y=0, R=H), XXLG (x=0, y=1, R=H), and XLLG (x=1, y=1, R=H). XXFG (x=0, y=1, R=α-1,2 L-fucose) is prevalent in xyloglucans from dicot primary cell walls. Oligosaccharide nomenclature is according to Fry *et al.* (1993).

In primary cell wall xyloglucans, the first galactose moiety in the oligosaccharide repeat is commonly substituted with α-1,2 L-fucose (Reiter, 2002). Considerable interest

in the structure, biosynthesis, and enzymatic modification of xyloglucans has been sustained because of the important role these polysaccharides play in plant cell wall morphogenesis (Reiter, 2002; Scheible and Pauly, 2004; Yokoyama *et al.*, 2004; Cosgrove, 2005; Thompson, 2005), as well as the emerging technical applications of xyloglucans in food products, pharmaceutical delivery (Miyazaki *et al.*, 1998; Yamatoya and Shirakawa, 2003), cellulose fibre modification (Brumer *et al.*, 2004; Zhou *et al.*, 2005; Zhou *et al.*, 2006) and biofuel production (Demain *et al.*, 2005).

Microorganisms have evolved sophisticated mechanisms to degrade plant cell wall polysaccharides and consequently exploit this rich carbon and energy source. Aerobic bacteria and fungi secrete several individual enzymes that synergistically degrade plant cell walls (Milkowski *et al.*, 2004). Some anaerobic microorganisms, notably Clostridia, utilize a large multi-enzymatic complex called the cellulosome (Bayer *et al.*, 1985), as explained in Chapter 1. The cellulosome displays a consortium of hydrolytic plant cell wall degrading enzymes, which may change with time, including cellulases, hemicellulases, pectinases and various esterases. The *Clostridium thermocellum* cellulosome is one of the best studied cellulosome systems. It is a multiprotein complex of about 3 MDa and displays endoglucanase, cellobiohydrolase (exoglucanase), xylanase, chitinase and β-glucanase (lichenase) enzymatic activity (Bayer *et al*., 1998). Cellulosome enzymes are tethered to the scaffolding protein of the complex through the interaction of dockerin domains with one of the nine cohesin platforms of the scaffoldin (Carvalho *et al.*, 2003), Figure 2.2A.

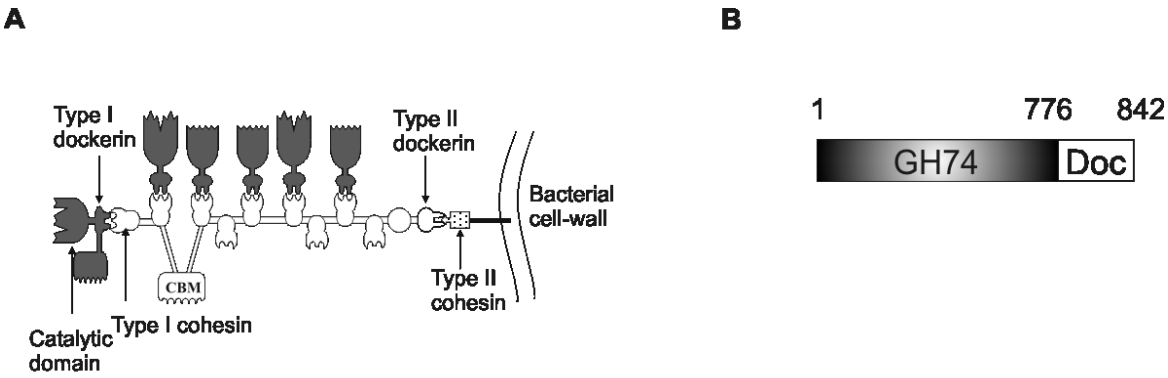

Figure 2.2. (A) Schematic representation of the cellulosome from *Clostridium thermocellum*; (B) xyloglucanase Xgh74A.

There is particular interest in the exploitation of the cellulosome from *C. thermocellum*, primarily because of the potential it offers for the degradation of lignocellulosic waste and subsequent generation of ethanol (reviewed in Demain *et al.*, 2005). To date, 71 open reading frames have been identified as cellulosomal components in this microorganism and about 30 of the encoded enzymes can be ascribed a direct role in cellulose hydrolysis (Zverlov *et al.*, 2005a). About half of the other proteins in the cellulosome are described as hemicellulolytic enzymes, highlighting the importance of these accessory enzymes in the processing of cellulosic composites. Recently, two major enzymes implicated in hemicellulose degradation by the *C. thermocellum* cellulosome have been characterized: an endo-*β*-xylanase and a xyloglucanase (Zverlov *et al.*, 2005b). Xyloglucanase Xgh74A, Figure 2.2B, is the first xyloglucanase identified in *C. thermocellum* and the first active xyloglucanase in a cellulosome and likely plays an important role in the degradation of dicot plant cell wall polysaccharides.

Many xyloglucanases have been classified into GH family 74 (hereafter GH74) in the Carbohydrate Active Enzyme (CAZy) classification (Henrissat, 1991; http://afmb.cnrs-mrs.fr/CAZY; April 2008); recently reviewed in Davies *et al.* (2005). From the structural point of view, the only member of the GH74 family published to date (2006) is the reducing end-specific cellobiohydrolase (OXG-RCBH) from *Geotrichum sp* M128 (Yaoi *et al.*, 2004). OXG-RCBH recognizes the reducing end of various xyloglucan-derived oligosaccharides and releases two glucosyl residues of the type GG, XG or LG (nomenclature according to Fry *et al.*, 1993, see also Figure 2.1) suggesting the presence of at least four subsites (Yaoi and Mitsuishi, 2002). Additionally, it was noted that the glucosyl residue at position +2 (nomenclature according to Davies *et al.*, 1997b) has to be unsubstituted, while the one at -1 should preferentially posses a xylosyl substituent. The OXG-RCBH structure consists of a tandem repeat of two seven-bladed *β*-propeller motifs with the catalytic center formed by the interface of these two domains.

The recent enzymatic characterization of the endoxyloglucanase Xgh74A, shows that this enzyme hydrolyses the glycosidic bond of the unbranched glucosyl residues in xyloglucan, to yield XXXG, XLXG (or XXLG) and XLLG oligosaccharides (Zverlov *et al.*, 2005b). Here the crystal structure of Xgh74A is described, both as an uncomplexed apoenzyme at 2.1 Å resolution, and of an inactive variant, Xgh74-D70A, with a single molecule of XLLG and one of XXLG bound in the active site cleft. These structures, in

light of kinetic and hydrolysis data, reveal the specificity determinants responsible for xyloglucan recognition and provide insight into the hydrolysis of this important plant cell wall polysaccharide.

2.2. EXPERIMENTAL PROCEDURES

2.2.1. Cloning of *Ct*Xgh74A catalytic domain

The cellulosomal xyloglucanase A from *C. thermocellum* is a bi-modular enzyme containing an N-terminal family 74 GH catalytic domain followed by C-terminal dockerin (Zverlov *et al.*, 2005b). To express the xyloglucanase catalytic module in *Escherichia coli*, Xgh74A hereafter, the DNA fragment encoding the protein domain was amplified by polymerase chain reaction (PCR) from *C. thermocellum* YS genomic DNA with the thermostable DNA polymerase *Pfu* Turbo (Stratagene). The primers used are in Table 2.1 and incorporated *Nhe*I and *Xho*I restriction sites. The amplified product was ligated into pMOS*Blue* (GE Health) and sequenced to ensure that no mutations had occurred during the PCR. The recombinant pMOS*Blue* derivative was digested with *Nhe*I and *Xho*I and the excised Xgh74A encoding gene was cloned into the similarly restricted expression vector pET21a (Novagen) to generate pCG1. *Ct*Xgh74A encoded by pCG1 contains a C-terminal His_6-tag.

Table 2.1. Primers used for the cloning and mutagenesis of Xgh74A.

Construct	Primers	Sequence (5'-3')
pCG1	Forward	CTC**GCTAGC**ATTTCCAGCCAGGCTGTA[a]
	Reverse	CAC**CTCGAG**ATCTGAAGCAGGTTCGCC
D70A	Forward	GATTTATGCACGTGCCGCTATCGGAGGAGCGTACC
	Reverse	GGTACGCTCCTCCGATAGCGGCACGTGCATAAATC
D480A	Forward	CTTGTAAGTGCAGTTGGGGCCCTTGTCGGTTTTGTTC
	Reverse	GAACAAAACCGACAAGGGCCCCAACTGCACTTACAAG

[a]The *Nhe*I and *Xho*I restriction sites introduced in the forward and reverse primers, respectively, are shown in bold.

2.2.2. Site directed mutagenesis

Mutants of Xgh74A were generated using a PCR-based QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions and using pCG1 as the template DNA. The sequences of the primers used to generate the protein

mutants are in Table 2.1. The mutated DNA sequences were sequenced to ensure that only the appropriate mutations had been incorporated into the nucleic acid.

2.2.3. Production of recombinant Xgh74A and mutants

E. coli Tuner (DE3) cells (Novagen) harboring the pET21a-Xgh74A plasmid were cultured in Luria-Bertani broth medium (LB) containing ampicillin to mid-exponential phase ($OD_{600\ nm}$=0.6), at which point cultures were transferred to 20 ºC and induced by addition of 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) whereupon they were grown for a further 20 h. Seleno-methionine (SeMet)-labeled Xgh74A was produced in *E. coli* B834 (DE3) (Novagen) containing the pET21a-Xgh74A plasmid with recombinant protein expression induced by 1 mM IPTG and incubation at 20 ºC for 20 h. Cells were harvested by centrifugation and disrupted by sonication in 20 mM HEPES-NaOH, 400 mM NaCl, pH 7.5 buffer. The cell-free extract was incubated at 65 ºC for 15 min and centrifuged to remove insoluble material. Samples were further purified by immobilized metal ion affinity chromatography (IMAC) and buffer exchanged to 10 mM HEPES-NaOH pH 7.5. Xgh74A samples thus purified were assessed pure by SDS-PAGE and were used for crystallization experiments. Xgh74A-D70A and D480A mutants were produced and purified following the native Xgh74A expression and purification protocols.

2.2.4. High-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD)

Oligosaccharides were analyzed on a Waters HPLC system with a Dionex Carbopac PA100 column. A Waters Concorde electrochemical detector was used in PAD mode with a 3-mm gold electrode and a HyREF platinum reference electrode. Two optimized gradients were used for different sizes of xylogluco-oligosaccharides (XGOs).

Gradient A (for analysis of Glc_4-based XGOs) - Solvent A: 100 mM NaOH; solvent B: 100 mM NaOH 200 mM NaOAc. Gradient program: 0-4 min, 100 mM NaOH, 60 mM NaOAc; 4-10 min, linear gradient of 60 mM NaOAc to 117 mM NaOAc, 10-11 min, 200 mM NaOAc. The system was then re-equilibrated for 4 min with the initial conditions prior to the next injection.

Gradient B (for analysis of higher order XGOs) - Solvent A: 100 mM NaOH; solvent B: 100 mM NaOH 500 mM NaOAc. Gradient program: 0-3 min, 100 mM NaOH, 40 mM NaOAc; 3-11 min, linear gradient of 40 mM NaOAc to 170 mM NaOAc; 11-18 min, linear gradient of 170 mM NaOAc to 200 mM NaOAc; 18-19 min, 500 mM NaOAc. The system was then re-equilibrated for 4 min with the initial conditions prior on the next injection.

2.2.5. Mass spectrometry of XGOs

Mass spectrometric analysis was performed with a Q-TOF™ 2 mass spectrometer fitted with a nanoflow ion source (Waters Corporation, Micromass MS Technologies, Manchester, UK). External calibration of the TOF analyser (single-reflectron mode, resolution >10000 FWHM) was obtained over the *m/z* range 50-1000 using a solution of NaI (1.5 g/L) in 1:1 2-propyl alcohol/water. Solutions of XGOs (typical concentration 0.01-0.1 g/L in 1:1 MeOH/water containing 0.5 mM NaCl) were infused into the ion source (3 kV) at 200 nL/min (syringe pump). The cone voltage was varied between at 35 V and 130 V to optimize the intensity of $[M+Na]^+$ and $[M+2Na]^{2+}$ ions. Argon was present in the collision cell at all times, and the collision energy was 10 V. A scan time of 2.5 s with an interscan delay of 0.1 s was used, and continuum data were collected until an acceptable signal-to-noise ratio was achieved after the combination of individual spectra (typically 1-30 spectra).

2.2.6. Preparation of XGOs from deoiled tamarind kernel powder

2.2.6.1. Mixture of XGOs based on a Glc_4 backbone (XLLG, XLXG, XXLG, XXXG)

Twenty gram of deoiled tamarind kernel powder (Saiguru Food Gum Manufacturer, Mumbai, India) were suspended in ammonium acetate buffer (1 L; 10 mM; pH 4.5) at 60 ºC and was vigorously stirred until homogenous. Then the suspension was cooled to 30 ºC and crude cellulase from *Trichoderma reesi* (Fluka) was added (100 mg, 500 U). The resulting solution was incubated at 30 ºC during 18 h under gentle stirring. The progress of the digestion was monitored by HPAEC-PAD (gradient B). The solution was filtered on a glass fibre filter, 1 mL NH_3 (37% in H_2O) was added, and the basic solution was pumped over a Q-sepharose (GE Healthcare) column (10 cm high, 2.6 cm diameter) to remove the cellulase. The resulting solution was freeze-dried to yield a mixture of XGOs as a white powder (typical yield 8-9 g). The following

oligosaccharide composition was obtained by HPAEC-PAD (gradient A): XXXG, XLXG, XXLG, XLLG (2:1:3:3).

2.2.6.2. Mixture of higher-order XGOs (Glc_8-Glc_{16} backbone)

A modified protocol based on that described by Vincken *et al.* (1995) was used. One gram of deoiled tamarind kernel powder was dissolved in ammonium acetate buffer (50 mL; 10 mM; pH 4.5) at 60 ºC for one hour. The solution was cooled to 30 ºC and 1 mg (1 unit) of crude cellulase (*T. reseei*, Fluka) was added. The progress of the digestion was monitored by HPAEC-PAD (gradient B). When HPAEC-PAD analysis indicated the presence of predominantly Glc_4 and Glc_8 oligosaccharides, the digestion was stopped (16 h) by boiling for 30 min. The solution was cooled to room temperature, filtered on a glass fibre filter, and the filtrate concentrated *in vacuo* to a volume of 10 mL. The oligosaccharides were separated by size exclusion chromatography on two Bio-Gel P6 (Bio-Rad) columns (2×90 cm, 2.6 cm diameter) connected in series and maintained at 60 ºC. The products were eluted with a flow rate of 0.5 mL/min with ultrapure water. Fractions (5 mL) were analyzed by HPAEC-PAD (gradient B), pooled; concentrated *in vacuo* and finally freeze dried (typical yield 130 mg Glc_4 oligosaccharides, 190 mg Glc_8 oligosaccharides, 98 mg Glc_{12} oligosaccharides, 20 mg Glc_{16} oligosaccharides and 8 mg Glc_{20} oligosaccharides).

2.2.6.3. XXXGXXXG and partially degalactosylated Glc_{12}-based XGOs

Degalactosylated higher order oligosaccharides were produced exactly as described in the preceding paragraph, except that 4 units of β-galactosidase (*Aspergillus niger*, Megazyme®, Ireland) was added immediately after boiling and cooling the crude cellulase digestion mixture. The degalactosylation reaction was carefully controlled by HPAEC-PAD analysis (gradient B) because of contaminating isoprimeverase activity in the commercial β-galactosidase (Kato *et al.*, 1985).

2.2.7. Crystallization, data collection, structure solution and refinement

Crystals of SeMet-Xgh74A and Xgh74A-D70A in complex with xyloglucan-derived oligosaccharides were grown by the hanging-drop method. SeMet-Xgh74A (10 mg/mL) was crystallized in 5 mM $CdCl_2$, 12% PEG 4000, 100 mM sodium cacodylate, pH 6.5. Xgh74A-D70A-XXLG complex crystals were prepared by co-crystallization of

the Glc_4 XGO mixture and the inactive mutant (46 mg/mL) in 200 mM KSCN, 22% PEG 3350, 100 mM HEPES-NaOH, pH 7.5. In both cases, crystals were cryoprotected by the addition of 20% glycerol (v/v) to the crystallization conditions. A 2.1 Å resolution dataset was collected from an uncomplexed SeMet-Xgh74A crystal at -173 ºC on beamline ID14.3 (European synchrotron radiation facility, ESRF, Grenoble, France). Diffraction data of the Xgh74A-D70A-XXLG complex were collected on beamline ID23.1 (ESRF). Both data sets were integrated with MOSFLM and scaled and reduced with SCALA from the CCP4 suite of programs (Collaborative Computational Project Number 4, 1994).

Although SeMet data were collected, Xgh74A structure was solved by molecular replacement because the structure of the homologous *Geotrichum* sp. M128 OXG-RCBH (Protein Data Bank, PDB, code 1SQJ) became available at the time these studies were carried out. The program PHASER (McCoy *et al.*, 2005) was used to place the 4 molecules of Xgh74A in the asymmetric unit. The Xgh74A model was built and refined, against the SeMet data using Arp-wARP (CCP4, 1994). Subsequently, the refined SeMet apo enzyme structure was used as a molecular replacement model for the solution of the Xgh74A-D70A ligand complex, which crystallized in a $P4_32_12$ form with two molecules in the asymmetric unit. Manual rebuilding and ligand placement were carried out with COOT (Emsley *et al.*, 2004). Solvent model was build with Arp-wARP with maximum likelihood refinement using REFMAC (Murshudov *et al.*, 1997).

2.2.8. Activity of Xgh74A on xyloglucan

2.2.8.1. Kinetics of Xgh74A

Reducing sugars released from the hydrolysis of plant cell wall polysaccharides by Xgh74A were quantified following the method described by Nelson (Nelson, 1944) and Somogyi (Somogyi, 1945). Enzyme activities were measured in 25 mM potassium phosphate buffer, pH 7.0, at 50 ºC. The assay mixture (100 µL) contained 10 µL of enzyme (with appropriate dilution) and concentrations of substrate ranging from 0.05 to 15 g/L. Activities were determined in triplicate in the linear range of the reactions.

2.2.8.2. Limit digest of tamarind xyloglucan and Glc_4-, Glc_8-, and Glc_{12}-based XGOs by Xgh74A

Tamarind xyloglucan (1 g/L) and Xgh74A (0.02 mg/L) were incubated in 20 mM potassium phosphate buffer pH 7.0 (total volume 25 µL) at 50 ºC for 30 min. The

reaction was stopped by incubation at 95 ºC for 10 min. 10 µL of the sample was analyzed by HPAEC-PAD (gradients A and B). A mixture of Glc_4-based XGOs (XXXG, XLXG, XXLG, XLLG) (3.5 g/L) and Xgh74A (2.3 g/L) was incubated in 20 mM potassium phosphate buffer pH 7.0 (total volume 25 µL) at 50 ºC for 30 min. The reaction mixture was then diluted 1:100 into ultrapure water and directly injected for HPAEC-PAD analysis (gradient A). Glc_8- (6.3 g/L) or Glc_{12}-based (9.5 g/L) XGOs and Xgh74A (0.02 g/L) were incubated in 20 mM potassium phosphate buffer pH 7.0 (total volume 25 µL) at 50 ºC for 30 min. The reaction mixture was then diluted 1:100 into ultrapure water and directly injected for HPAEC-PAD analysis (gradients A and B). For each of the above cases, comparative experiments were carried out by substituting *Trichoderma longibrachiatum* endoglucanase (0.25 units, EGII, Lot 50201, Megazyme®, Ireland) for Xgh74A under identical conditions.

2.3. RESULTS AND DISCUSSION

The *Clostridium thermocellum* β-1,4-xyloglucan hydrolase Xgh74A is a 842-residue protein that consists of a N-terminal catalytic module (residues 1-776) and a C-terminal dockerin module (residues 777-842), as presented in Figure 2.2B. For structural and kinetic studies, the Xgh74A catalytic module was cloned into the pET21a vector from residue Ile^{28} to Glu^{762} and overexpressed in *E. coli* Tuner cells (Novagen).

2.3.1. Kinetics and specificity of Xgh74A

Xgh74A was active on tamarind xyloglucan, on lichenan, and on the artificial substrates carboxymethyl cellulose (CMC 4M, Megazyme®, Ireland) and hydroxyethylcellulose (HEC), Table 2.2.

Table 2.2. Kinetics of Xgh74A on different polysaccharides.

Polysaccharide	k_{cat} [a]	K_m	k_{cat}/K_m
	min^{-1}	g/L	$(g/L)^{-1}\ min^{-1}$
Xyloglucan	47.54	0.76	62.55
CMC 4M	43.19	4.61	9.37
Lichenan	35.69	6.57	5.43
HEC	13.05	5.10	2.55

[a] k_{cat} are in molecules of product per molecule of enzyme per minute.

Xgh74A shows considerably greater catalytic efficiency on xyloglucan (k_{cat}/K_m 63 $(g/L)^{-1}$ min^{-1}) than on any of the other substrates including CMC 4M (k_{cat}/K_m 9 $(g/L)^{-1}$ min^{-1}) which, like xyloglucan, is substituted at the 6 position, in this case with an average degree of substitution of 4 carboxymethyl groups every 10 sugars. The enzyme showed no activity on xylan, polygalacturonic acid, wheat arabinoxylan, rhamnogalacturan, curdlan, laminarin, galactomannan, galactan and arabinan with a small unquantifiable activity on glucomannan. Taken together the results are strongly suggestive that Xgh74A is a true xyloglucanase.

Limit digest analysis of tamarind xyloglucan hydrolysis by Xgh74A indicates the liberation of XXXG, XLXG, XXLG and XLLG (Figure 2.3) as the major products.

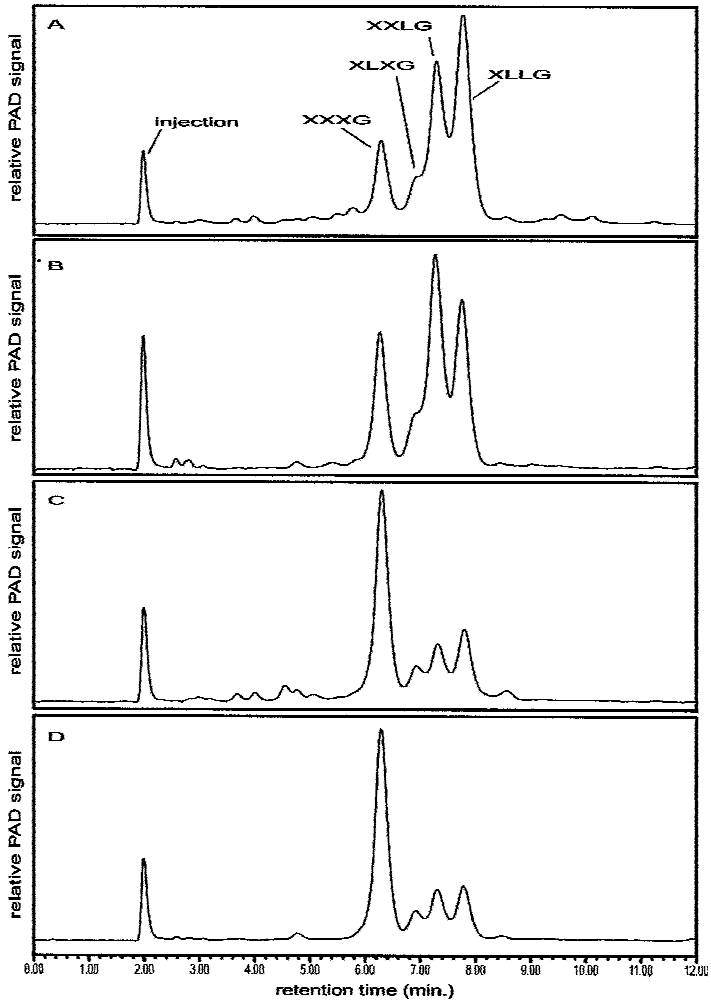

Figure 2.3. HPAEC-PAD analysis (gradient A) of endoglucanase hydrolysis products.
(A) Xgh74A limit digest of tamarind xyloglucan. (B) *Trichoderma* EGII limit digest of tamarind xyloglucan. (C) Xgh74A limit digest of partially degalactosylated Glc_{12}-based XGOs. (D) *Trichoderma* EGII limit digest of partially degalactosylated Glc_{12}-based XGOs.

Longer oligosaccharides did not accumulate, but were observed as transient species by HPAEC-PAD. In contrast to previous work on this enzyme (Zverlov *et al.*, 2005), no evidence was seen for the production of XXG, XXX, or XXGG by mass spectrometry (Figure 2.4). It can be speculated that the acidic ESI conditions and/or different ion optics employed by Zverlov *et al.* (2005b) may have contributed to the formation of fragment ions in the ESI source. Here, the use of MeOH/H_2O/NaCl as an ESI solvent minimized oligosaccharide fragmentation and yielded exclusively $[M+Na]^+$ or $[M+2Na]^{2+}$ adducts, depending upon the applied cone voltage (Figure 2.4).

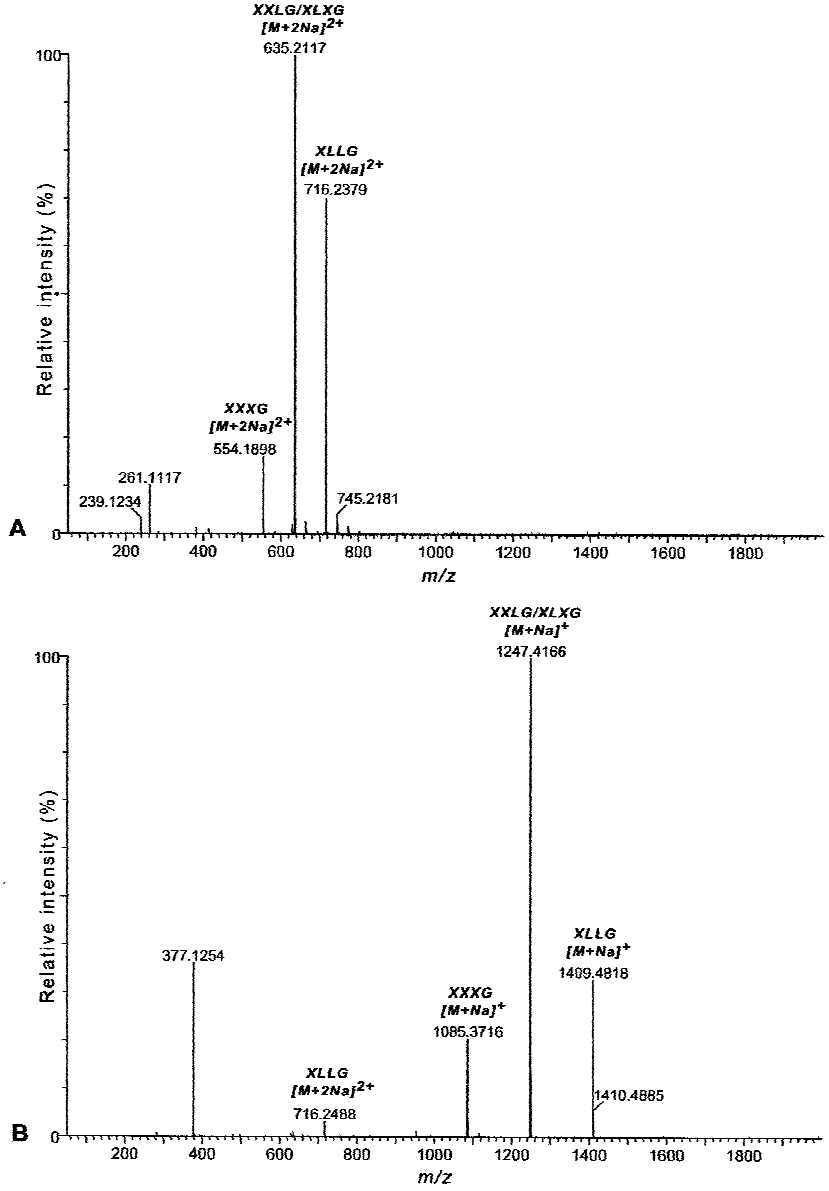

Figure 2.4. Mass spectrometric analysis of the hydrolysis products of Xgh74A acting on tamarind xyloglucan.
(A) cone voltage 35 V, (B) cone voltage 130 V.

To investigate whether Xgh74A cleaves tamarind xyloglucan at a position other than the anomeric carbon of the unbranched glucosyl moiety (Figure 2.1), limit digest experiments were performed on XGOs based on Glc_4-, Glc_8-, and Glc_{12} backbones. Cleavage of the β-1,4 glycosidic bond between two branched Glc units bearing α-1,6 Xyl residues has been previously reported in the related GH74 OXG-RCBH from *Geotrichum* sp. (Yaoi and Mitsuishi, 2002). Incubation of a high concentration of Xgh74A with a 2:1:3:3 mixture of XXXG/XLXG/XXLG/XLLG showed no formation of smaller oligosaccharides by HPAEC-PAD (data not shown) demonstrating that Xgh74A, in contrast to OXG-RCBH, cannot cut at substituted glucosyl moieties, at least on these short substrates. Similarly, incubation of Xgh74A with XXXGXXXG or a variably galactosylated Glc_8 XGO mixture, yielded only XXXG or the expected XXXG/XLXG/XXLG/XLLG mixture, respectively, as determined by HPAEC-PAD (data not shown). Likewise, Xgh74A digestion of a partially degalactosylated Glc_{12} XGO mixture yielded predominantly XXXG and minor amounts of other Glc_4-based XGOs (Figure 2.3). In all cases the action of Xgh74A, as determined by HPAEC-PAD and/or MS analysis, was indistinguishable from that of commercially available *T. longibrachiatum* endoglucanase (Figure 2.3 and Figure 2.4).

2.3.2. Structure of *Clostridium thermocellum* xyloglucanase Xgh74A

The structure of the catalytic module of Xgh74A was solved in a $P2_1$ crystal form by molecular replacement using the homologous *Geotrichum* sp. M128 OXG-RCBH (PDB code 1SQJ) as a search model. The final Xgh74A model, an apoenzyme incorporating SeMet in place of methionine, comprises residues Val^{33} to Ser^{760} and was refined to crystallographic R-factors of 17.5% (R_{cryst}) and 21.0% (R_{free}) with diffraction data to a resolution of 2.1 Å. Data collection and refinement statistics are summarized in Table 2.3. The asymmetric unit contains four copies of the polypeptide chain that can be superimposed with an average root-mean-square deviation (RMSD) of 0.25 Å showing no significant conformational differences due to crystal packing. Some of the contacts between molecules in the asymmetric unit are mediated through the coordination of cadmium ions (added as a crystallization component).

Table 2.3. Data collection and refinement statistics for Xgh74A and Xgh74A-D70A-XLLG-XXLG complex.

	Xgh74A	**Xgh74A-D70A-XLLG-XXLG complex**
Data Collection		
Space Group	$P2_1$	$P4_32_12$
Cell dimensions		
a, b, c (Å)	100.4, 97.9, 199.1	141.4, 141.4, 193.4
α, β, γ (º)	90.0, 97.6, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.93100	0.97565
Resolution	45.79-2.10 (2.21-2.10)	54.15-1.95 (2.06-1.95)
R_{merge}	0.121 (0.515)	0.148 (0.533)
$\langle I/\sigma_I \rangle$	9.9 (2.5)	14.3 (4.6)
Completeness (%)	99.0 (98.3)	100.0 (100.0)
Redundancy	3.9 (3.9)	11.5 (11.5)
Refinement		
Resolution (Å)	19.90-2.10	54.15-1.95
Number of reflections	208,964	135,354
R_{cryst}/R_{free}	0.18/0.21	0.18/0.20
Number of atoms		
Protein	21792	11294
Ions (Cd^{2+})	7	NA[b]
Ligand	NA[b]	354
Water	2138	1277
B-factors		
Protein	19.7	15.0
Ions (Cd^{2+})	25.8	NA[b]
Ligand	NA[b]	
XLLG (minus subsites)	NA[b]	16.0
XXLG (positive subsites)	NA[b]	19.0
Water	26.4	25.5
RMSD		
Bond lengths (Å)	0.012	0.010
Bond angles (º)	1.175	1.251
Ramachandran statistics (%)[a]	97/3/0.07	96.5/3.5/0
PDB code	2CN2	2CN3

[a]Ramachandran statistics indicate the fraction of residues in the most favored, additionally allowed and disallowed regions of the Ramachandran diagram, as defined by the program MOLPROBITY (Lovell *et al.*, 2003). [b]NA, not applicable.

Xgh74A consists of two seven-bladed β-propeller domains (Figure 2.5A), as expected from the sequence homology with OXG-RCBH. The N-terminal domain of Xgh74A comprises residues 63 to 459 whilst the C-terminal domain involves residues 33-62 and 460-760. Similarly to OXG-RCBH, the Xgh74A N-terminal domain is orientated at angle of approximately 90 degrees relative to the C-terminal domain and interactions between these domains occur primarily through hydrogen-bonding and hydrophobic interactions over a shared contact area of about 7530 $Å^2$. The N- and C-terminal domains, which can be superimposed with an RMSD of 3.1 Å, exhibit 19% sequence identity, which most likely reflects an ancient gene duplication event. The Xgh74A N- and C-terminal domains are connected by two loop segments, one located

in the N-terminus and the other in the middle of the sequence. In the OXG-RCBH structure, the N- and C-terminal domains are linked by three segments; one in the N-terminus, the second in the middle and the third in the C-terminus of the sequence (this C-terminal segment adds a fifth strand to the second blade of the N-terminal propeller which is absent in Xgh74A).

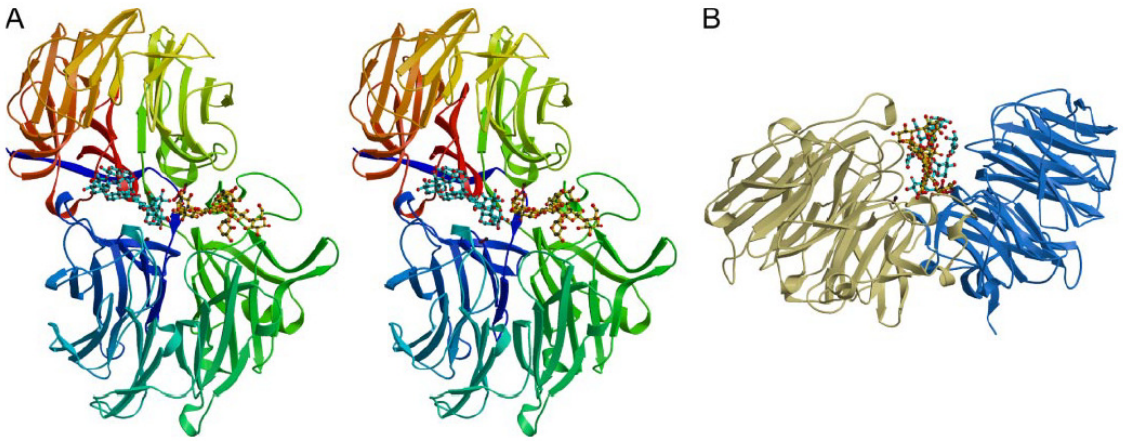

Figure 2.5. Three-dimensional structure of *C. thermocellum* xyloglucanase Xgh74A.
(A) divergent (wall-eyed) stereo cartoon of the structure, color-ramped from N-terminus (blue) to C-terminus (red) with the ligands in ball-and-stick representation. (B) mono view of the structure showing the N- and C-terminal domains, presumably resulting from an ancestral gene duplication event, colored khaki and blue, respectively. These figures were drawn with MOLSCRIPT (Kraulis, 1991).

The overall topology of Xgh74A is thus very similar to OXG-RCBH with all the secondary structure elements linked through identical connectivity. The Cα traces superimpose with a RMSD of 1.8 Å for 664 equivalent residues, reflecting 39% sequence identity (calculation performed with DALI; Holm and Sander, 1993). Not surprisingly, greater structural divergence is found in the loops connecting the blades of the β-propeller architecture. The most important of these differences is the different conformation adopted by the loop Thr^{397}-Pro^{406} in Xgh74A compared with its structural equivalent Asn^{374}-Thr^{391} in OXG-RCBH, which may contribute to the significantly different substrate specificity of the two enzymes. These structural details are discussed below, in the light of the ligand complexes of Xgh74A.

2.3.3. Active site structure

It is immediately apparent, Figure 2.5, that the substrate-binding region of Xgh74A lies in an open cleft. This groove is formed at the intersection of the N- and C-terminal domains. The surface of the cleft is formed by the loops connecting the β-propeller blades in both domains. In the apo Xgh74A structure some of these loops (Tyr^{206}-Asp^{217}, Thr^{291}-Asn^{298} and Asp^{524}-Asp^{527}) are disordered, whereas in the ligand complexed forms they become ordered and participate in substrate binding (described below). Catalysis by family GH74 enzymes occurs with inversion of anomeric configuration, *i.e.*, the stereochemistry of the product is inverted with respect to the β-linkage of the substrate. A classical interpretation of glycoside hydrolysis with inversion of anomeric configuration implicates two key residues: a catalytic acid, to facilitate leaving-group departure by protonation and a catalytic base, to activate the incoming water molecule for nucleophilic attack by deprotonation (glycosidase catalytic mechanisms are reviewed by Davies *et al.*, 1997a). In GH74 enzymes, it is believed (Yaoi *et al.*, 2004) that two aspartate residues play the role of Brønsted acid and base; in Xgh74A these are believed to be Asp^{480} and Asp^{70}, respectively. Asp^{70} and Asp^{480} are located in the middle of the active centre cleft, lying on opposite sides, and deep within the cavity with their carboxylate groups approximately 10 Å apart. Site-directed mutagenesis of either of these pivotal residues, to alanine, results in an inactive enzyme, which within the sensitivity of the assay suggests at least 1000-5000 times less activity than wild-type enzyme. Asp^{70} is located in the middle of the loop connecting the second and third strand of the first propeller blade of the N-terminal domain. The peptide sequence in this region is strictly conserved among the members of the GH74 family. In the apoenzyme, Asp^{70} forms a hydrogen bonding interaction with the side-chain of Glu^{459} and two molecules of water. Asp^{480} is located in the C-terminal domain in an equivalent position to Asp^{70}. However, in contrast to Asp^{70}, Asp^{480} does not form H-bonds with protein atoms and points directly into the cleft. It is not immediately apparent what might contribute to an elevated pK_a for this catalytic acid in the absence of bound substrate.

2.3.4. Structure of the Xgh74A-D70A mutant in complex with Glc_4-based XGOs

To probe the structural determinants of xyloglucan recognition we attempted to co-crystallize Xgh74A and two inactive Xgh74A variants (D70A and D480A) with preparations of XGOs based on Glc_4 backbones. Crystals of the Xgh74A-D70A mutant,

complexed with a mixture of Glc_4–based oligosaccharides, were obtained by co-crystallization and diffracted to 1.95 Å resolution. Crystals were indexed in the $P4_32_12$ space group and contained two molecules in the asymmetric unit. The complex structure is essentially identical to the ordered parts of the apo structure resulting in an RMSD value of 0.3 Å for the Cα atoms. In the ligand complex structure, however, it is also possible to build the previously disordered loops, which all interact directly with the bound oligosaccharide. The electron density map displays well-defined density for seventeen sugar rings, corresponding to a molecule of XLLG and another of XXLG, either side of the catalytic centre (Figure 2.6). The two molecules sit in an extended conformation at the bottom of the cleft on both sides of the catalytic residue Asp^{480}. It is possible that the desolation afforded by ligand binding contributes to the pK_a elevation of the catalytic acid. The glucosyl backbones extend about 20 Å in opposite directions from the centre point described by the Asp^{480} residue. The interaction of the two-ligand molecules extends over an area of approximately 316 $Å^2$.

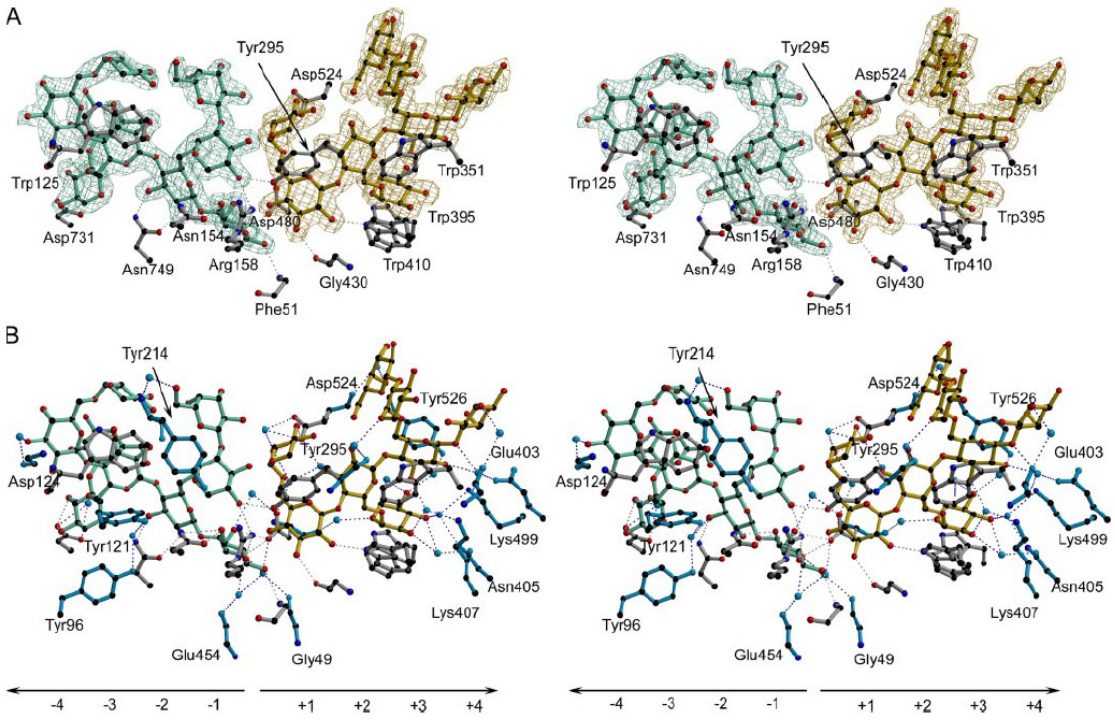

Figure 2.6. Interactions of *C. thermocellum* xyloglucanase Xgh74A (D70A) with two XGOs.
(A) Observed $2F_{obs}$-F_{calc} electron density for the ligand with interacting residues, drawn with BOBSCRIPT (Esnouf, 1997); "negative" subsite sugars are shown in cyan with "positive" (leaving group) subsite sugars in yellow; residues making direct interactions with the ligands are shown in grey. (B) Ligand binding to Xgh74A (D70A); in this case, solvent-mediate hydrogen bonds are included with their protein partners colored in blue and with the residues making direct contacts remaining in grey.

Despite co-crystallization with a mixture of oligosaccharides (in relative proportions XXXG 2: XLXG 1: XXLG 3: XLLG 3) the species that we observe bound to the enzyme corresponds to XLLG in the "minus" subsites (subsite nomenclature according to Davies *et al.*, 1997b; see Chapter 1) and XXLG in the "positive" leaving-group subsites. Thus, four β-1,4 glucosyl moieties of one XLLG molecule are located in the negative binding sites -1 (Glc^{-1}), -2 (Glc^{-2}), -3 (Glc^{-3}) and -4 (Glc^{-4}) in an extended conformation. Three α-1,6 linked xylose residues branch from the -2 ($Xyl^{-2'}$), -3 ($Xyl^{-3'}$) and -4 ($Xyl^{-4'}$) glucosyl units, with both $Xyl^{-2'}$ and $Xyl^{-3'}$ also bearing a β-1,2 linked galactosyl ($Gal^{-2''}$) unit, the latter partially disordered.

The mean temperature factor for this oligosaccharide is 16 $Å^2$ and its interaction area with the enzyme is approximately 162 $Å^2$. Glc^{-1} is positioned in the middle of the diagonal line connecting the Cα atoms of the catalytic residues Asp^{70} and Asp^{480} (Ala^{70} in the complex structure). Ala^{70} in the Xgh74A-D70A mutant lies below the plane of the Glc^{-1} ring at a distance of about 5.8 Å between Glc^{-1} C1 and Ala^{70} Cα, consistent with the position demanded for the catalytic base in an inverting mechanism. Asp^{480} is located above this plane at a distance of 4.3 Å. Glc^{-1} forms a network of H-bonding interactions where all its oxygen atoms except O4 are involved (Figure 2.6 and Figure 2.7). Glc^{-1} O1, O2 and O3 atoms interact with the nitrogen main chain of Phe^{51}, Arg^{158} side-chain and Asn^{154} side-chain respectively, whereas the putative catalytic acid, Asp^{480}, forms H-bonds with Glc^{-1} O5 and O6. All the side-chains involved in the recognition of Glc^{-1} appear conserved in the multiple sequence alignment of GH74 family, Figure 2.8. Superposition of the Xgh74A and OXG-RCBH structures shows similar environments around the Glc^{-1} O6 position (-1' subsite). In both structures, the loops enclosing this region adopt similar conformations and some of the side-chains lining the cavity are conserved, leaving room to accommodate a xylose at the -1' position. Accordingly, it is not evident what constitutes the structural basis supporting the fact that endoxyloglucanases prefer to cleave the xyloglucan chain at unbranched glucosyl positions (-1) in contrast with the ability of exoxyloglucanases to process substrates with a xylose ramification at the -1' position (Yaoi and Mitsuishi, 2002; Bauer *et al.*, 2005).

At subsite -2, the plane of the Glc^{-2} ring is rotated 180 degrees relative to the plane of Glc^{-1} contacting the enzyme through two H-bonding interactions mediated by Glc^{2-} O2 and O3 with the side-chain of Asn^{749}. $Xyl^{-2'}$ ring is located almost parallel to the plane of Tyr^{214} aromatic ring and its interaction with the protein is mediated by two

H-bonds between Xyl$^{-2'}$ O4 and the side-chain oxygen of Tyr^{295} and side-chain NH1 of Arg^{158}. Gal$^{-2''}$ does not contact the protein directly but forms an H-bond interaction with a water molecule which in turn interacts with the main chain nitrogen atom of Tyr^{214}. The Gal$^{-2''}$ moiety does not appear to be specifically recognized by the enzyme; instead it is located close to the border of the cleft and exposed to the solvent.

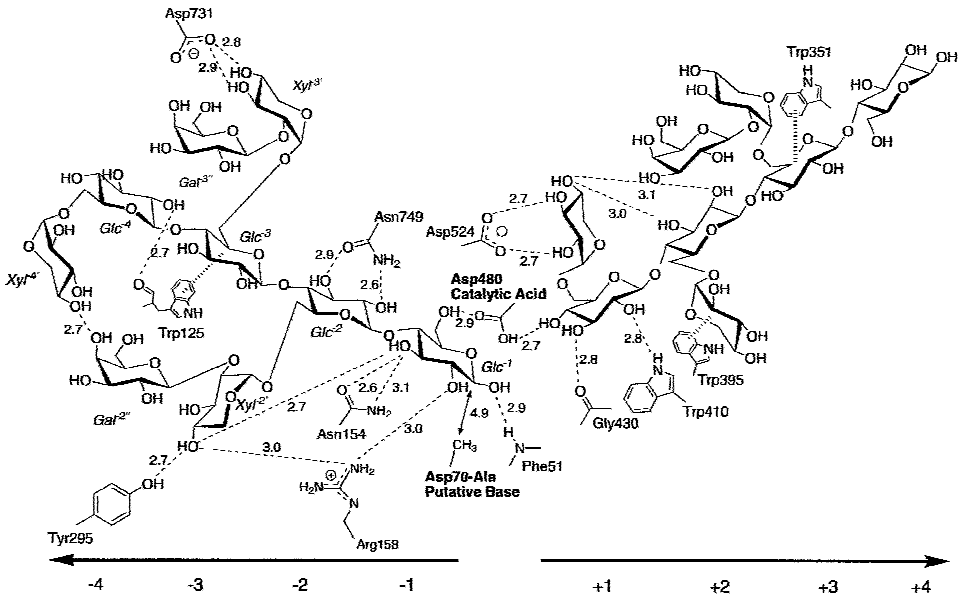

Figure 2.7. Schematic diagram of the direct interactions of Xgh74A (D70A mutant) with two xyloglucan-derived oligosaccharides.

The -3 subsite glucoside is "stacked" on the side-chain of Trp^{125} and does not form any H-bonding interactions with the enzyme. The Xyl$^{-3'}$ ring in Xgh74A describes an angle of approximately 90 degrees with respect to the plane of Glc^{-3} ring pointing directly to the middle of the cleft in the N-terminal to C-terminal direction. This residue forms two H-bonds between its O3 and O4 atoms and the side-chain of Asp^{731}. An additional partially occupied Gal (Gal$^{-3''}$) residue was modelled in the electron density adjacent to Xyl$^{-3'}$ O2 atom. This Gal$^{-3''}$ residue is located at the entrance of the cleft and exposed to solvent. Gal$^{-3''}$ does not interact directly with any protein atom but forms one H-bond between its O4 and the O3 of Glc^{-3}. This structural feature is in agreement with the described absence of specificity towards Gal residues on positions -2 and -3. The sugar moieties at position -4, Glc^{-4} and Xyl$^{-4'}$ are located at the periphery of the cleft. The plane of the Glc^{-4} ring lies at an angle of 180 degrees with respect to the plane of

Glc^{-3}. This glucosyl unit contacts the enzyme through an H-bond between O2 and main chain oxygen of Trp^{125}. The $Xyl^{-4'}$ residue does not contact any protein atom and only makes H-bonding interactions between its O4 atom and $Gal^{-2'}$ O4 and Glc^{-2} O2 atoms.

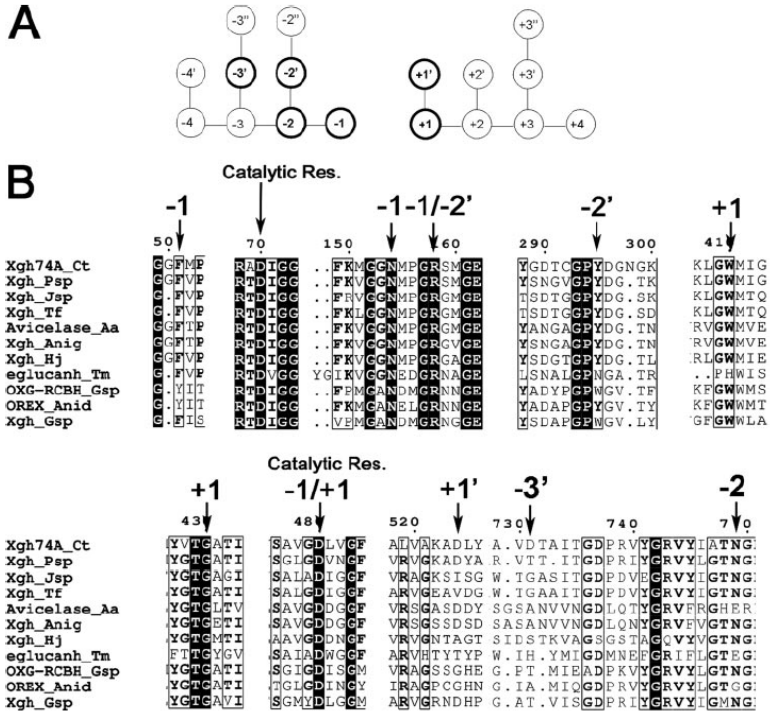

Figure 2.8. Sequence conservation around the XGOs recognition sites in family GH74.
(A) Schematic representation of the sugar subsites specifically recognised with Xgh74A (bold circles). (B) Fragments of the multiple sequence alignment of members of family GH74 whose enzymatic characterization has been reported. The sequence alignment was calculated with ClustalW (Thompson *et al.*, 1994) and represented with ESPript (Gouet *et al.*, 1999). Abbreviations: *Ct*, *Clostridium thermocellum*; *Psp*, *Paenibacillus* sp KM21; *Jsp*, *Jonesia* sp DSM 14140; *Tf*, *Thermobifida fusca* YZ; *Aa*, *Aspergillus aculeatus*; *Anig*, *Aspergillus niger*; *Hj*, *Hypocrea jecorina* QM6a; *Tm*, *Thermotoga maritime*; *Gsp*, *Geotrichum* sp M128; *Anid*, *Aspergillus nidulans* FGSC A4.

In the positive or "leaving-group" subsites of Xgh74A, a second XXLG molecule binds in an extended conformation in which an imaginary line drawn through the glucosyl residues at the positive subsites makes an angle of about 60 degrees with respect to the trajectory of the glucosyl backbone bound in the negative subsites, *i.e.*, the chain is bent. The interaction area with the enzyme is smaller (154 $Å^2$) than the area of contact over the minus subsites. Direct H-bonding contacts between this molecule and the protein atoms are also fewer and are confined just to the +1 site. Catalytic acid, Asp^{480}, contacts the glucosyl unit at the +1 and -1 sites forming H-bonding interactions with the O4 of Glc^{+1} unit consistent with its role in aiding departure of this group,

through proton donation, during catalysis. The +1 sugar residue also interacts with the aromatic nitrogen of Trp^{410} (O2) and the main chain oxygen of Gly^{430}. Asp^{524} forms two H-bonding interaction with O2 and O3 of the $Xyl^{+1'}$ residue and additionally, the O5 atom of this residue contacts the O3 of $Xyl^{+2'}$.

The glucosyl backbone starts to emerge from the binding cleft from site +2. None of the sugar residues at sites +2, +3 and +4 makes direct H-bonding contacts with protein atoms but are found to be involved in a complex network of interactions mediated by water molecules as is represented in Figure 2.6B. In the Xgh74A complex described here there is no density, not even at very low levels, indicative of a galactosyl residue attached to the +2' xyloside. Indeed, inspection of the structure would suggest that there are steric blockages for the accommodation of a +2" galactosyl moiety. The interpretation of limit digest patterns, in particular the observation of XLLG, demands that the +2" region must be able to accommodate a galactosyl moiety during catalysis. One possibility is that the binding mode for XLLG is more flexible, at either protein or ligand levels, than that observed here for XXLG.

2.3.5. Conservation of xyloglucan recognition sites in members of GH74 family

GH74 family groups enzymes that are able to hydrolyse xyloglucan oligosaccharides but also are active on non-branched substrates like barley β-glucans (β-1,3/1,4 glucan), CMC, Avicel (microcrystalline cellulose) or galactomannan (β-1,4 mannose). A wide spectrum of activities is observed among members of this family. This spectrum covers enzymes that can, apparently, only process xyloglucan up to enzymes that actually prefer non-branched substrates such as barley β-glucan. For example, *Paenibacillus* sp KM21 (Yaoi *et al.*, 2005) and *Thermobifida fusca* YZ (Irwin *et al.*, 2003) xyloglucanases are only active on xyloglucans from various sources but not on non-branched polymers such as barley β-glucan, CMC, Avicel or xylan. *Geotrichum* sp M128 xyloglucanase displays its highest activity on the more branched xyloglucans from tamarind or pea, but it is less efficient on barley xyloglucan that contains fewer xylose decorations and the enzyme shows no activity on non-branched substrates (Yaoi *et al.*, 2004). At the other end of the spectrum, *Thermotoga maritima* Cel74 shows its highest activity on barley β-glucan and is about 75% less efficient on tamarind xyloglucan (Chhabra and Kelly, 2002).

The specificity of the interaction between xylo-oligosaccharide ligands and Xgh74A interpreted in light of direct H-bonding contacts with the enzyme is reduced to

positions -3', -2', -2, -1, +1 and +1' (Figure 2.8). Apart from the strict conservation of the catalytic residues Asp^{70} and Asp^{480} and their sequence equivalents in the GH74 family, the residues responsible for the recognition of the glucosyl units at positions -2, -1 and +1 appear highly conserved in the multiple sequence alignment of GH74 members (Figure 2.8). Only two substitutions are observed at the glucosyl recognition site -1 and +1 in the multiple sequence alignment. The first one is in the equivalent position to Xgh74A Phe^{51} where a Tyr residue appears in *Geotrichum* sp OXG-RCBH and in *Aspergillus nidulans* oligoxyloglucan reducing end-specific xyloglucanobiohydrolase (OREX). This interaction is not mediated by the side-chain but through an H-bond between the carbonyl oxygen of the main chain and the Glc^{-1} O1 atom. The second substitution is observed in the equivalent position of Xgh74A Trp^{410} where a His residue appears in *Thermotoga maritima* endoglucanase. The sequence around this region in *T. maritima* Cel74 is also different, a Pro residue appears in the position of an otherwise strictly conserved Gly residue in a two residues shorter loop. The sequence equivalent position of Xgh74A Asn^{749} responsible for the interaction with the O2 and O3 of glucosyl residue at position -2 appears also highly conserved in the GH74 alignment.

The equivalent positions of Xgh74 residues responsible for xyloglucan recognition at the prime subsites are not as well conserved at the sequence level as the positions at the -/+ 1 and -2 subsites. The degree of sequence variation increases from position -2' where some conservation among the GH74 members is observed to position -3' where no conservation is observed among the family GH74 members. Thus, the equivalent to Xgh74 Tyr^{295} that recognises a xylose residue at the -2' subsite appears conserved in most of the sequences with the exception of the endoglucanase from *T. maritima* where an Asn residue is found instead. The equivalent position of Xgh74 Asp^{524} that contacts the xylose residue at the -1' subsite appears as Ile, Val or Ala in the xyloglucanases from *Jonesia* sp, *Thermobifida fusca* and *Hypocrea jecorina*, respectively. At the -3' subsite, Xgh74 Asp^{731} is found not conserved in any of the family members, the loop in which this residue is located displays variations in length and overall amino acid composition making difficult to assess the presence or not of an equivalent interaction in the absence of structural data on these other family members.

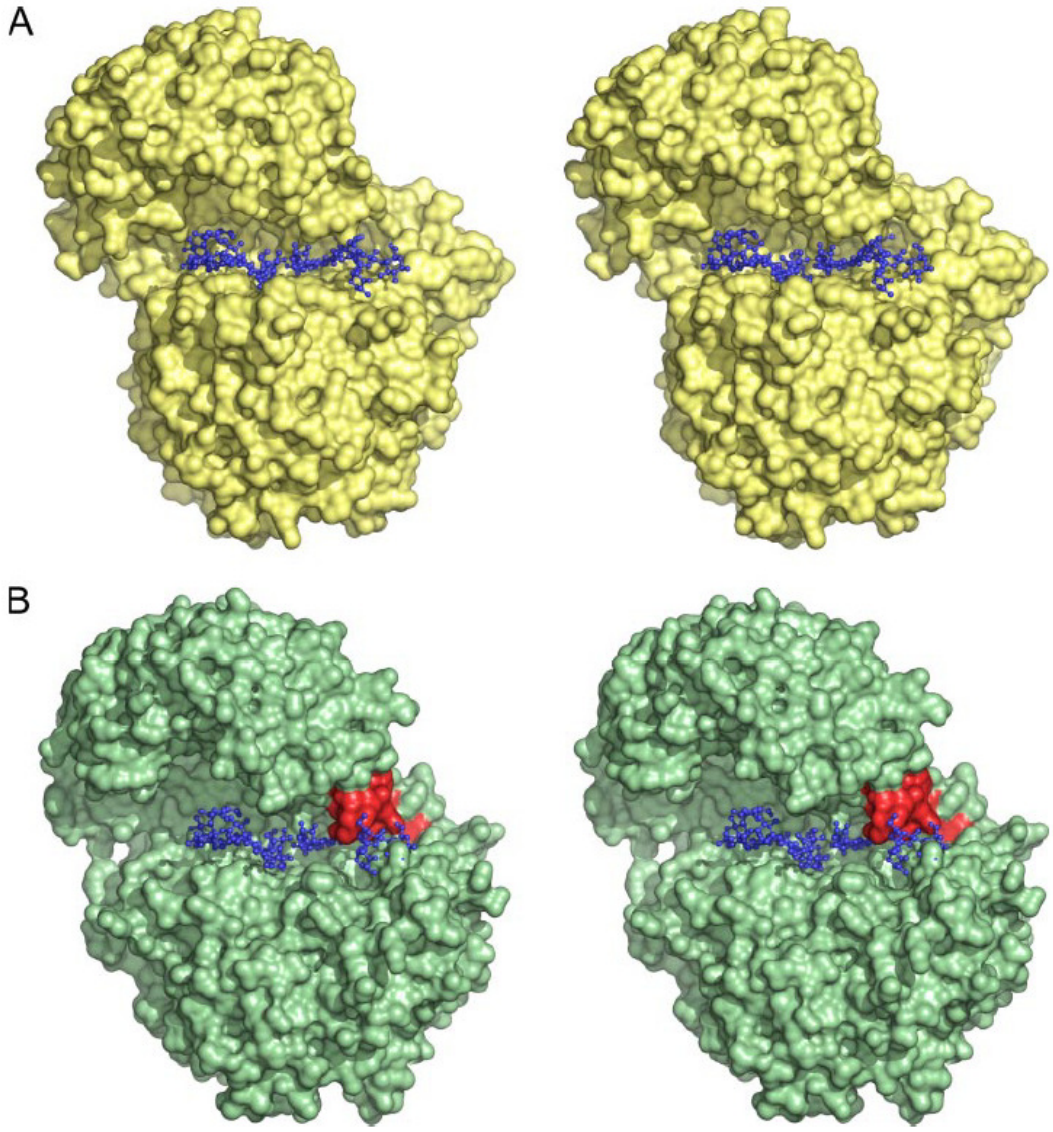

Figure 2.9. Comparison of endoxyloglucanase Xgh74A from *Clostridium thermocellum* and the reducing end-specific GH74 enzyme OXG-RCHB from *Geotrichum* sp M128.
(A) Van der Waals' surface for Xgh74A (yellow) with the ligands in blue showing the open cleft of this endoenzyme. (B) Van der Waals' surface for OXG-RCHB (pale green) with the ligands from Xgh74A superimposed in blue ball-and-stick. The surface of the OXG-RCHB loop Gly^{375}-His^{385} that closes off the substrate binding surface and is responsible for the exo mode of action is shaded in red. The figures are shown in divergent ("wall-eyed") stereo.

2.3.6. Exo versus Endo specificity in family GH74

In both Xgh74A and OXG-RCBH the substrate binding cavities are open grooves well exposed to solvent (Figure 2.9). OXG-RCBH is an exoglucanase that releases two glucosyl residues from the reducing end of the xyloglucan polymer, suggesting the presence of at least two negative reducing end subsites and two positive leaving group subsites (Yaoi and Mitsuishi, 2002; Yaoi *et al.*, 2004) and demanding the ability to

cleave at xylose-substituted glucosyl moieties. In contrast, Xgh74A processes the xyloglucan chain in an endo fashion releasing four glucosyl residue segments (Zverlov *et al.*, 2005b and this work).

Xgh74A residues Trp^{125} and Asp^{731} appear to contribute to specificity for the sugars located at subsite -3 where they interact directly with Glc^{-3} and $Xyl^{-3'}$ (Figure 2.5 and Figure 2.6). In the OXG-RCBH structure, the equivalent structural determinants for sugar recognition at the -3 site are absent. The OXG-RCBH equivalent to Trp^{125} of Xgh74A is Asp^{89}, which is situated in a loop two residues shorter than in Xgh74A and is consequently too distant to interact with the Glc^{-3} residue. Similarly, while Xgh74A Asp^{731} interacts directly with $Xyl^{-3'}$ O2 and O3 atoms, the structural equivalent of Xgh74 Asp^{731} in OXG-RCBH is Thr^{736}, but its position is again distant from $Xyl^{-3'}$ most likely as a result of conformational constraints imposed in the loop by the presence of residues Gly^{734}-Pro^{735} in OXG-RCBH. These features likely contribute to the reported differences between the two enzymes with respect to the number of reducing end subsites.

Differences between the two enzymes at the +3 sub-site (Figure 2.9) are more pronounced. The conformations of the (structurally-equivalent) loops Xgh74A Thr^{397}-Pro^{406} and OXG-RCBH Asn^{374}-Thr^{391} are dramatically different, the latter closing the binding cleft immediately after the subsite +2 and presumably restricting OXG-RCHB to exo-hydrolysis. The cleft in Xgh74A is thus open in both extremes and differs from OXG-RCBH in which the loop Gly^{375}-His^{385} blocks one-half of the substrate binding landscape. The Xgh74A structure reveals both a complex binding architecture in which subsites accommodate seventeen distinct sugar moieties accounting for the pattern of xyloglucan recognition observed by limit hydrolysis of tamarind xyloglucan. Catalysis occurs with inversion of anomeric configuration in a mechanism in which (kinetically essential) aspartates 480 and 70 likely play the role of catalytic acid and base, respectively. Given the growing importance of plant biomass conversion, especially in the context of the demand for clean energy sources, the Xgh74A structure provides the first insights into the recognition and hydrolysis of xyloglucan, a crucial component of the plant cell wall.

Acknowledgements

Authors thank Gustav Sundqvist (KTH Biotechnology) for MS analysis. This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC). CG was funded by Fundação para a Ciência e a Tecnologia (Portugal) through individual grant SFRH/BD/16731/2004. Funding from the Swedish Foundation for Strategic Research and the KTH Biofibre Materials Centre is also acknowledged. HB is a Fellow (Rådsforskare) of the Swedish Research Council. CMF solved the Xgh74A three-dimensional structure and MB executed the mass spectrometric studies.

CHAPTER 3 Functional and structural studies on CBM30 and CBM44 from *Ct*Cel9D-Cel44A

This chapter describes the functional and structural characterization of the xyloglucan-binding CBMs, termed CBM30 and CBM44, of the major cellulosomal cellulase *Ct*Cel9A-Cel44A. Data relating to the production of CBM44 crystals will be initially presented in section 3.1. In section 3.2 the structure and function of CBM30 and CBM44 will be fully described.

3.1. OVEREXPRESSION, PURIFICATION AND CRYSTALLIZATION OF THE TWO C-TERMINAL DOMAINS OF THE BI-FUNCTIONAL CELLULASE *Ct*CEL9D-CEL44A FROM *Clostridium thermocellum*

Shabir Najmudin[1], Catarina Guerreiro[2], Carlos M.G.A. Fontes[2], Luís M.A. Ferreira[2], Maria J. Romão[1], and José A.M. Prates[2]

[1]REQUIMTE, Departamento de Química, FCT-UNL, 2829-516 Caparica, Portugal; [2]CIISA, Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal

Adapted from *Acta Cryst.* (2005) F61, 1043-1045

Abstract

Clostridium thermocellum produces a highly organized multi-enzyme complex of cellulases and hemicellulases for the hydrolysis of plant cell-wall polysaccharides, which is termed the cellulosome. The bifunctional multi-modular cellulase *Ct*Cel9D-Cel44A is one of the largest components of *C. thermocellum* cellulosome. The enzyme contains two internal catalytic domains belonging to GH families 9 and 44. The C-terminus of this cellulase, comprising a PKD module and a CBM (CBM44), has been crystallized. The crystals belong to the tetragonal space group $P4_32_12$, containing a single molecule in the asymmetric unit. The native and SeMet-derivative crystals diffracted to 2.1 and 2.8 Å, respectively.

3.1.1. Introduction

Life on Earth is possible because nature has evolved mechanisms for recycling the carbon stored in organic compounds. Recalcitrant polysaccharides organized in the plant cell wall pose numerous obstacles to hydrolysis through the development of complex chemical structures and also by becoming inaccessible to biocatalysts in the intricacy of this macromolecular structure. Anaerobic bacteria and fungi have developed an extremely complex and dynamic extracellular macromolecular assembly of cellulases and hemicellulases, termed the cellulosome (see Chapter 1), which is used for the production of soluble sugars from plant cell wall polysaccharides (for reviews, see Bayer *et al*., 2004; Doi and Kosugi, 2004). The cellulosome of the thermophilic bacterium *Clostridium thermocellum* has been extensively studied (Beguin and Lemaire, 1996; Bayer *et al*., 1998; Carvalho *et al,* 2003). In this thermostable complex, the multifunctional scaffoldin (CipA) contains nine reiterated cohesin domains that individually bind to a 23-residue tandemly repeated complementary module, called the dockerin, present in the various cellulosomal catalytic subunits. Therefore, the dockerins of each cellulosomal enzyme bind specifically to one of the nine cohesin domains in CipA and thus the cellulosome comprises one molecule of CipA and nine discrete enzymes. The draft genome of *C. thermocellum* suggests that there are at least 71 cellulosomal proteins (http://www.jgi.doe.gov; April 2008). Since dockerins cannot discriminate between the individual cohesins, it is believed that the enzymatic composition of each individual cellulosome is primarily regulated at expression level (Bayer *et al*., 2004). A second type of cohesin-dockerin interaction is responsible for the integration of cellulosomes into supramacromolecular structures referred to as polycellulosomes.

The largest catalytic component of *C. thermocellum* cellulosome is the bifunctional cellulase *Ct*Cel9D-Cel44A, previously designated CelJ (Ahsan *et al.*, 1996). This enzyme comprises, sequentially from the N-terminus, a family 30 CBM (*Ct*CBM30), internal GH family 9 and 44 catalytic domains, a Type I dockerin, a PKD module and a C-terminal family 44 CBM (*Ct*CBM44). *Ct*CBM30 and *Ct*CBM44 direct the associated catalytic domains to the proximity of their target cellulosic substrates, therefore maximizing catalytic activity by promoting an intimate and prolonged interaction between enzyme and substrate, as referred in Chapter 1 (Boraston *et al*., 2004). Interestingly, the CBMs from *Ct*Cel9D-Cel44A recognize with equal efficiency undecorated and highly branched β-1,4 glucosidic ligands, such as cellulose and

xyloglucans, respectively (Guerreiro *et al.*, 2006), as it will be fully described in section 3.2. However, the structural determinants that may allow the binding of CBMs at a single binding site to such different polysaccharides remain unknown. To gain insights into the structural properties that govern the promiscuity in ligand recognition revealed by these CBMs, this work aims to determine the crystal structures of *C. thermocellum* CBM30 and CBM44. In the present communication, the overexpression, purification, crystallization and preliminary X-ray analysis of the *C. thermocellum Ct*Cel9D-Cel44A C-terminal region comprising both PKD and CBM44 modules is described.

3.1.2. PKD-CBM44 expression and purification

To express PKD-CBM44 in *E. coli*, the region of the *Ct*Cel9D-Cel44A gene encoding PKD and CBM44 was amplified by PCR from *C. thermocellum* YS genomic DNA, using the thermostable DNA polymerase *Pfu* Turbo (Stratagene). The primers used are depicted in Table 3.2, construct pCG2 (see 3.2.2.1). The PCR product was cloned into pGEM T-easy (Promega) and sequenced to ensure that no mutations had occurred during the PCR. The recombinant pGEM T-easy derivative was digested with *Nhe*I and *Xho*I and the excised clostridial gene was cloned into the similarly restricted expression vector pET21a (Novagen), to generate pCG2. PKD-CBM44 encoded by pCG2 contains a C-terminal His_6-tag.

To express the recombinant protein, *E. coli* BL21 (DE3) (Novagen) or the methionine auxotroph B834 (DE3) (Novagen) transformed with pCG2 were cultured in 12 L LB containing 100 μg/mL ampicillin or in 4 L liquid growth medium prepared as described by Carvalho *et al*. (2004), respectively, at 37 ºC to mid-exponential phase ($OD_{550\ nm}$=0.6). At this point, IPTG was added to a final concentration of 1 mM and the cultures were incubated for a further 5 h. Levels of recombinant protein expression were in both cases approximately 10 mg of PKD-CBM44 per litre of medium. The His_6-tagged recombinant protein was purified from cell-free extracts by IMAC as described previously (Dias *et al*., 2004). Purified PKD-CBM44 protein was buffer-exchanged, using PD-10 Sephadex G-25M gel-filtration columns (GE Health), into Buffer A (see 3.2.2.1), concentrated to 20 mg/mL with Amicon 10 kDa molecular-weight centrifugation membranes and subjected to gel filtration using a HiLoad 16/60 Superdex 75 column (GE Health) with protein eluted at 1 mL/min in Buffer A. The recombinant protein was concentrated and washed three times with 5 mM DTT in water (for the

SeMet protein) or water (for the native protein), using the same centrifugal membrane, and the final protein concentration was adjusted to 60 mg/mL. The purity of the protein, which was found to be nearly homogenous, was estimated using SDS-PAGE (Figure 3.1).

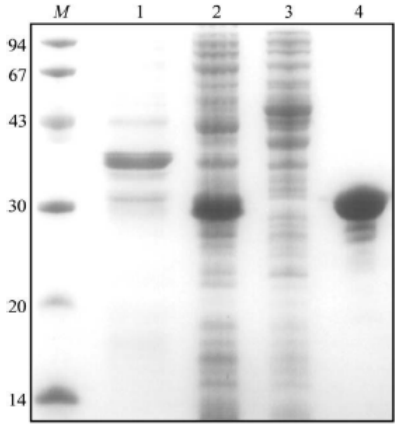

Figure 3.1. 14% SDS-PAGE evaluation of PKD-CBM44 purity during purification.
Lane *M*, molecular-weight markers (kDa); lane 1, cell debris after sonication; lane 2, cell-free extract after sonication; lane 3, eluate from the column after sample loading; lane 4, purified PKD-CBM44 protein after Ni-affinity column chromatography.

3.1.3. PKD-CBM44 crystallization conditions

The crystallization conditions were screened by the hanging-drop vapour-phase diffusion method using an in-house modified version of the sparse-matrix method of Jancarik and Kim (1991) and the Hampton Crystal screen, Crystal screen 2 and PEG/Ion screen (Hampton Research, Alison Viejo, CA). Drops consisting of 1 µL 13, 30, 50 and 60 mg/mL PKD-CBM44 and 1 µL reservoir solution were prepared at 4 ºC and 20 ºC. Clusters of tiny bipyramidal protein crystals (maximum dimension ~50 µm) grew at 4 ºC within a week in the following three conditions: (i) 0.2 M $MgCl_2$, 0.1 M HEPES pH 7.0, 30% ethanol; (ii) 0.2 M $CaCl_2$, 0.1 M sodium acetate pH 4.5 and 30% ethanol and (iii) 0.2 M $CaCl_2$, 0.1 M sodium acetate pH 4.5 and 20% 2-isopropanol. These crystals were very unstable and difficult to handle owing to the volatility of the precipitants. Crystals did not diffract using the in-house source and only diffracted to 3.5 Å at the ESRF. These crystals were improved on by fine-screening and by using the Hampton Additive Screens I and II. Glycerol [6% (v/v)], $CdCl_2$ (10-30 mM), xylitol [3-25% (v/v)] and $SrCl_2$ (10 mM) produced fewer but larger and better diffracting crystals

(0.1×0.1×0.2 mm) over a period of 1-8 weeks at 4 ºC (Figure 3.2). These crystals were stabilized by adding 1 µL cryoprotectant solutions containing 30% (v/v) ethylene glycol or glycerol to the crystallization buffer stepwise over a few days, allowing them to equilibrate between each addition. The crystals were flash-frozen in liquid nitrogen after soaking in the cryoprotectant, either 30% (v/v) ethylene glycol or glycerol added to the crystallization buffer, for a few seconds.

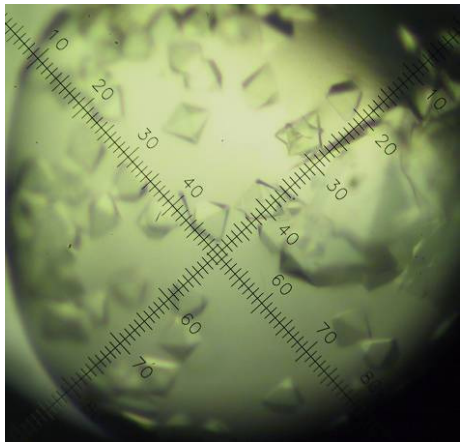

Figure 3.2. Crystals of PKD-CBM44.
The crystals were obtained by hanging-drop vapour diffusion in the presence of 0.2 M $CaCl_2$, 0.1 M sodium acetate pH 4.5 and 30% ethanol with 6% glycerol as additive. The largest crystals are approximately 0.1×0.1×0.2 mm in size.

Crystals of SeMet-containing protein were also obtained by vapour-phase diffusion using the hanging-drop method with equal volumes (1 µL) of protein solution (30 or 60 mg/mL in 5 mM DTT) and reservoir solution from a fine-screen based around the above three successful conditions for the native crystals. Crystals grew at 4 ºC over a period of one week in two different conditions: (i) 0.2 M $CaCl_2$, 0.1 M sodium acetate pH 4.5 and 22.5% ethanol and (ii) above plus 10 mM $SrCl_2$. The crystals were treated in the same way as the native crystals for stabilization and flash-freezing.

3.1.4. Data collection and processing

The final statistics for data collection and processing are summarized in Table 3.1. Structure determination will be described in the following section (Guerreiro *et al.*, 2006). Initial data sets were collected on a home source with Cu Kα X-ray radiation

(Enraf-Nonius rotating-anode generator operated at 5 kW, 100 mA equipped with a MAR imaging plate). A native data set (at λ=1.2915 Å) and SeMet-labelled (at the selenium-edge for MAD) data were collected on beamline ID14-EH4 at the ESRF (Grenoble, France) using a Quantum 4 charge-coupled device detector (ADSC) with the crystal cooled at -173 ºC using a Cryostream (Oxford Cryosystems Ltd.). All data sets were processed using the programs MOSFLM (Leslie, 1992) and SCALA (Kabsch, 1988a, 1988b) from the CCP4 suite (1994). All crystals belong to the tetragonal space group. The Matthews coefficient (V_m=3.6 $Å^3$ Da^{-1}) indicated the presence of one molecule in the asymmetric unit and a solvent content of 65% (Matthews, 1968). At the home source, the improved crystals diffracted to 3.05 Å. A complete and highly redundant data set was collected for a native crystal grown in crystallization buffer (ii) with 6% (v/v) glycerol. A total of 360º of data were collected with 1º oscillations over approximately 36 h. Only the first 256 images were used in processing as the crystal suffered radiation damage. Attempts to solve the native structure by molecular replacement using an ensemble of PKD domains (PDB codes 1B4R, 1CTN, 1LOQ; sequence identities <30%) and of various CBM (sequence identities <20%) structures were unsuccessful. This data was also not good enough to give a sufficient signal for the S atoms for a SAD experiment.

Table 3.1. Crystallization and data-collection statistics for PKD-CBM44.
Values in parentheses are for the highest resolution shell.

Data set	**Native 1**	**SeMet (peak)**	**Native 2**
Main precipitant[a]	30% ethanol + 6% glycerol	22.5% ethanol	20% 2-propanol + 10 mM $CdCl_2$
Cryoprotection	30% ethylene glycol	30% glycerol	30% ethylene glycol
X-ray source	Cu *Kα* rotating anode, in-house	ID14-EH4, ESRF	ID14-EH4, ESRF
Wavelength (Å)	1.5418	0.97934	1.2915
Spacegroup	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit-cell parameters (Å)			
a=*b*	87.21	87.10	87.47
c	106.14	107.67	103.19
Resolution limits (Å)	45-3.05	67.7-2.8	61.9-2.1
No. of observations	180868 (18185)	127725 (17834)	145323 (10261)
No. of unique observations	8203 (1101)	19840 (2884)	23914 (3311)
Multiplicity	22.0 (16.5)	6.4 (6.2)	6.1 (3.1)
Completeness (%)	99.1 (94.5)	99.8 (99.7)	99.5 (96.7)
($I/\sigma(I)$)	20.8 (3.2)	20.3 (2.6)	15.1 (1.3)
R_{merge}[b]	20.7 (91.3)	9.1 (56.0)	6.7 (72.5)

[a]Plus 0.1 *M* sodium acetate pH 4.5 and 0.2 M $CaCl_2$ to make up the crystallization buffer.
[b]$R_{merge}=\Sigma_h \Sigma_i |I(h,i) - (I(h))|/\Sigma_h \Sigma_i$ I(h,i), where $I(h,i)$ is the intensity of the measurement of reflection h and $(I(h))$ is the mean value of $I(h,i)$ for all i measurements.

On beamline ID14-EH4 at the ESRF, the SeMet crystal diffracted to a resolution of 2.8 Å and the native crystal to 2.1 Å. Location of anomalous scatterers, phasing and density modification were carried out using SHELXD and SHELXE (Sheldrick and Schneider, 1997). The SeMet crystal suffered severe radiation damage during the edge data collection. Thus, only the SeMet peak data was used to determine the positions of 4/5 Se atoms. The correlation coefficient output by SHELXD was 40%. The pseudo-free correlation coefficient, contrast and connectivity figures of merit given by SHELXE for the correct heavy-atom enantiomer ($P4_32_12$) were 48.55%, 0.807 and 0.908, respectively, as opposed to 24.85%, 0.148 and 0.789, respectively for the wrong hand ($P4_12_12$).

Acknowledgements

This work was supported in part by Fundação para a Ciência e a Tecnologia (Portugal) through POCI/BIA-PRO/59118/2004 and individual grants SFRH/BD/16731/2004 (CG) and SFRH/BPD/20357/2004 (SN). The authors would like to thank Dr. Elspeth Gordon for her help during synchrotron data collection at beamline ID14-EH4 at the ESRF (Grenoble, France).

3.2. XYLOGLUCAN IS RECOGNIZED BY CARBOHYDRATE-BINDING MODULES THAT BIND TO β-GLUCAN CHAINS

Catarina I.P.D. Guerreiro[1*], Shabir Najmudin[2*], Ana L. Carvalho[2], José A.M. Prates[1], Márcia A.S. Correia[1], Victor D. Alves[1], Luís M.A. Ferreira[1], Maria J. Romão[2], Harry J. Gilbert[3], David N. Bolam[3], and Carlos M.G.A. Fontes[1]

[1]CIISA, Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; [2]REQUIMTE, Departamento de Química, FCT-UNL, 2829-516 Caparica, Portugal; [3]Institute for Cell and Molecular Biosciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, UK

Adapted from *J. Biol. Chem.* (2006) 281, 8815-8828

Abstract

Enzyme systems that attack the plant cell wall contain non-catalytic CBMs that mediate attachment to this composite structure and play a pivotal role in maximizing the hydrolytic process. Although xyloglucan, which comprises a backbone of β-1,4-glucan decorated primarily with xylose residues, is a key component of the plant cell wall, CBMs that bind to this polymer have not been identified. Here it is shown that the C-terminal domain of the modular *Clostridium thermocellum* enzyme *Ct*Cel9D-Cel44A (formerly known as CelJ) comprises a novel CBM (designated CBM44) that binds with equal affinity to cellulose and xyloglucan. It is also shown that accommodation of xyloglucan side-chains is a general feature of CBMs that bind to single cellulose chains. The crystal structures of CBM44 and the other CBM (CBM30) in *Ct*Cel9D-Cel44A display a β-sandwich fold. The concave face of both CBMs contains a hydrophobic platform comprising three tryptophan residues that can accommodate up to five glucose residues. The orientation of these aromatic residues is such that the bound ligand would adopt the twisted conformation displayed by cello-oligosaccharides in solution. Mutagenesis studies confirm that the hydrophobic platform located on the concave face of both CBMs mediates ligand recognition. In contrast to other CBMs that bind to single polysaccharide chains, the polar residues in the binding cleft of CBM44 play only a minor role in ligand recognition. The mechanism by which these proteins are able to recognize linear and decorated β-1,4-glucans is discussed based on the structures of CBM44 and the other CBMs that bind single cellulose chains.

[*]Contributed equally to this work

3.2.1. Introduction

The recycling of photosynthetically fixed carbon by the action of microbial plant cell wall hydrolases is a fundamental biological process that is integral to one of the major geochemical cycles and, in addition, has considerable industrial potential (Coughlan, 1985). The complex chemical and physical structure of the plant cell wall restricts enzyme access to the polysaccharides, primarily cellulose and hemicellulose, which comprise this composite macromolecule. Microbial cellulases and hemicellulases generally contain non-catalytic CBMs that target these enzymes to specific sites within the plant cell wall (Boraston *et al.*, 2004). As referred in Chapter 1, CBMs potentiate the hydrolytic activity of these biocatalysts by bringing the appended catalytic modules into intimate contact with their target substrates, thereby reducing the "accessibility problem" (Bolam *et al.*, 1998; Gill *et al.*, 1999; Boraston *et al.*, 2004). Based on primary structure similarity, CBMs have been grouped into 51 sequence-based families (Coutinho and Henrissat, 1999; http://afmb.cnrs-mrs.fr/CAZY; April 2008).

Structural studies on representatives of the majority of CBM families demonstrate that these protein modules display a β-sandwich fold, which has led to the identification of a CBM superfamily (Boraston *et al.*, 2004). This fold consists of two β-sheets each consisting of 3-6 antiparallel β-strands. The topology of CBM ligand-binding sites, which may vary in proteins that are members of the same family, complements the conformation of the target polysaccharide. Thus, in Type A modules, which interact with the flat surfaces of crystalline polysaccharides, the binding site consists of a planar hydrophobic platform that contains three exposed aromatic amino acids (Kraulis *et al.*, 1989; Tormo *et al.*, 1996; Raghothama *et al.*, 2000). These CBMs show no significant affinity for soluble polysaccharides, and the ligand specificity of CBM families that contain Type A modules is usually invariant. In contrast, Type B CBMs, which bind to single polysaccharide chains, accommodate the ligands within extended clefts of varying depths (Xu *et al.*, 1995; Charnock *et al.*, 2000; Szabo *et al.*, 2001; Boraston *et al.*, 2002). In these CBMs, polysaccharide recognition is variable within different members of these families and normally reflects the catalytic specificity of their cognate catalytic modules. Indeed, Type B CBM families display plasticity in their capacity to accommodate heterogeneity both in the branches that may decorate the sugar backbone and the composition and linkage of the sugar backbone. This variation in ligand recognition is exemplified in CBM family 6 (CBM6), which contains proteins that

recognize xylan, cellulose (β-1,4-linked glucose homopolymer), laminarin (β-1,3-linked glucose homopolymer), and β-1,4- and β-1,3-mixed linked β-glucans such as lichenan (Czjzek *et al.*, 2001; Pires *et al.*, 2004; Bueren *et al.*, 2005). In addition, plasticity in ligand recognition may result from subtle molecular determinants in a single CBM-binding site. Thus, polar residues in CBM29 are able to hydrogen bond to the axial O2 in mannose and the equatorial O2 in glucose enabling the protein to bind cellulose, glucomannan and mannan (Charnock *et al.*, 2002). Finally, studies on CBMs that recognize the branched hemicelluloses xylan and galactomannan indicate that the side-chains of decorated polysaccharides are usually solvent-exposed and do not restrict ligand binding (Notenboom *et al.*, 2001a; Charnock *et al.*, 2002; Pires *et al.*, 2004; Flint *et al.*, 2005) nor represent specificity determinants. It is believed that this feature of Type B CBMs enables these proteins to interact with similar affinities to decorated and undecorated polysaccharides. The generality of this hypothesis, however, remains to be established particularly with respect to the key polysaccharides that contain a backbone of β-1,4-linked glucose polymers that can be undecorated, exemplified by cellulose, or contain extensive side-chains such as xyloglucan. Both these polysaccharides are abundant in plant cell walls, particularly in the primary cell wall of dicotyledons, where they provide tensile strength and thus undergo modification in rapidly expanding cells.

To investigate whether CBMs can display high affinity for heavily branched β-1,4-glucans such as xyloglucan, and to interrogate the molecular determinants that govern this binding specificity, we have studied the structure and function of the N- and C-terminal domains of *Ct*Cel9D-Cel44A from *Clostridium thermocellum*, a component of the bacterium's cellulosome (Ahsan *et al.*, 1996). This enzyme contains internal GH families 9 and 44 catalytic domains in addition to an N-terminal CBM30, a Type I dockerin, and a C-terminal module of unknown function. Here it is shown that the C-terminal domain of *Ct*Cel9D-Cel44A comprises a novel CBM (designated CBM44) that displays high affinity for β-1,4 glucose polymers, including xyloglucan. The three-dimensional structures of CBM30 and CBM44, combined with mutagenesis studies, reveal that both modules contain a single ligand cleft with a strip of three aromatic residues that are displayed to the edge of the respective ligand-binding site. The mechanism by which CBMs that recognize cellulose are also able to accommodate the side-chains of xyloglucan is discussed.

3.2.2. Experimental Procedures

3.2.2.1. Protein expression and purification

*Ct*Cel9D-Cel44A is a modular enzyme containing an N-terminal CBM30, followed by two catalytic domains belonging to GH9 and GH44, a dockerin and a C-terminal domain of unknown function (Figure 3.3). The internal region of the C-terminal domain of *Ct*Cel9D-Cel44A displays homology with polycystic kidney disease (PKD) domains found in a variety of eukaryotic and prokaryotic proteins.

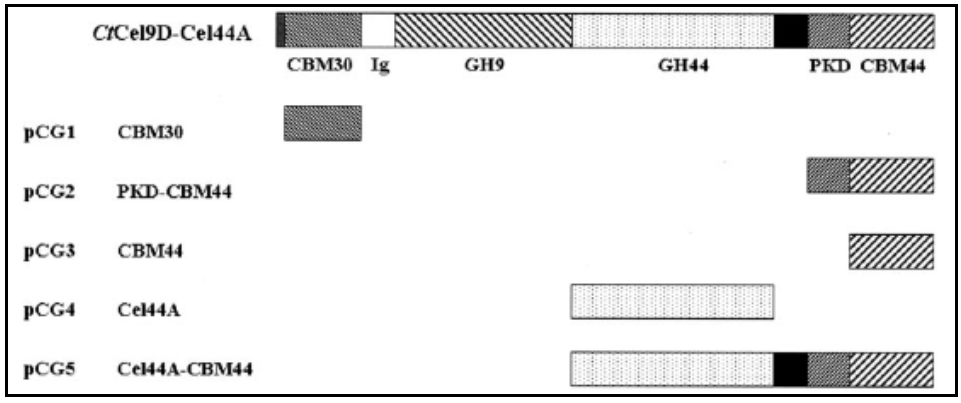

Figure 3.3. Molecular architecture of *Ct*Cel9D-Cel44A and its truncated derivatives.
The modules encoded by the defined recombinant plasmids are indicated. The grey and black boxes represent the signal peptide and the dockerin, respectively. Domain boundaries are specified below the modular representation of full-length *Ct*Cel9D-Cel44A.

DNA encoding CBM30, PKD-CBM44, CBM44, Cel44A and Cel44A-CBM44 was amplified by PCR from *C. thermocellum* YS genomic DNA, using the primers listed in Table 3.2, and cloned into *Nhe*I and *Xho*I restricted pET21a to generate pCG1, pCG2, pCG3, pCG4 and pCG5, respectively (Figure 3.3). The recombinant plasmids were sequenced to ensure that no mutations had occurred during the PCR that used the thermostable DNA polymerase *Pfu* Turbo (Stratagene).

The recombinant proteins, which contain a C-terminal His_6-tag, were expressed in *Escherichia coli* BL21 (DE3) (Novagen) cells, harbouring the appropriate recombinant plasmid, cultured in LB containing 100 μg/mL ampicillin at 37 ºC to mid-exponential phase ($OD_{550\ nm}$=0.6). IPTG was then added to a final concentration of 1 mM and the cultures were incubated for a further 3-5 h. The His_6-tagged recombinant proteins were purified from cell-free extracts by IMAC as described previously (Dias *et al.*, 2004). Preparation of *E. coli* to generate SeMet-PKD-CBM44 was performed as described by

Carvalho *et al*. (2004), and the protein was purified by using the same procedures employed for the native CBMs, except that the buffer also contained 2 mM *β*-mercaptoethanol.

Table 3.2. Primers used for the cloning and mutagenesis of CBM30 and CBM44.

Construct	Primers	Sequence (5'-3')
pCG1	Forward	CTC**GCTAGC**AGTGCCGAAACAGTTGCT[a]
	Reverse	CAC**CTCGAG**CTTGATTGCAGGAGCGGA
pCG2	Forward	CTC**GCTAGC**GTTCCGGAAAACCAGGCA
	Reverse	CAC**CTCGAG**CCAGTCAATAGCATCTAC
pCG3	Forward	CTC**GCTAGC**TTTACAGCTACCATAAAAG
	Reverse	CAC**CTCGAG**CCAGTCAATAGCATCTAC
pCG4	Forward	CTC**GCTAGC**GCAAAAGTGGTTGACATC
	Reverse	CAC**CTCGAG**CTCCGCAGCTTCAAGCAC
pCG5	Forward	CTC**GCTAGC**GCAAAAGTGGTTGACATC
	Reverse	CAC**CTCGAG**CCAGTCAATAGCATCTAC
$W22A_{30}$	Forward	TCGCCTGTAGTCGGAGCGTCAGGAAGCGGTATG
	Reverse	CATACCGCTTCCTGACGCTCCGACTACAGGCGA
$W63A_{30}$	Forward	GTTCAGTCAGGAGCGTGGATATCTCTTCTTAC
	Reverse	GTAAGAAGAGATATCCACGCTCCTGACTGAAC
$W73A_{30}$	Forward	ACATTAAGAGGAGCGAACACCCATGAC
	Reverse	GTCATGGGTGTTCGCTCCTCTTAATCT
$W171A_{30}$	Forward	CCGTTTACAGTAGCGTTCAGTGATATAAAG
	Reverse	CTTTATATCACTGAACGCTACTGTAAACGG
$Q179A_{44}$	Forward	CCTATTGCTGCCATCGCGCCGTATATTATGC
	Reverse	GCATAATATACGGCGCGATGGCAGCAATAGG
$M183A_{44}$	Forward	CCAGCCGTATATTGCGCCTCATACACC
	Reverse	GGTGTATGAGGCGCAATATACGGCTGG
$D188A_{44}$	Forward	CCTCATACACCTGCGTGGTCGGAAGTC
	Reverse	GACTTCCGACCACGCAGGTGTATGAGG
$W189A_{44}$	Forward	CTCATACACCTGATGCGTCGGAAGTCCTC
	Reverse	GAGGACTTCCGACGCATCAGGTGTATGAG
$V192A_{44}$	Forward	GATTGGTCGGAAGCGCTCTGGAATTCG
	Reverse	CGAATTCCAGAGCGCTTCCGACCAATC
$W194A_{44}$	Forward	CGGAAGTCCTCGCGAATTCGACATGG
	Reverse	CCATGTCGAATTCGCGAGGACTTCCG
$S196A_{44}$	Forward	GTCCTCTGGAATGCGACATGGAAAGG
	Reverse	CCTTTCCATGTCGCATTCCAGAGGAC
$W198A_{44}$	Forward	GGAATTCGACAGCGAAAGGATACACCATGG
	Reverse	CCATGGTGTATCCTTTCGCTGTCGAATTCC
$T202A_{44}$	Forward	GACATGGAAAGGATACGCGATGGTGAAGACCG
	Reverse	CGGTCTTCACCATCGCGTATCCTTTCCATGTC
$M203A_{44}$	Forward	GGAAAGGATACACCGCGGTGAAGACCG
	Reverse	CGGTCTTCACCGCGGTGTATCCTTTCC
$Q227A_{44}$	Forward	CGACTTGGCCGCAGGCGATGGGTATACAGG
	Reverse	CCTGTATACCCATCTCCTGCGGCCAAGTCG
$Q231A_{44}$	Forward	GCAGATGGGTATAGCGGTACAGACCATAG
	Reverse	CTATGGTCTGTACCGCTATACCCATCTGC

[a]The *Nhe*I and *Xho*I restriction sites introduced in the forward and reverse primers, respectively, are shown in bold.

For crystallization, CBM30 and PKD-CBM44 were further purified by size exclusion chromatography. Following IMAC, the proteins were buffer-exchanged into 50 mM HEPES-HCl buffer, pH 7.5, containing 200 mM NaCl (Buffer A or Buffer A + 5 mM DTT, for SeMet protein) and then subjected to gel filtration using a HiLoad 16/60 Superdex 75 column (GE Health) with protein eluted at 1 mL/min in Buffer A. Purified CBM30 and PKD-CBM44 were concentrated using an Amicon 10-kDa molecular mass centrifugal concentrator and washed three times with 5 mM DTT (for the SeMet proteins) or water (for the native proteins). *Cellulomonas fimi* CBM4-1 (*Cf*CBM4-1), *C. thermocellum* CBM11 (*Ct*CBM11), *Clostridium cellulovorans* CBM17 (*Cc*CBM17), *Bacillus* spp CBM28 (*Bsp*CBM28) and *Piromyces equi* CBM29-2 (*Pe*CBM29-2) were all prepared as described previously (Kormos *et al.*, 2000; Carvalho *et al.*, 2004; Boraston *et al.*, 2000; Boraston *et al.*, 2002 and Charnock *et al.*, 2002, respectively).

3.2.2.2. Source of sugars used

All soluble polysaccharides were purchased from Megazyme® (Ireland), except oat spelt xylan, laminarin and HEC, which were obtained from Sigma, and pustullan which was obtained from Calbiochem. Cello-oligosaccharides were from Seikagaku Corp. (Japan). Avicel (PH101) was obtained from Serva, and acid-swollen cellulose was prepared as described previously (Wood, 1988).

3.2.2.3. Binding assays

The affinity of CBM30, PKD-CBM44 and CBM44 for a range of soluble polysaccharides was determined by affinity gel electrophoresis (AGE). The method was essentially as described by Tomme *et al.* (2000) using the polysaccharide ligands at a concentration of 0.1% (w/v) unless stated otherwise. Electrophoresis was carried out for 4 h at room temperature in native polyacrylamide gels containing 10% (w/v) acrylamide. The non-binding negative control protein was bovine serum albumin (BSA). Quantitative assessment of binding was carried out as described previously (Takeo, 1984), using polysaccharide concentrations ranging from 0.001 to 0.5% (w/v). Qualitative assessment of PKD-CBM44 and CBM44 binding to Avicel was carried out as follows: 30 μg of protein in 50 mM Tris-HCl buffer, pH 7.5, containing 0.05% (v/v) Tween 20 and 5 mM $CaCl_2$ (Buffer B) were mixed with 1 mg of ligand in a final reaction volume of 200 μL. The reaction mixture was incubated for 2 h at room

temperature with gentle shaking, after which time the insoluble ligand was precipitated by centrifugation at 13,000*g* for 5 min. The supernatant, comprising the unbound fraction, was removed, and the pellet was washed three times with 200 µL of Buffer B. The bound protein was eluted by boiling the polysaccharides in 200 µL of 10% (w/v) SDS containing 10% (v/v) β-mercaptoethanol for 10 min. Bound and unbound fractions were analysed by SDS-PAGE using a 14% (w/v) acrylamide gel.

3.2.2.4. Isothermal titration calorimetry (ITC)

ITC measurements were made at 25 ºC following standard procedures (Flint *et al.*, 2004) using a Microcal Omega titration calorimetre. Proteins were dialyzed, extensively, against either 50 mM HEPES-HCl buffer, pH 8.0, or 50 mM sodium phosphate buffer, pH 7.0, and the ligand was dissolved in the same buffer to minimize heats of dilution. During a titration experiment the protein sample (40-250 µM), stirred at 300 rpm in a 1.4331-mL reaction cell, was injected with a single 1 µL aliquot, followed by 29 successive 10 µL aliquots of ligand comprising 2-5 mg/mL polysaccharide or 0.5-5 mM oligosaccharide at 200 s intervals. Integrated heat effects, after correction for heats of dilution, were analyzed by nonlinear regression using a single site binding model (Microcal Origin, version 7.0). The molar concentration of CBM-binding sites present in the polysaccharide ligands was determined as described previously (Bolam *et al.*, 2004). The fitted data yield the association constant (K_a) and the enthalpy of binding (ΔH). Other thermodynamic parameters were calculated using the standard thermodynamic Equation 1,

$$-RT\ln K_a = \Delta G = \Delta H - T\Delta S \qquad \text{(Equation 1)}$$

3.2.2.5. Site-directed mutagenesis

Mutants of CBM30 and CBM44 were generated using the PCR-based QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions, using pCG1 and pCG3 as the template DNA, respectively. The sequences of the primers used to generate these mutants are displayed in Table 3.2. The mutated DNA sequences were sequenced to ensure that only the appropriate mutations had been incorporated into the nucleic acid.

3.2.2.6. Enzyme assays

The activity of truncated derivatives of *Ct*Cel9D-Cel44A against various polysaccharides was determined as described previously (Dias *et al.*, 2004) by measuring the rate of release of reducing sugars using the Nelson (Nelson, 1944) and Somogyi (Somogyi, 1945) reagent. Assays were carried out at 50 ºC, in 50 mM sodium phosphate buffer, pH 5.0, containing 1 mg/mL BSA. To determine the pH profile of Cel44A, 50 mM sodium phosphate, pH 4.5-7.0 and 50 mM Tris-HCl, pH 7.0-9.5, buffers were used in the enzyme assays. The linearity of the reactions was confirmed by measuring the release of reducing sugars at three time points. All reported results are the mean of three separate experiments.

3.2.2.7. Crystallization and data collection

Crystallization conditions for both PKD-CBM44 and CBM30 were screened using the Hampton Crystal screen, Crystal screen 2 and PEG/Ion screen (Hampton Research, Alison Viejo, CA). Native or SeMet crystals of PKD-CBM44 were grown by vapor-phase diffusion using the hanging drop method with an equal volume (1 µL) of protein (50 mg/mL in water or 5 mM DTT, respectively) and reservoir solution (0.2 M $CaCl_2$, 0.1 M sodium acetate, pH 4.5, and 22.5% ethanol (v/v)). Crystals, which grew over a period of 1 week, were stabilized by adding 1 µL of cryoprotectant solutions containing 30% (v/v) ethylene glycol or glycerol in the crystallization buffer, stepwise over a few days for equilibration, and flash-frozen in liquid nitrogen. The detailed protocols used to obtain PKD-CBM44 crystals are described later in this chapter (Najmudin *et al.*, 2005). Initial PKD-CBM44 data sets were collected on a home source with CuK_α X-ray radiation from an Enraf-Nonius rotating anode generator operated at 5 kW, with a MAR research image-plate detector. Native (at λ=1.2915 Å) and SeMet (at the selenium-edge for MAD) data sets were collected on the beamline ID14-EH4 at the ESFR using a Quantum 4 charge-coupled device detector (ADSC) with the crystals cooled at -173 ºC using a cryostream (Oxford Cryosystems Ltd.). Crystals belong to the tetragonal space group. The Matthews coefficient (V_M=3.6 $Å^3$/Da) indicated the presence of one molecule in the asymmetric unit and a solvent content of 65% (Matthews, 1968). Crystals of CBM30 native protein were grown with an equal volume (1 µL) of protein (66 mg/mL in water) and reservoir solution (0.8 M sodium and potassium tartrate in 0.1 M NaHEPES, pH 7.5). A CBM30 crystal, which grew over a

period of 3-4 days, was soaked for a few seconds in a modified crystallization solution containing 30% (v/v) glycerol and flash-cooled in a nitrogen stream at -173 ºC for data collection. Preliminary crystal characterization was performed in-house, and the diffraction experiments showed that CBM30 crystals belonged to the primitive orthorhombic space group with unit cell parameters *a*=66.6 Å, *b*=85.5 Å, *c*=88.9 Å. A complete data set of a single crystal was collected at beamline ID23-1 at the ESRF, using a marCCD detector. At a fixed wavelength of 1.072 Å, the crystal diffracted beyond 2.27 Å resolution. Systematic absence of the odd reflections in the *h*, *k* and *l* axis indicates that the crystal belonged to the $P2_12_12_1$ space group. A Matthews coefficient of 3 $Å^3$/Da suggested the presence of two molecules in the asymmetric unit, with a solvent content of approximately 60% (Matthews, 1968). All data sets were processed using the programs MOSFLM (Leslie, 1992) and SCALA (Kabsch, 1988a, 1988b) from the CCP4 suite (1994) and the statistics are shown in Table 3.3.

3.2.2.8. Phasing, model building and refinement

The location of four of the five selenium sites in PKD-CBM44, phasing, and density modification were performed using the SHELXD/E package (Sheldrick and Schneider, 1997). SHELXD gave a correlation coefficient of 40% and SHELXE confirmed the space group to be unambiguously $P4_32_12$. The selenium sites were refined with SOLVE (Terwilliger and Berendzen, 1999), and the resulting phases fed into RESOLVE to give a partial (~50%) structural model. Automated model building using ARP/wARP (Perrakis *et al.*, 1999) and the RESOLVE partial structure and the high energy remote SeMet data generated a model with R_{work}=29.2%. This initial model consisted of 243 of a total of 260 amino acid residues. This model was used to fit into the 2.17 Å native data after rigid-body refinement in REFMAC5 (Murshudov *et al.*, 1997), treating each domain as a rigid body, with R_{work}=42.9%. Iterative model building with O (Jones *et al.*, 1991), together with restrained refinement in REFMAC5 (Murshudov *et al.*, 1997) resulted in a final model with R_{work}=16.9% (R_{free}=21.6%).

The four residues from the PKD-CBM44 N-terminus (Met, Ala, Ser and Val) were disordered and were not included in the final model. However, Leu^{250} and Glu^{251} in the C-terminal linker were defined, although the six residues of the His_6-tag were not. The final model includes 250 amino acid residues, two calcium ions, 12 ethylene glycol molecules, and 380 water molecules. The final refined native model was used to complete the refinement of the 2.1 Å SeMet data, after rigid-body fitting, using O

(Jones *et al.*, 1991) and REFMAC5 (Murshudov *et al.*, 1997). The final SeMet model included 250 amino acid residues, two calcium ions, three ethylene glycol molecules, and 381 water molecules. It is similar to the native form, with significant deviations (RMSD>0.4 Å) in two loop regions (residues 47-55 and 187-194).

Table 3.3. Data collection and structure statistics for natives CBM30 and PKD-CBM44 and SeMet PKD-CBM44.
Values in parentheses are for the highest resolution shell.

Data set	Native CBM30	SeMet (peak)[a]	SeMet (edge)[a]	SeMet (higher energy edge)[a]	SeMet (high energy remote)[a]	Native PKD-CBM44
Wavelength (Å)	1.072	0.97934	0.97936	0.97936	0.94645	1.2915
Spacegroup	$P2_12_12_1$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit cell parameters						
a (Å)	66.6	87.10	86.87	86.92	86.90	87.47
b (Å)	85.5	87.10	86.87	86.92	86.90	87.47
c (Å)	88.9	107.67	108.31	108.39	108.24	103.19
Resolution limits (Å)	50-2.27	67.7-2.8	67.7-2.15	67.7-2.10	67.7-2.1	66.7-2.17
No. observations	152,855 (22,036)	127,725 (17,834)	265,692 (16,590)	270,229 (14,500)	188,468 (26,191)	139,247 (12,180)
No. unique observations	24,116 (3,452)	19,840 (2,884)	22,678 (2,816)	23,856 (2,729)	24,857 (3,557)	21,826 (3,121)
Multiplicity	6.3 (6.4)	6.4 (6.2)	11.7 (5.9)	11.3 (5.3)	7.6 (7.4)	6.4 (3.9)
Completeness (%)	99.9 (100)	99.8 (99.7)	97.6 (85.7)	95.7 (77.2)	99.9 (99.9)	99.9 (99.8)
$(I/\sigma(I))$	18.6 (5.9)	20.3 (2.6)	31.6 (2.2)	33.2 (2.4)	24.4 (3.0)	16.4 (2.1)
R_{merge}[b]	7.6 (31.8)	9.1 (56.0)	4.9 (60.8)	4.5 (52.0)	5.3 (62.2)	6.6 (54.9)

Structure refinement

	Native CBM30	SeMet[a] (high energy remote)	Native PKD-CBM44
Resolution range	20.0-2.27 (2.328-2.27)	67.7-2.1 (2.155-2.1)	66.7-2.17 (2.226-2.17)
Reflections used	21,633 (1,545)	23,573 (1,729)	20,683 (1,483)
R_{work}[c] (%)	21.6 (29.5)	17.4 (24.5)	16.9 (26.5)
R_{free}[c] (%)	26.5 (35.4)	21.0 (28.4)	21.6 (28.5)
No. protein atoms	2778	1935	1935
No. water molecules	218	381	380
No. calcium ions	0	2	2
No. ethylene glycol molecules	0	3	12
RMSD from ideal geometry			
Bond lengths	0.007	0.009	0.010
Bond angles (º)	1.031	1.218	1.245

[a]SeMet data are for PKD-CBM44 incorporated with seleno-methionine. The peak data were collected from a different crystal from the drop. It suffered radiation damage during edge data collection, so another crystal from the same drop was used. For the higher energy edge data collection, beam attenuation was increased from 20% to 40%. The crystal was translated along the beam for the high energy remote data collection.

[b]$R_{merge}=\Sigma_h \Sigma_i |I(h,i)-<I(j)>|/\Sigma_h\Sigma_i I(h,i)$, where $I(h,i)$ is the intensity of the i^{th} measurement of reflection h and $<I(h)>$ is the mean value of $I(h,i)$ for all i measurements.

[c]$R_{work}=\Sigma||F_{calc}|-|F_{obs}||/\Sigma|F_{obs}|\times100$, where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. R_{free} is calculated for a randomly chosen 5% of the reflections for PKD-CBM44 and 10% for CBM30.

In relation to CBM30, subsequent to the generation of protein crystals, a structural model of the CBM was deposited in the PDB under the accession code 1WMX (Horiguchi *et al.*, unpublished results[1]). Therefore, this structure served as a model for the Patterson search methods, using program MOLREP (Vagin and Teplyakov, 1997) from the CCP4 suite (1994). Twenty cycles of rigid-body refinement of the molecular replacement solution, using program REFMAC5 (Murshudov *et al.*, 1997), gave an overall R_{work}=36.3% and R_{free}=36.9% (for a free set of 10% of the reflections). Iterative cycles of model building and restrained refinement brought the R_{work} and R_{free} to the final values of 21.7% and 26.4%, respectively. The final model contains 346 amino acid residues, belonging to two polypeptide chains, and 218 water molecules. All amino acid residues are in the allowed regions of the Ramachandran plot. The refinement statistics are summarized in Table 3.3.

3.2.3. Results and Discussion

3.2.3.1. A novel CBM family

A previous study showed that CelJ from *C. thermocellum* (designated *Ct*Cel9D-Cel44A in this study to reflect the modern nomenclature of GHs by Henrissat, 1998), contains GH9 and GH44 catalytic modules in addition to an internal dockerin that targets the enzyme to the clostridial cellulosome (Ahsan *et al.*, 1996), an N-terminal family 30 CBM (CBM30) and a C-terminal domain of unknown function (Figure 3.3). CBM30, which displays affinity for β-1,4-gluco-polymers (Arai *et al.*, 2003), plays a pivotal role in the function of GH9, a typical processive endoglucanase, whereas GH44 was assigned as displaying endo-xylanase activity (Ahsan *et al.*, 1996). The first 90 amino acids of the 250-residue C-terminal region of *Ct*Cel9D-Cel44A comprises a PKD module (the nature of this module is discussed below), whereas the remaining 160 amino acids display no extensive sequence identity with other modules present in GHs. The data presented in this study show that this C-terminal region of *Ct*Cel9D-Cel44A binds to plant structural polysaccharides (see below) and is thus classified as a novel CBM (designated CBM44), representing the founder member of CBM family 44. Alignment of CBM44 with CBM families 4, 6, 9, 16, 17, 22, 27, 29, 33, 35 and 37,

[1]Horiguchi, Y., Kono, M., Suzuki, A., Yamane, T., Anai, M., Sakka, K., and Ohmiya, K., unpublished results.

revealed the presence of three very short conserved motifs at the N and C termini of these domains (Figure 3.4) which, in CBM44 (see below) and in several other CBM families, play a key role in the coordination of a structural calcium. This observation provides further support for the grouping of CBM families that display a β-sandwich fold into a superfamily (Boraston *et al.*, 2004), analogous to the clan classification of GHs (see Chapter 1). Although the superfamily reflects an ancestral link between these CBM families, it provides no significant functional or mechanistic information, as ligand specificity and even the identity of the ligand-binding residues can vary within a family.

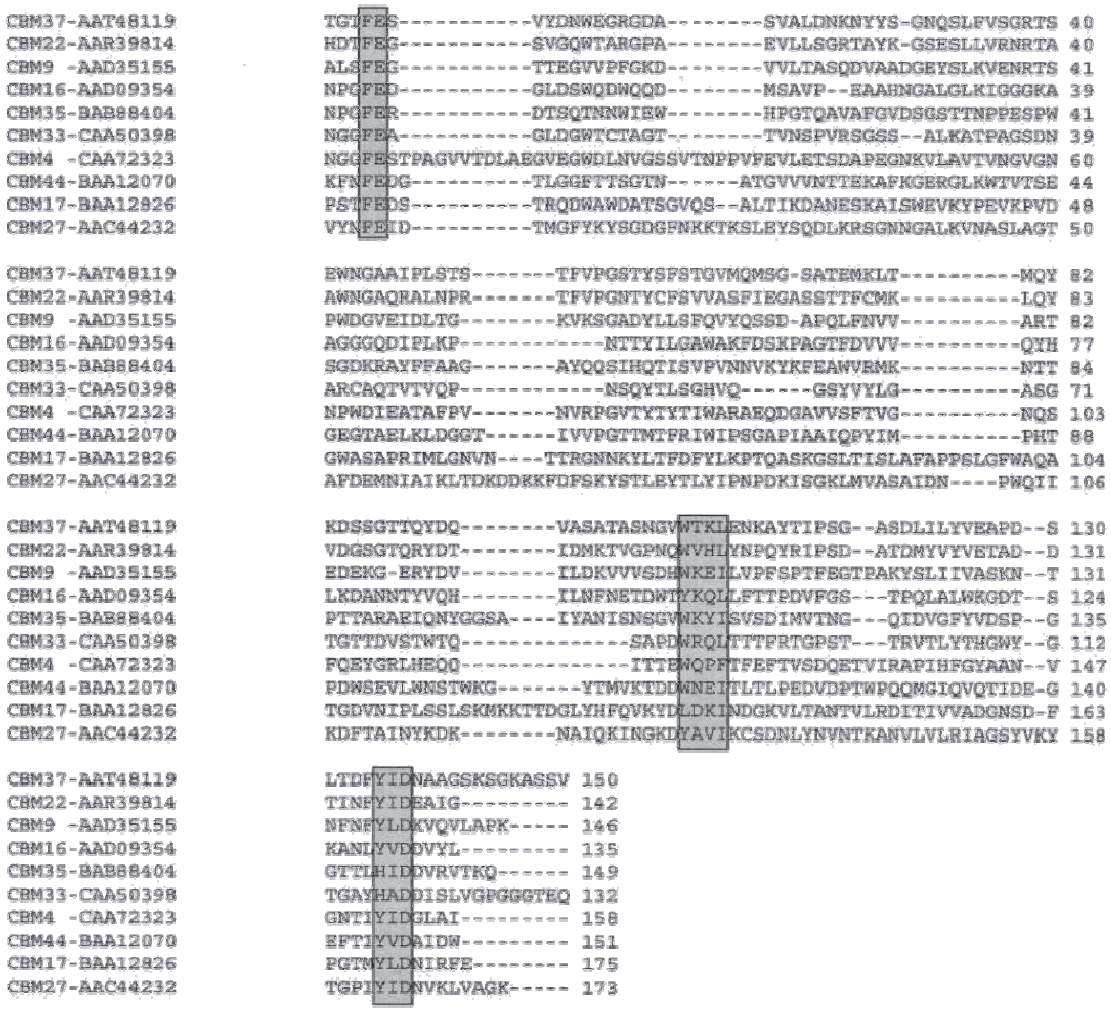

Figure 3.4. Alignment of CBM44 with CBMs from other related families.
Sequences used in the alignment are labelled with the accession number indicating their respective CBM family. The three conserved motifs responsible for calcium bonding in CBMs from the β-sandwich superfamily are boxed.

3.2.3.2. Ligand specificity of CBM30 and CBM44

CBM44 and PKD-CBM44 were purified to electrophoretic homogeneity, and their biochemical properties evaluated. AGE revealed that both proteins displayed significant affinity for xyloglucan, barley β-glucan, lichenan, HEC and konjac glucomannan, although exhibiting a significantly lower K_a for oat spelt xylan (Figure 3.5 and Table 3.4).

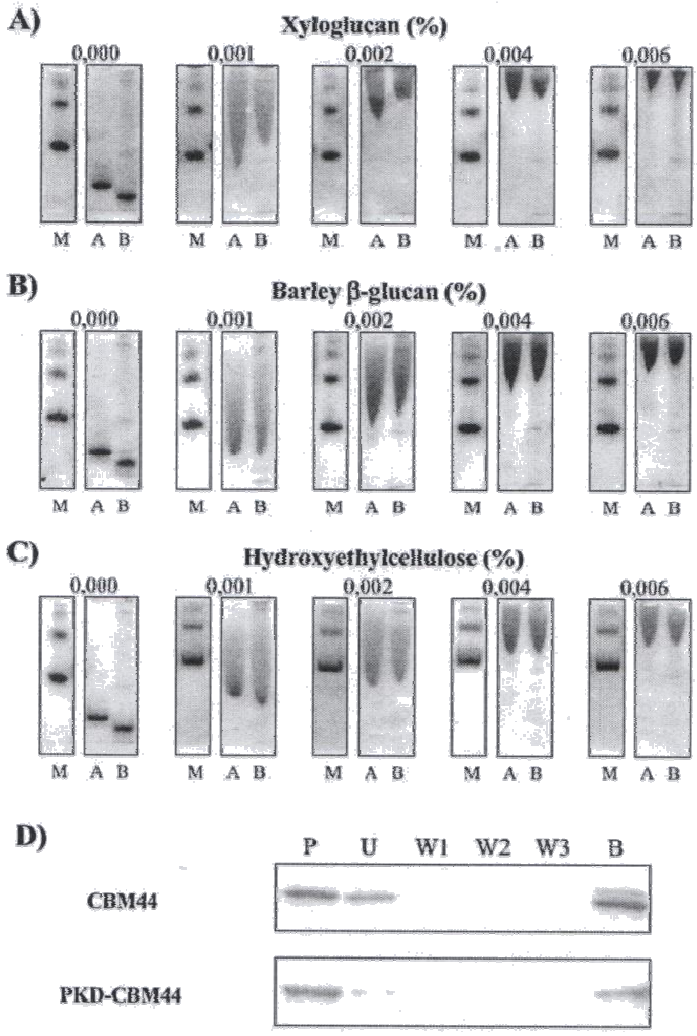

Figure 3.5. Quantitative AGE of PKD-CBM44 and CBM44 with xyloglucan, barley β-glucan and HEC as the ligand (A-C) and qualitative binding to the insoluble polysaccharide Avicel (D).
BSA (lane M), PKD-CBM44 (lane A) and CBM44 (lane B) were subjected to nondenaturing electrophoresis in gels containing a range of different concentrations of xyloglucan (A), barley β-glucan (B) and HEC (C). The percentage concentrations of the various soluble polysaccharides are displayed. Lanes in D: P, purified protein; U, unbound fraction; W1, W2 and W3, pellet washes; B, bound fraction.

No binding to laminarin, curdlan, carob galactomannan, potato galactan, pullulan, or pustulan was detected, suggesting that the CBM targets β-1,4 glucose polymers, consistent with the capacity of both proteins to interact with the insoluble cellulose preparation Avicel (Figure 3.5). The affinities of CBM44 and PKD-CBM44 for the various ligands were similar, suggesting that the PKD domain does not contribute to carbohydrate recognition. CBM44 is thus a Type B CBM, binding individual β-1,4-glucan chains both in soluble and insoluble polysaccharides. AGE revealed that CBM30 displays a similar ligand specificity profile to CBM44, although the affinity constants are about 10-15 fold lower for the N-terminal module (Table 3.4).

Table 3.4. Ligand specificity of CBM30, PKD-CBM44 and CBM44 modules quantified by AGE.

Soluble polysaccharide	K_a[a]		
	CBM30	**PKD-CBM44**	**CBM44**
Xyloglucan	43.6	686.3	726.3
Barley β-glucan	43.0	493.5	513.3
Lichenan	36.4	484.8	502.5
HEC	26.7	250.0	266.3
Konjac glucomannan	5.6	188.3	226.7
Oat spelt xylan	NB[b]	11.4	18.0
Carob galactomannan	NB	NB	NB
Laminarin	NB	NB	NB
Curdlan	NB	NB	NB
Galactan	NB	NB	NB
Pullulan	NB	NB	NB
Pustulan	NB	NB	NB

[a]The values are given as the ligand concentration (% w/v).
[b]K_a below 2.

3.2.3.3. Quantitative assessment of CBM30 and CBM44 ligand binding by ITC

The binding of CBM30 and CBM44 to their respective ligands was also investigated by ITC. The data, presented in Table 3.5 and Figure 3.6, confirm that CBM44 binds to xyloglucan, lichenan, β-glucan, HEC and glucomannan with similar affinity, although xyloglucan is its preferred polysaccharide ligand. The observation that both CBM30 and CBM44 do not interact with galactomannan indicates that the axial O2 of mannose makes steric clashes with the protein at one or more sugar-binding sites. The much reduced affinity of both CBM30 and CBM44 for xylan may reflect the importance of a direct interaction between the exocyclic O6 of glucose with the proteins, although it is also possible that the orientation of the aromatic platform in the

binding site discriminates against ligands that adopt the 3-fold helical conformation displayed by the xylose polymer (Atkins, 1992).

Table 3.5. Affinity of CBM44 and CBM30 for oligo- and polysaccharides as determined by ITC.

Protein	Ligand	$K_a \times 10^4$	ΔG	ΔH	$T\Delta S$	n^a
		M^{-1}	*kcal mol*$^{-1}$	*kcal mol*$^{-1}$	*kcal mol*$^{-1}$	
CBM44	Cellohexaose	72.8 ± 7.2	-8.0 ± 0.1	-15.9 ± 0.3	-7.9 ± 0.4	1.1 ± 0.1
CBM30	Cellohexaose[b]	6.4 ± 0.8	-6.2 ± 0.1	-8.0 ± 0.5	-1.8	1.2 ± 0.1
CBM44	Cellopentaose	6.6 ± 1.3	-6.6 ± 0.1	-14.5 ± 0.5	-7.9 ± 0.6	1.0 ± 0.0
CBM30	Cellopentaose[b]	1.2 ± 0.8	-5.3 ± 0.3	-6.9 ± 0.5	-1.7	1.3 ± 0.1
CBM44	Xyloglucan	81.6 ± 9.8	-8.1 ± 0.1	-16.3 ± 0.6	-8.2 ± 0.7	1.0 ± 0.0
CBM30	Xyloglucan	7.2 ± 1.4	-6.6 ± 0.1	-10.4 ± 0.3	-3.8 ± 0.2	1.0 ± 0.0
CBM44	HEC	12.2 ± 3.3	-6.9 ± 0.2	-12.5 ± 0.5	-5.6 ± 0.7	1.0 ± 0.0
CBM30	HEC	4.5 ± 0.5	-6.3 ± 0.1	-10.0 ± 0.2	-3.7 ± 0.3	1.0 ± 0.0
CBM44	β-glucan	22.5 ± 3.6	-7.3 ± 0.1	-17.7 ± 0.6	-10.4 ± 0.7	1.0 ± 0.0
CBM30	β-glucan	2.8 ± 0.3	-6.1 ± 0.1	-11.2 ± 0.2	-5.1 ± 0.3	1.0 ± 0.0
CBM44	Lichenan	12.3 ± 2.8	-6.9 ± 0.1	-22.6 ± 1.3	-15.7 ± 1.4	1.0 ± 0.0
CBM30	Lichenan	3.6 ± 0.4	-6.2 ± 0.1	-11.5 ± 0.4	-5.3 ± 0.4	1.0 ± 0.0
CBM44	Glucomannan	9.0 ± 2.0	-6.7 ± 0.1	-15.9 ± 0.8	-9.2 ± 0.9	1.0 ± 0.0
CBM30	Glucomannan	~ 0.4[c]	-	-	-	-

[a]Number of binding sites on the protein; [b]Data are from Arai *et al*. (2003); [c]Value is an estimate because affinity was too low to obtain accurate value.

Titration of CBM44 with β-1,4-linked gluco-oligosaccharides showed that the protein displays no affinity for cellotriose but binds with increasing K_a to cellotetraose, cellopentaose and cellohexaose. The stoichiometry of binding for each cello-oligosaccharide is ~1 indicating that CBM44 has a single ligand-binding site. The interaction of both CBM30 and CBM44 with oligosaccharides and polysaccharides is enthalpy-driven, with entropy making an unfavorable contribution to ligand binding. This pattern of energetics is typical of the binding of CBMs to soluble saccharides and the molecular basis for these thermodynamic changes has been widely discussed (Notenboom *et al.*, 2001a; Xie *et al.*, 2001a; Boraston *et al.*, 2002; Charnock *et al.*, 2002). The observation that cellotetraose and the xyloglucan heptasaccharide XXXG (the most common repeating unit of xyloglucan) bind to CBM44 with similar affinities is entirely consistent with the polysaccharide data, indicating that the side-chains in xyloglucan do not make steric clashes with the CBM but also do not contribute to binding and thus do not represent specificity determinants.

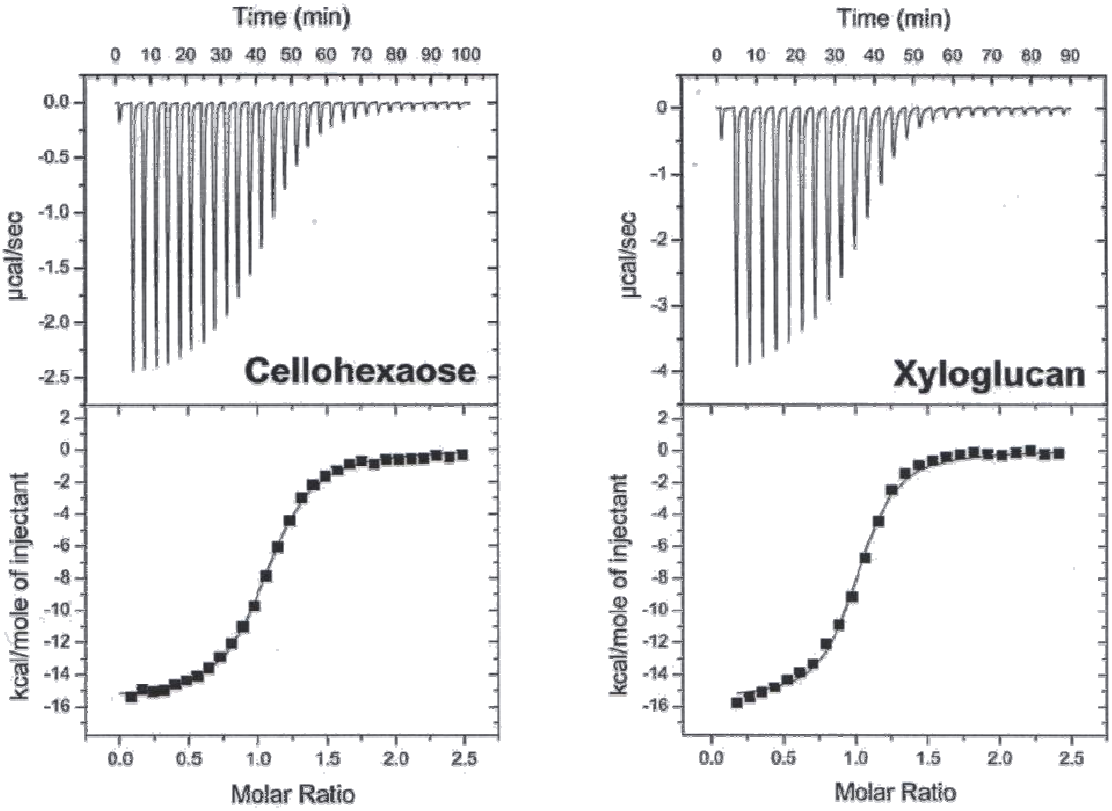

Figure 3.6. Representative ITC data of CBM44 binding to cellohexaose and xyloglucan.
The upper parts of each panel show the raw binding heats, the lower parts show the integrated binding heats minus the dilution control heats fitted to a single-site binding model. In the cellohexaose titration, the ligand was at 0.5 mM and CBM44 at 40 µM. In the xyloglucan titration, the ligand is at 2 mg/mL and CBM44 at 75 µM.

The observation that CBM30 and CBM44 bind to xyloglucan provides the first evidence that CBMs are able to accommodate the side-chains of this decorated glucan. The specificity of the two CBMs is entirely consistent with the substrate specificity of the appended GH44 (see below). Because previous studies on CBMs have not questioned their affinity for xyloglucan, we explored whether the capacity to bind this polysaccharide is a generic feature of CBMs that recognize single β-glucan chains. We used ITC to determine the affinity of *Cf*CBM4-1, *Ct*CBM11, *Cc*CBM17, *Bsp*CBM28 and *Pe*CBM29-2 (which all bind to β-linked glucans) for xyloglucan. The data show (see Table 3.6) that all five modules recognize xyloglucan. It is likely, however, that the xyloglucan side-chains make some steric clashes with the protein surface of *Ct*CBM11, *Cc*CBM17 and *Bsp*CBM28 because these modules bind to the decorated glucan less tightly than cellohexaose, which fully occupies the respective ligand-binding sites. The similar affinity of xyloglucan and cellohexaose for *Cf*CBM4-1 indicates that the side-chains of the decorated glucan do not make significant steric clashes with the protein.

The modest elevation in affinity of *Pe*CBM29-2 for xyloglucan, compared with cellohexaose, may indicate that the xylose side-chains are making a positive contribution to binding. It should be noted, however, that *Pe*CBM29-2 binds, through cooperative interactions, more tightly to polysaccharides than the corresponding oligosaccharides (which occupy the complete binding site), and thus the 3-fold increase in affinity for xyloglucan over cellohexaose may not be related to the capacity of the protein to recognize the xylose side-chains (Charnock *et al.*, 2002).

Table 3.6. Affinity of glucan-binding CBMs for xyloglucan and cellohexaose as determined by ITC. Abbreviations: *Cf*, *Cellulomonas fimi*; *Ct*, *Clostridium thermocellum*; *Cc*, *Clostridium cellulovorans*; *Bsp*, *Bacillus* spp; *Pe*, *Piromyces equi*.

Protein	$K_a \times 10^4$	
	Cellohexaose	Xyloglucan
	M^{-1}	
*Cf*CBM4-1	8.7	4.4
*Ct*CBM11	9.7	0.6
*Cc*CBM17	8.7	1.7
*Bsp*CBM28	4.8	0.4
*Pe*CBM29-2	1.8	5.9

Inspection of the three-dimensional crystal structures of *Cf*CBM4-1 (Boraston *et al.*, 2002), *Cc*CBM17 (Notenboom *et al.*, 2001a) and *Pe*CBM29-2 (Charnock *et al.*, 2002) in complex with cello-oligosaccharides provides some insight into the mechanism by which these proteins bind xyloglucan. The most common repeating unit in xyloglucan is XXXG. It is likely that the undecorated glucose (G) binds to subsite +2 in *Cf*CBM4-1 and subsite +3 in *Cc*CBM17 and *Pe*CBM29-2, because at these locations the exocyclic O6 of the hexose sugar is pointing directly at the protein surface and thus a xylose side-chain would make a steric clash with the CBM preventing ligand binding. In *Cc*CBM17, the other three subsites could all accommodate a glucose substituted at O6 with a xylose. The protein, however, displays maximal binding to cellohexaose (~50-fold tighter than cellotetraose). Thus, the steric clashes that reduce affinity for xyloglucan (Table 3.6) are likely to be in subsites distal to those identified in the crystal structure of the CBM in complex with cellotetraose. In the *Cf*CBM4-cellopentaose complex the exocyclic O6 of Glc^{+4} (in addition to Glc^{+2}) appears to point toward the protein surface, and thus decoration of this sugar with xylose would cause a significant steric clash, which is contrary to the biochemical data that show a similar affinity of

*Cf*CBM4-1 for xyloglucan and cellohexaose (Table 3.6). Glc^{+4} (and Glc^{+5}), however, makes no interactions with CBM4, and thus it is highly likely that this sugar is able to rotate such that O6 is pointing into solvent, without causing a significant loss in affinity. In *Pe*CBM29-2 Glc^{+1}, in addition to Glc^{+2}, would clash with the protein if it is decorated with xylose at O6. As Glc^{+1} makes a single hydrogen bond with the CBM, the loss of this interaction, which would be required in a rotation of the sugar such that O6 points into solvent, is likely to incur only a modest energetic penalty. It should be noted, however, that the determination of the precise conformation of xyloglucan when bound to CBMs requires the resolution of the crystal structure of these proteins in complex with oligosaccharides derived from the polysaccharide.

3.2.3.4. CBM44 potentiates cellulase activity of Cel44A through a targeting effect

To evaluate the role of CBM44 in the function of *Ct*Cel9D-Cel44A, truncated derivatives of the enzyme comprising the GH44 catalytic module (Cel44A) and Cel44A fused to CBM44 (Cel44A-PKD-CBM44) were expressed in *E. coli* and purified to electrophoretic homogeneity. The temperature optimum of Cel44A is ~75 ºC with maximal activity at pH 5.0.

The biochemical properties of truncated derivatives of *Ct*Cel9D-Cel44A indicate that Cel44A displays mixed endo-β-1,4-glucanase/xylanase activity, hydrolysing cellulosic substrates such as CMC and mixed linked β-1,4-β-1,3-glucans, while exhibiting reduced but still significant activity against oat spelt xylan (Table 3.7). The enzyme did not hydrolyse laminarin, pustulan, galactan or galactomannan. Most interestingly, the enzyme displays significant activity against glucomannan, and its activity against the branched β-1,4-glucose polymer xyloglucan is comparable with its capacity to hydrolyse CMC or lichenan (Table 3.7). Cel44A is a GH44 enzyme, and glycoside hydrolases in this family display predominantly endo-β-1,4-glucanase or xyloglucanase activity (http://afmb.cnrs-mrs.fr/CAZY; April 2008). The data reported here show that Cel44A is indeed a cellulase/xyloglucanase that displays some xylanase activity and not a xylanase, as reported previously (Ahsan *et al.*, 1996).

Table 3.7. Specific activities of Cel44A and Cel44A-PKD-CBM44 for a range of plant cell wall polysaccharides and targeting effect of CBM44.

Substrate	Enzyme activity (k_{cat}[a])	
	Cel44A	**Cel44A-PKD-CBM44**
Specific activity[b]		
Avicel	2.9	3.9
Xyloglucan	149.0	159.0
Lichenan	299.8	330.3
CMC	228.9	283.1
Glucomannan	112.6	117.6
Oat spelt xylan	83.2	87.6
Laminarin	0	0
Pustulan	0	0
Galactomannan	0	0
Targeting effect of CBM44[c]		
Xyloglucan	105.9	97.4
Xyloglucan + laminarin + pustulan	39.4	92.5
CMC	105.6	91.0
CMC + laminarin + pustulan	60.1	106.5

[a]Values are mol of product formed per mol of enzyme min^{-1}.
[b]Assays for determining specific activity performed at 50 ºC with 0.5% (w/v) of substrate except for Avicel where 2% (w/v) substrate was used.
[c]For testing the targeting effect of CBM44, enzyme activities were measured at 50 ºC with 0.15% (xyloglucan) or 0.25% (CMC) of the target polysaccharide (w/v). Where appropriate, 0.68% (w/v) of laminarin and pustulan were included in the reaction.

A proposed role for CBMs is to maintain the proximity of the cognate catalytic domains within the context of a complex macromolecular structure such as the plant cell wall (Boraston *et al.*, 2004). This so-called targeting function could orchestrate the distribution of a large repertoire of enzymes to the different polysaccharides located in plant cell walls. Therefore, in *Ct*Cel9D-Cel44A, CBM44 would direct Cel44A to its target substrates, the branched and unbranched β-1,4-glucose cell wall polymers, in the presence of a range of other polysaccharides. To test this hypothesis, the catalytic efficiency of both Cel44A and Cel44A-PKD-CBM44 against xyloglucan and CMC was evaluated in the presence and absence of polysaccharides that do not act as substrates for this enzyme. The data presented in Table 3.7 suggest that CBM44 "rescues" the activity of Cel44A in the presence of the polysaccharides laminarin and pustulan**,** which, although not acting as substrate can presumably make non-productive interactions with the active site of the enzyme. Therefore, we propose that in the context of the plant cell wall, CBM44 potentiates the activity of *Ct*Cel44A by targeting the enzyme to the regions of this macromolecular structure where Cel44A substrates are located.

3.2.3.5. Crystal structure of CBM44

The three-dimensional structure of PKD-CBM44 was solved using MAD methods with a SeMet derivative crystal. The final model of PKD-CBM44 contains two calcium ions (Ca1 and Ca2). The structure reveals the presence of two separated domains, PKD and CBM44, both adopting a β-sandwich fold (Figure 3.7a). Ca1, located in the PKD module, has an octahedral coordination; the side-chain O atoms of Asn^{4} (Oδ1), Asp^{33} (bidentate coordination from Oε1 and Oε2), Asp^{35} (Oδ1), Asp^{76} (Oδ1), and the main chain O atom of Gly^{5} interact with the metal ion with a single water molecule completing the coordination sphere (Figure 3.7b). A second calcium binding site (Ca2) was identified in CBM44. Ca2 also shows octahedral coordination and is bound to residues Asn^{101}, Lys^{130} and Arg^{133} (main chain O atoms), Asp^{96} (Oδ1), Glu^{103} (Oε1) and Asp^{245} (bidentate coordination from Oε1 and Oε2). The distance between the ligands and the calcium ions varies from 2.3 to 2.6 Å (Figure 3.7b).

In general, β-sandwich CBMs have at least one calcium-bound metal, which is often solvent inaccessible and is suggested to stabilize the protein fold (Boraston *et al.*, 2004). The importance of calcium in maintaining the structural integrity of a family 4 CBM was demonstrated by showing that removal of the metal from the xylan-binding module reduced its melting temperature by 23 ºC, although this had no effect on ligand binding (Abou-Hachem *et al.*, 2002). It has been shown, however, that in a family 35 and a family 36 CBM calcium plays a direct role in ligand recognition (Jamal-Talabani *et al.*, 2004; Tunnicliffe *et al.*, 2005). The calcium-binding sites in the PKD domain and CBM44 are buried inside the protein structures supporting, in both cases, a structural role for the metal. It is interesting to note that CBM44 Glu^{103} and Asp^{245}, which play a key role in Ca2 coordination, are residues belonging to the first and third amino acid consensus motifs, respectively, and are present in several members of the β-sandwich CBM superfamily (see 3.2.3.1. and Figure 3.4). Furthermore, all the other residues belonging to these motifs are in the vicinity of the calcium ion, including the structural Trp^{209} that belongs to the second motif and is highly conserved in the β-sandwich superfamily. Together, the data suggest that these three motifs fulfil a structural role of extreme importance for the CBM architecture by creating the required topological environment for calcium coordination. Therefore, it can be predicted that the primary sequence of all β-sandwich CBMs containing a structural calcium ion that is on the opposite face to the ligand-binding site should contain these three consensus motifs.

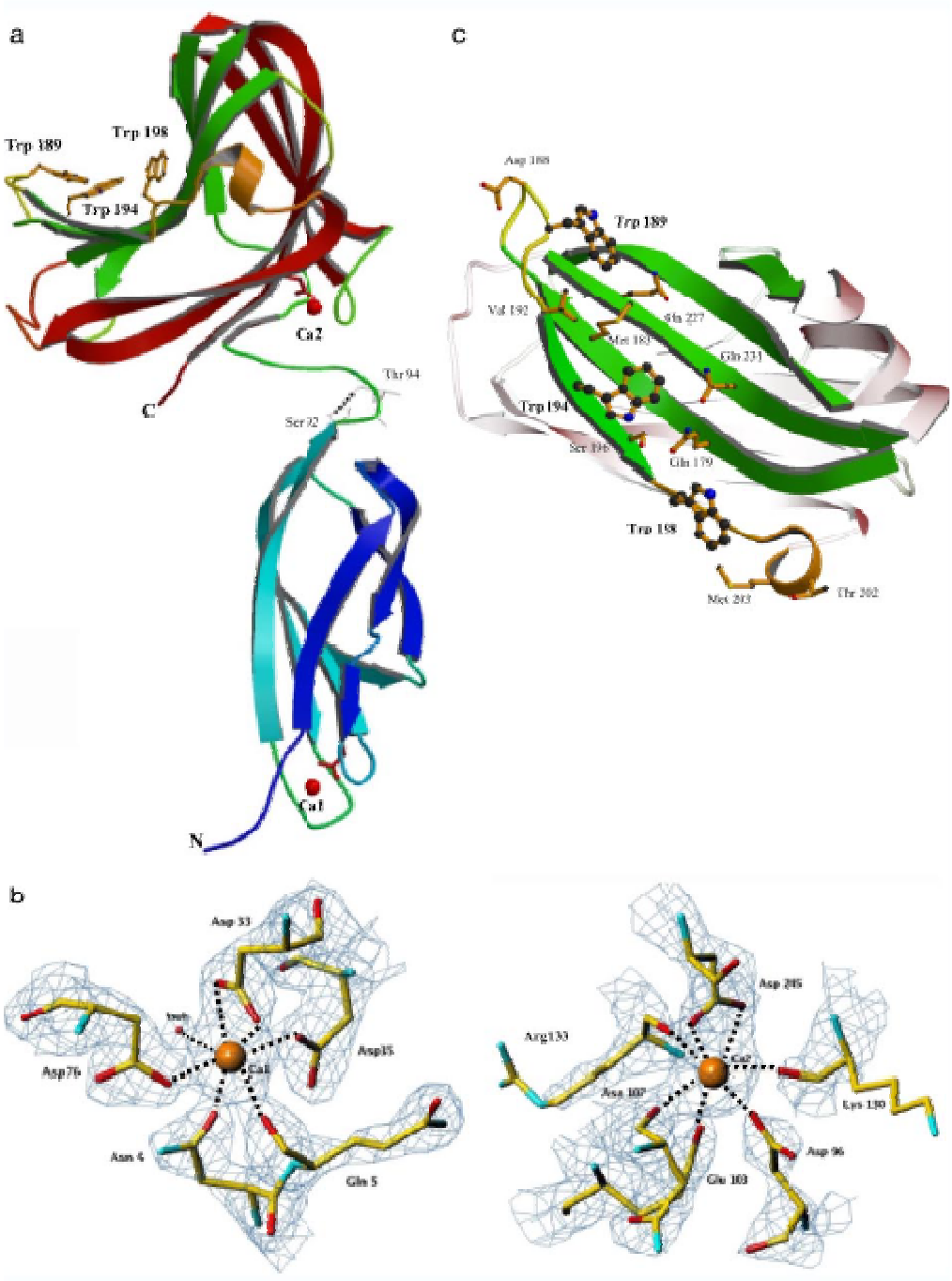

Figure 3.7. The three-dimensional structure and the hydrophobic platform of PKD-CBM44.
The overall structure of PKD-CBM44 highlighting the secondary structural elements, with the β-strands of each β-sheet coloured the same is shown in (a). The tryptophans in the binding cleft are depicted as sticks, and the calcium atoms depicted as red spheres. The H-bond between the OH of Ser^{92} and carbonyl O of Thr^{94} in the linker region is shown in cyan. Note: β-strand 2 of PKD has a kink at residues 29/30. (b) Comparison of the calcium-binding sites Ca1 in the PKD (left) and Ca2 in the CBM44 (right) domains and their corresponding coordinating amino acid residues. The electron density was contoured at 2σ. (c) Top down view of the CBM44 binding cleft showing the amino acids (as sticks) that are in the vicinity of the three tryptophan residues (as balls-and-sticks) that comprise the hydrophobic platform. All the ribbon figures in Figure 3.7, Figure 3.8 and Figure 3.10 were prepared using MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Bacon, 1997) and the electron density figures with TURBO-FRODO (Roussel and Cambillau, 1991).

The PKD domain is built from two β-sheets with three and four β-strands packed face to face and displays features typical of a β-sandwich (Figure 3.7a). The four-stranded β-sheet has a slightly concave topology. The number and arrangement of these strands are identical to several members of the immunoglobulin and fibronectin Type III superfamilies, and thus the PKD domain can be defined as displaying an immunoglobulin-like β-sandwich fold (Bycroft *et al.*, 1999). The core of the PKD β-barrel is highly hydrophobic, including four phenylalanines, three tyrosines and one tryptophan. All these residues are conserved in the PKD domain family (data not shown). PKD domains are so-called because 16 copies of this module were originally identified in the extracellular segment of polycystin-1, a large cell surface glycoprotein encoded by the *pkd1* gene, which is mutated in autosomal dominant polycystic kidney disease (Bycroft *et al.*, 1999). One or more copies of the PKD domain are also found in several other extracellular proteins from higher organisms, eubacteria and archaebacteria. PKD domains are present in ~39 GHs (primarily GH18 chitinases), although their role in the function of these enzymes have not been extensively investigated. It is interesting to note, however, that a recent study showed that the PKD module of chitinase A from *Alteromonas* sp strain O-7 binds to chitin domain and the residues Trp^{30} and Trp^{67} mediate the interaction with the polysaccharide (Orikoshi *et al.*, 2005). Biochemical data presented here suggest that the PKD domain in *Ct*Cel9D-Cel44A does not modulate the function of CBM44 when binding to soluble and insoluble polysaccharides nor does it bind carbohydrates *per se*.

The sugar-binding Trp^{30} of the chitin-binding PKD, described above, is not conserved in *Ct*Cel9D-Cel44A PKD, whereas the equivalent residue to Trp^{67} (plays a role in sugar recognition in the chitinase PKD) in the clostridial module (Trp^{43}) is buried and therefore belongs to the hydrophobic core of the protein. It is possible that this domain functions as a non-flexible spacer domain in *Ct*Cel9D-Cel44A, although the role of these modules in GHs is clearly variable. The PKD domains of polycystin-1, however, are involved in inter-molecular interactions, playing an important role in intercellular adhesion (Beskrovnaya *et al.*, 2000).

The PKD-CBM44 structure forms a dimer related by a crystallographic 2-fold. The four-stranded β-sheet of the PKD domain interacts with CBM44 via four hydrogen bonds between the parallel strands 7 of the PKD domain (residues 81-90) and strand 13 of the CBM44 (residues 190-196). Therefore, it is possible that PKD modules

participate in specific protein-protein interactions within the cellulosome, which may be important in orchestrating the binding of defined repertoires of plant cell wall hydrolases in each individual multienzyme complex. Residues Ser^{92} to Gly^{95} constitute the interdomain linker sequence connecting the PKD domain to CBM44. Most interestingly, Ser^{92} hydrogen bonds to the main chain O of Thr^{94} suggesting that, at least under some conditions, the linker sequence does not possess significant flexibility (Figure 3.7a). The CBM44 structure (Figure 3.7a) reveals a classic distorted β-jelly roll fold consisting of two five-stranded anti-parallel β-sheets, which form a convex (β-strands 9, 12, 14, 15 and 17) and concave face (β-strands 10, 11, 13, 16 and 18). The core of the β-barrel is highly hydrophobic and includes five phenylalanine and three tryptophan residues. As in other CBMs that bind individual polymer chains, the concave side of CBM44 forms a cleft. In CBM44, this surface depression is defined by polypeptide stretches Thr^{111}–Thr^{115}, Gly^{145}–Leu^{151}, Ala^{177}–His^{185}, Val^{192}–Met^{203}, and Gln^{226}–Thr^{234} and strongly resembles the equivalent grooves observed in other Type B CBMs. This cleft is decorated with the side-chains of Thr^{111}, Ser^{113}, Thr^{115}, Glu^{144}, Thr^{146}, Glu^{148}, Lys^{150}, Asp^{152}, Gln^{179}, Tyr^{181}, Met^{183}, His^{185}, Trp^{189}, Trp^{194}, Ser^{196}, Trp^{198}, Gln^{227}, Gln^{231} and Gln^{233} (Figure 3.7c). DALI structural similarity searches reveal that CBM44 is most similar to CBMs of families 22 (1DYO), 15 (1GNY), 27 (1 of 3), 11 (1V0A), 4 (1GUI and 1ULO), 29 (1 W8U), 28 (1UWW) and 17 (1J83) with more than 129 matching Cα positions and a RMSD of less than 3.1 Å. All these proteins are members of the β-sandwich CBM superfamily and, apart from CBM15 and CBM29, which operate in nature at lower temperatures than the other modules, contain a structural calcium ion in a similar position to CBM44.

3.2.3.6. Crystal structure of CBM30

Structure solution of CBM30, to a resolution of 2.27 Å, revealed two copies (chains A and B) of the polypeptide chain in the asymmetric unit of the crystal, related by a non-crystallographic 2-fold axis. Chain A of the CBM30 model consists of 174 amino acids, whereas the purified protein contains 207 amino acids, including the C-terminal His_6-tag (Figure 3.8a).

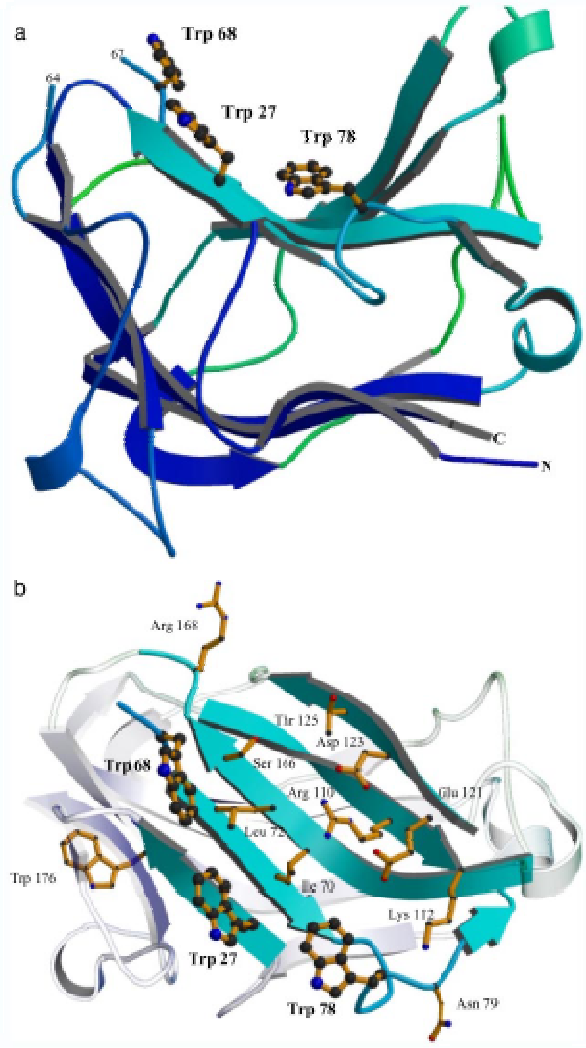

Figure 3.8. The three-dimensional structure and the hydrophobic platform of CBM30.
(a) The overall structure of CBM30 monomer looking side-on down the binding cleft, highlighting the secondary structural elements, with the β-strands of each β-sheet colored the same; the aromatic residues involved in ligand binding are shown in ball-and-stick. (b) The binding cleft of CBM30 highlighting the amino acids, depicted as sticks, that are in the vicinity of the three tryptophan residues, depicted as ball-and-stick.

The N-terminal stretch from residue Ser^{1}A to Lys^{12}A, Gln^{65}A, Ser^{66}A, and the C-terminal region extending from Glu^{188}A to His^{207}A, which includes the His_6-tag, were disordered and thus were not included in the final model. The model for chain B was identical to chain A, except for the inclusion of one additional residue, Lys^{12}B. The model includes 218 water molecules. CBM30 adopts a β-sandwich fold in which the two β-sheets each contain five β-strands (Figure 3.8a). The two β-sheets are connected mainly by loops, although one of these linking regions, Val^{44}-Thr^{49}, is a β-strand. The first β-sheet (β-sheet 1) includes β-strands 1 (Val^{19}-Phe^{19}), 4 (Pro^{54}-Val^{60}), 11 (Phe^{173}-Asp^{179}), 9 (Gln^{139}-Pro^{144}) and 6 (Tyr^{91}-Gly^{98}), whereas the second β-sheet (β-sheet 2)

consists of strands 2 (Gly^{26}-Ser^{30}), 5 (Trp^{69}-Thr^{74}), 7 (Val^{106}-Asp^{111}), 8 (Asp^{123}-Val^{127}) and 10 (Val^{160}-Ser^{166}). Each contains a hydrophobic core, which comprises 11 leucine, 8 isoleucine and 8 phenylalanine residues. The contacts between the two monomers (water-mediated and interactions between the loop Lys^{112}-Ile^{122} in both molecule A and B) is not biologically relevant because the protein is monomeric in solution.

As observed in other Type B CBMs, the concave side of the β-sandwich forms a cleft that is defined by β-sheet 2. Decorating the CBM30 groove are the residues Trp^{27}, Trp^{68}, Ile^{70}, Leu^{72}, Trp^{78}, Asn^{79}, Arg^{110}, Lys^{112}, Glu^{121}, Asp^{123}, Thr^{125}, Ser^{166} and Arg^{168} (Figure 3.8b). The aromatics are aligned at the edge of the cleft and comprise the hydrophobic platform. Structural comparison of CBM30 with the 1WMX model, which was used to solve the phase problem by molecular replacement methods, shows no significant differences in the overall structure. All β-sheets and loops are conserved, and the structural differences are restricted to some side-chains. The 1WMX structure (deposited in PDB) includes a stretch of 16 residues in the C-terminal region of only one of the monomers (from residue Glu^{180}B to Pro^{198}B). This sequence lies in between the interface of a symmetry-related dimer. Because of disorder, the corresponding region (from residue Glu^{188}B to Pro^{203}B) is not visible in the structure of CBM30 reported here.

A DALI data base search for homologous macromolecules reveals structural homology of CBM30 to several members of the β-sandwich superfamily, especially CBM11 (1V0A), CBM29 (1W8U), CBM4 (1GUI), CBM27 (1PMH), CBM15 (1GNY), CBM22 (1DYO), CBM6 (1UXZ) and CBM36 (1UX7). The RMSD of these structures, when superimposed on CBM30, vary in the range of 2.3-3.7 Å. Most interestingly, although a member of the β-sandwich superfamily of CBMs, CBM30, like CBM15 and CBM29, does not contain any calcium ion in its structure (Figure 3.8a). Thus, although bound calcium is a common feature of thermostable members of this superfamily, it is not a requirement for stability at elevated temperatures. The lack of a bound calcium explains why CBM30 does not contain the three consensus amino acid motifs of the superfamily that coordinate the metal (see 3.2.3.5.).

3.2.3.7. Probing the location of the ligand-binding residues in CBM30 and CBM44

Inspection of the surface clefts of CBM30 and CBM44 (Figure 3.7 and Figure 3.8) reveals the presence of three aromatic residues on the edge of both putative binding

sites. Most intriguingly, of these residues only Trp^{27} and Trp^{68} are invariant in family CBM30 members that contain the complete sequence (two CBM30 sequences lack the region that includes these two residues; see below), whereas Trp^{78} is conserved in only two of the seven sequences that comprise this family, suggesting differential binding mechanisms between members of CBM30. To probe the importance of the solvent-exposed residues Trp^{27}, Trp^{68}, and Trp^{78} in CBM30, the mutant proteins $W27A_{30}$, $W68A_{30}$ and $W78A_{30}$ were produced, and their biochemical properties were compared with wild-type CBM30. The data (Figure 3.9a) show that $W27A_{30}$ and $W68A_{30}$ display no significant affinity for decorated and undecorated β-glucans or glucomannan, whereas the control protein $W176A_{30}$ with a mutation on a surface Trp that is not located on the putative cleft, exhibits similar ligand binding properties to the wild-type protein. By contrast, $W78A_{30}$ displays reduced, but still significant, affinity for xyloglucan and barley β-glucan (~10-fold lower than wild-type CBM30) and no detectable binding to glucomannan, although a 10-fold reduction in the K_a for glucomannan is likely to result in no retardation of electrophoretic migration in gels containing 0.1% (w/v) of the polysaccharide. Taken together, these results confirm that the groove located on the concave surface of CBM30 comprises the ligand-binding site, and that the three aromatic residues located on the edge of the cleft are involved in ligand recognition.

In general, removal of the ligand-binding aromatic residues in CBMs completely abolishes polysaccharide binding (Pell *et al.*, 2003; Henshaw *et al.*, 2004; Flint *et al.*, 2005). In CBM30, however, it is apparent that Trp^{78} plays a less important role in ligand binding than the other two aromatic residues, which is entirely consistent with the observation that although Trp^{27} and Trp^{68} are invariant (except for two proteins that contain N-terminal truncations) in family 30 CBMs, Trp^{78} is not highly conserved. In CBM29 Tyr^{46}, located in a similar position to Trp^{78}, at one end of the binding site, was less important in ligand recognition than the other two tryptophan residues in the cleft (Flint *et al.*, 2005), Figure 3.10. It was suggested that the tyrosine makes less extensive hydrophobic interactions with the sugar, and here a similar explanation for the retention of some ligand recognition in the $W78A_{30}$ mutant is proposed. Most interestingly, two members of CBM30 family do not contain the N-terminal 80-residue sequence where the identified binding residues are located, and it is possible that these proteins do not retain a carbohydrate-binding function.

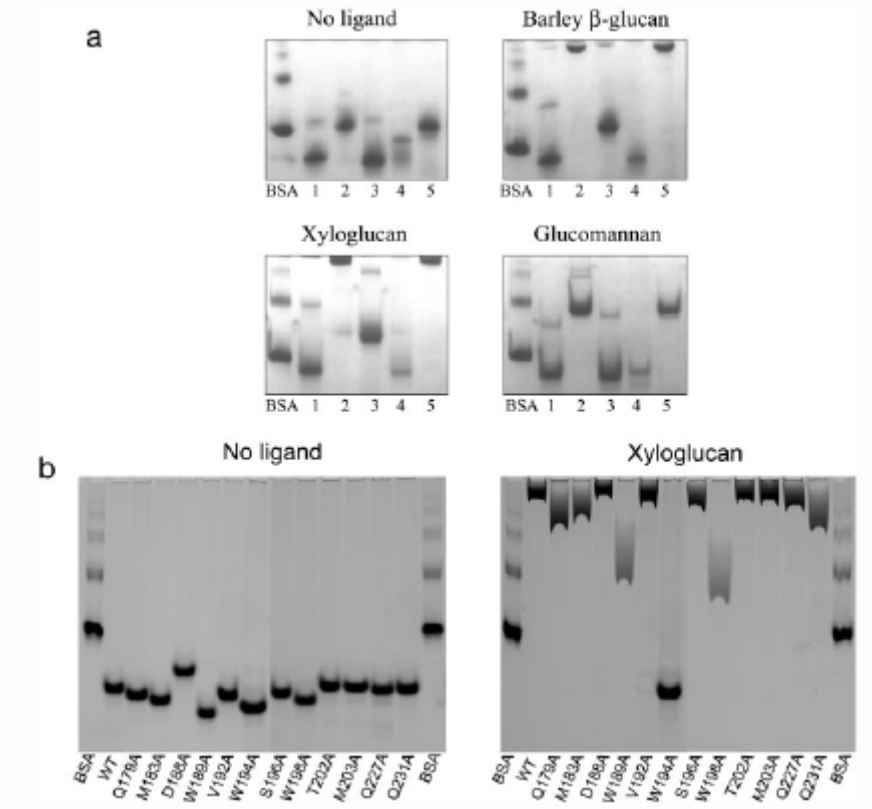

Figure 3.9. Binding of CBM30 (*a*) and CBM44 (*b*) alanine mutants to soluble polysaccharides analized by AGE.
(a) BSA is the non-interacting negative control (lanes BSA), $W68A_{30}$ (lanes 1), $W176A_{30}$ (lanes 2), $W27A_{30}$ (lanes 3), $W78A_{30}$ (lanes 4) and CBM30 wild-type (lanes 5) were subjected to non-denaturing electrophoresis in gels containing the specified ligands. (b) CBM44 wild-type and alanine mutants were subjected to non-denaturing electrophoresis in a gel containing 0.0125% (w/v) xyloglucan.

In CBM44 the loss of the central aromatic residue in the binding site, Trp^{194}, completely abrogates ligand recognition, confirming that the groove located on the concave surface of the protein comprises the ligand-binding site (Figure 3.8c and Figure 3.9b). In contrast to $W194A_{44}$, removal of the tryptophan residues that flank Trp^{194} (Trp^{189} and Trp^{198}) caused a relatively modest decrease in affinity (Table 3.8, Figure 3.8c and Figure 3.9b). It is possible that Trp^{194} makes a stronger hydrophobic interaction with the glucan chain than the flanking tryptophans, although the dominance of this amino acid may also reflect an additional hydrogen bond between the indole nitrogen of Trp^{194} and the polysaccharide. Indeed, the central aromatic amino acid, Trp^{54}, in the ligand binding site of the CBM2a of *Cellulomonas fimi* Cel6A also plays a more important role in cellulose binding than the flanking tryptophans. McLean *et al.* (2000) proposed that an additional hydrogen bond between the pyrrolic amine of this residue and cellulose explains why Trp^{54} plays a dominant role in ligand binding.

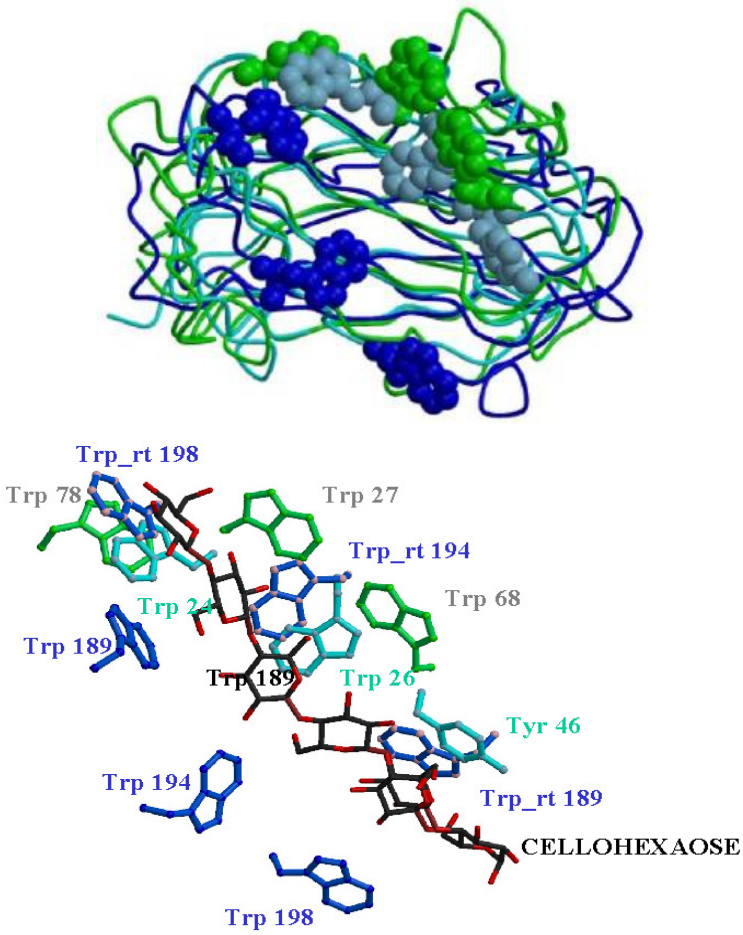

Figure 3.10. Superpositioning of CBM30 and CBM29 on CBM44.
CBM44 is shown in blue, CBM30 in green and CBM29 (PDB code 1GWM) in sky blue trace, with the aromatics (Trp^{189}, Trp^{194} and Trp^{198} for CBM44; Trp^{27}, Trp^{68} and Trp^{78} for CBM30; and Trp^{24}, Trp^{26} and Tyr^{46} for CBM29) as ball-and-sticks.

The data described above clearly show that the hydrophobic platform provided by the three tryptophan residues in both CBM30 and CBM44 play a pivotal role in ligand binding (Figure 3.10). The orientation of these residues in CBM44 is such that it will both bind a slightly twisted glucan chain, consistent with the conformation adopted by cello-oligosaccharides in free solution and when bound to variety of CBMs (Notenboom *et al.*, 2001a; Boraston *et al.*, 2002; Charnock *et al.*, 2002). The length of the hydrophobic platform in CBM44 is ~24 Å, and the position of the aromatic residues would enable binding to sugars *n*, *n+2* and *n+4*, which indicates that the protein can accommodate cellopentaose. Therefore, it is interesting to note that CBM44 displays ~10-fold higher affinity for cellohexaose than cellopentaose. It is possible that residues at either end of the hydrophobic platform of CBM44 interact with cellohexaose, although mutagenesis studies show that Asp^{188}, Thr^{202} and Met^{203}, which are distal to

Trp^{189} and Trp^{198} (Figure 3.7c), do not contribute to ligand binding (see Figure 3.9b, Table 3.8, and see below). Similarly, CBM4-1 from *Cellulomonas fimi* Cel9B displays 250-fold higher affinity for cellotetraose than cellotriose (Johnson *et al.*, 1996), even though the protein only interacts with three consecutive glucose residues (Boraston *et al.*, 2002). It is possible that the more extended interchain hydrogen bonding network afforded by these longer ligands stabilizes the conformation adopted by the oligosaccharide in the binding cleft. Alternatively, if the protein does not interact with O1 of the reducing end glucose, the flexible anomeric configuration adopted by this sugar may reduce binding affinity, and thus these CBMs bind optimally to internal regions of glucan chains. In CBM30 the orientation of the aromatic residues is different than CBM44. It is likely the Trp^{68}, Trp^{72} and Trp^{78} will bind the same face of the sugars *n, n+1* and *n+3* in an oligosaccharide chain that adopts a slightly twisted conformation. Most interestingly, the module binds maximally to cellohexaose, whereas the tetrasaccharide is sufficient to occupy the complete hydrophobic platform. The possible mechanism for the tighter binding of ligands that extend beyond the hydrophobic platform may be similar to *Cf*CBM4-1, described above.

A mutagenesis strategy was employed to probe the role in ligand recognition of residues in CBM44 that are solvent-exposed and in close proximity to the hydrophobic platform, which plays a key role in ligand binding, as shown above (Figure 3.7c). In general, exchanging these amino acids for alanine causes only a very modest reduction in affinity, as judged by AGE (Figure 3.9b) and ITC (Table 3.8). Furthermore, AGE showed that the influence of the nine mutants was similar for the five ligands tested, xyloglucan, laminarin, glucomannan, HEC and barley β-glucan (data not shown), indicating that these residues do not interact with specific ligands. It is particularly surprising that the affinities of the mutant proteins $Q179A_{44}$, $S196A_{44}$, $M183A_{44}$ and $Q227A_{44}$ are not severely affected because the mutated amino acids are likely to interact with sugars *n+1* and *n+3*, which are positioned between the three aromatic residues. Only the loss of Gln^{231} causes a significant decrease in affinity (~7-fold), a residue that is likely to hydrogen-bond to the sugar that stacks against the critical central aromatic amino acid Trp^{194}. Although polar residues do not play an important role in ligand recognition in Type A CBMs (McLean *et al.*, 2000) that bind to crystalline ligands, they do contribute to the binding of Type B CBMs to their target carbohydrates (Xie *et al.*, 2001a; Pell *et al.*, 2003; Henshaw *et al.*, 2004). As such CBM44 is an atypical Type B CBM. Therefore, it is likely that the protein does not make many direct hydrogen bonds

with its ligands and that glucan recognition is mediated primarily through the three aromatic residues in the binding cleft. This may explain why the side-chains in xyloglucan do not adversely affect the binding of the polysaccharide to CBM44, as there would be no steric clashes between the xylose decorations and amino acids that make direct hydrogen bonds with the glucan backbone.

Table 3.8. Affinity of CBM44 mutants for xyloglucan as determined by ITC.

Protein	$K_a \times 10^4$	ΔG	ΔH	$T\Delta S$
	M^{-1}	$kcal\ mol^{-1}$	$kcal\ mol^{-1}$	$kcal\ mol^{-1}$
Wild-type	81.6 ± 9.8	-8.1 ± 0.1	-16.3 ± 0.6	-8.2 ± 0.7
$Q179A_{44}$	21.2 ± 0.4	-7.3 ± 0.0	-11.8 ± 0.5	-4.5 ± 0.6
$M183A_{44}$	22.7 ± 1.7	-7.3 ± 0.0	-10.4 ± 0.1	-3.1 ± 0.1
$D188A_{44}$	106.0 ± 8.9	-8.2 ± 0.0	-16.7 ± 0.6	-8.5 ± 0.7
$W189A_{44}$	3.9 ± 0.6	-6.2 ± 0.1	-7.4 ± 0.4	-1.2 ± 0.6
$V192A_{44}$	71.6 ± 7.6	-8.0 ± 0.0	-13.9 ± 0.5	-5.9 ± 0.5
$W194A_{44}$	NB[a]			
$S196A_{44}$	49.0 ± 5.8	-7.8 ± 0.1	-11.2 ± 0.3	-3.4 ± 0.4
$W198A_{44}$	2.2 ± 0.2	-5.9 ± 0.0	-6.8 ± 0.5	-0.9 ± 0.6
$T202A_{44}$	77.0 ± 7.3	-8.0 ± 0.0	-14.1 ± 0.2	-6.1 ± 0.3
$M203A_{44}$	62.6 ± 6.4	-7.9 ± 0.0	-14.1 ± 0.4	-6.2 ± 0.5
$Q227A_{44}$	29.3 ± 2.5	-7.4 ± 0.0	-13.3 ± 0.3	-5.9 ± 0.3
$Q231A_{44}$	12.1 ± 0.7	-6.9 ± 0.0	-11.0 ± 0.6	-4.1 ± 0.6

[a]No binding was detected.

3.2.4. Conclusions

The plant cell wall is an intricate macromolecular structure formed by a complex repertoire of recalcitrant polysaccharides in which β-1,4 glucose polymers predominate. Plasticity is a remarkable property of cellulases that enables the same enzyme to cleave β-1,4-glucosidic bonds in a range of polysaccharides (*e.g.* cellulose, xyloglucan, glucomannan and mixed linked β-1,4-β-1,3-glucans). Here it is shown that Cel44A of *Ct*Cel9D-Cel44A is a typical endoglucanase that is capable of cleaving a variety of glucan-based plant cell wall polysaccharides such as cellulose, β-glucans, xyloglucans and glucomannans, while retaining significant activity against xylan. It is shown that the role of CBM44, previously a domain of unknown function, is associated with the targeting of Cel44A to its substrates, which are embedded in a complex polysaccharide matrix. To mediate efficient targeting, ligand recognition by CBM44 needs to mirror the substrate specificity of Cel44A, and it is shown here that this promiscuity in

carbohydrate recognition is intrinsic to CBM44. The properties of CBM44 are remarkably similar to the N-terminal CBM30, although binding affinities are lower for the latter CBM. As typical of Type B CBMs from the β-sandwich superfamily, the crystal structures of CBM30 and CBM44 suggest that the shallow groove identified in the protein structure constitutes the ligand-binding site. Site-directed mutagenesis studies revealed that the aromatic residues that comprise the hydrophobic platform in the respective clefts of CBM30 and CBM44 play a pivotal role in ligand binding. It is suggested that the location of the three binding residues on the edge of the cleft may enable CBM30 and CBM44 to bind unsubstituted β-1,4 glucan polymers, in addition to highly branched hemicelluloses such as xyloglucans. The topology of the binding site and the paucity of polar residues involved in ligand recognition, at least in CBM44, indicate that the decorations evident in xyloglucans will be solvent-exposed, contributing to the plasticity in ligand specificity displayed by CBM30 and CBM44. Whether the true biological function of CBMs that recognize single β-1,4-glucan chains is to target xyloglucan rather than the amorphous regions of cellulose is currently unclear. Finally, this work illuminates the importance of *Ct*Cel9D-Cel44A, a major cellulosomal enzyme, in the function of this highly efficient multienzyme complex. In a concerted action CBM44 and Cel44A target the hydrolysis of hemicellulosic polysaccharides (β-glucans, xyloglucans, and glucomannans) that are in intimate contact with cellulosic microfibrils. Once these polysaccharides are removed, both catalytic domains, Cel9D acting as an exocellulase on crystalline cellulose and Cel44A functioning as an endoglucanase on the amorphous sections of the polysaccharide, act synergistically to degrade the most abundant and recalcitrant plant cell wall polysaccharide, cellulose.

Coordinates and observed structure factor amplitudes have been deposited in the PDB with the accession codes 2C24 (CBM30), 2C26 (native CBM44) and 2C4X (SeMet CBM44).

Acknowledgments

The ESRF is thanked for provision of data collection facilities. This work was supported by grant POCT/BIA-PRO/59118/2004 from Fundação para a Ciência e Tecnologia (Portugal) and by individual grants SFRH/BPD/20357/2004 (to SN), SFRH/BD/16731/2004 (to CG) and SFRH/BD/23784/2005 (to MC) from the same institution. SN and AC solved the PKD-CBM44 and CBM30 structures, respectively. DB executed the ITC.

CHAPTER 4 How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides. Structure and activity of a *Clostridium thermocellum* lichenase, *Ct*Lic26A

Edward J. Taylor[1*], Arun Goyal[2*§], Catarina I.P.D. Guerreiro[2], José A.M. Prates[2], Victoria Money[1], Natalie Ferry[3], Carl Morland[3], Antoni Planas[4], James A. Macdonald[5], Robert V. Stick[5], Harry J. Gilbert[3], Carlos M.G.A. Fontes[2] and Gideon J. Davies[1]

[1]York Structural Biology Laboratory, Department of Chemistry, University of York, York, YO10 5YW, UK; [2]CIISA, Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; [3]Institute for Cell and Molecular Biosciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, UK; [4]Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, 08017 Barcelona, Spain; [5]Chemistry, School of Biomedical and Chemical Sciences M313, University of Western Australia, Crawley, Western Australia 6009, Australia

Adapted from *J. Biol. Chem.* (2005) 280, 32761-32767

Abstract

One of the most intriguing features of the 112 GHs families is the range of specificities displayed by different members of the same family, whereas the catalytic apparatus and mechanism are often invariant. Family GH26 predominantly comprises β-1,4-mannanases; however, a bifunctional *Clostridium thermocellum* GH26 member (hereafter *Ct*Lic26A) displays a markedly different specificity. Here it is shown that *Ct*Lic26A is a lichenase, specific for mixed $(Glc\beta1,4Glc\beta1,4Glc\beta1,3)_n$ oligo- and polysaccharides, and displays no activity on manno-configured substrates or β-1,4 linked homopolymers of glucose or xylose. The three-dimensional structure of the native form of *Ct*Lic26A has been solved at 1.50 Å resolution revealing a characteristic $(\beta/\alpha)_8$ barrel with Glu^{109} and Glu^{222} acting as the catalytic acid/base and nucleophile in a double-displacement mechanism. The complex with the competitive inhibitor, Glc-β-1,3-isofagomine (K_i 1 μM), at 1.60 Å, sheds light on substrate recognition in the -2 and -1 subsites and illuminates why the enzyme is specific for lichenan-based substrates. Hydrolysis of β-mannosides by GH26 members is thought to proceed through transition states in the $B_{2,5}$ (boat) conformation in which structural distinction of glucosides *versus* mannosides reflects not the configuration at C2 but the recognition of the pseudo-axial O3 of the $B_{2,5}$ conformation. Here it is suggested a different conformational itinerary for the GH26 enzymes active on gluco-configured substrates.

[*]Contributed equally to this work
[§]Present address: Dept. Biotechnology, Indian Institute of Technology Guwahati, North Guwahati, Guwahati 781 039, Assam, India

4.1. INTRODUCTION

GHs play a critical role in both eukaryotes and prokaryotes, where they fulfil numerous important functions such as substrate acquisition and the remodeling of cell walls and the glycan decorations of glycoproteins. GHs are currently grouped into 112 sequence-based families (Coutinho and Henrissat, 1999; http://afmb.cnrs-mrs.fr/CAZY; April 2008). A further complexity of these GH families is that many different families, defined by sequence, display similar three-dimensional structures and catalytic apparatus leading to a "clan" classification, analogous to that used for proteases (see 1.4.1.). Thus clan GH-A unites a large number of diversely related GHs including 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79 and 86 with a parallel divergence of substrate specificities (Henrissat *et al.*, 1995). All these enzymes adopt the same $(\beta/\alpha)_8$ fold and perform catalysis with net retention of the configuration of the anomeric carbon via a double displacement mechanism. Notably, the location of the catalytic apparatus is completely conserved, with the catalytic acid/base and nucleophile glutamates positioned on strand β-4 and β-7, respectively (Henrissat *et al.*, 1995; Jenkins *et al.*, 1995). The majority of these enzymes possess a conserved NEP motif (occasionally, such as in GH26, this is replaced by HEP) in which the glutamate is the acid/base and the Asn/His interacts with the O2 of the substrate in the critical -1 subsite, where the glycoside undergoes transition state formation during glycosidic bond hydrolysis. It is, therefore, intriguing how GH-A clan members recognize substrates differing in configuration at the C2 position while maintaining similar active-centre environments. This comparison is given particular poignancy by the selective hydrolysis of mannosides, whose ground-state axial O2 renders non-enzymatic mannoside chemistry particularly challenging.

GH26 (May 2005) contains 63 open reading frames, the vast majority of which are endo-β-1,4-mannanases (mannanases), although two members of this family display β-1,3-xylanase activity (Araki *et al.*, 2000; Okazaki *et al.*, 2002). In May 2005, the only reported three-dimensional structure for a GH26 member is for the *Cellvibrio japonicus* mannanase Man26A (Hogg *et al.*, 2001). Man26A hydrolyses manno-configured substrates via a $B_{2,5}$ transition state, indicating that the enzyme exploits the pseudo-axial O3 of the transition state as an important specificity determinant and avoids steric clashes of O2 during catalysis (Ducros *et al.*, 2002). Currently, the conformational itinerary of GH26 members that display activity against xylo- and gluco-configured

substrates is unclear. To address this question we have screened family GH26 for enzymes that exhibit β-glucanase activity. Here it is reported the cloning and overexpression of a GH26 enzyme that displays novel activity for this family, the *Clostridium thermocellum* lichenase, *Ct*Lic26A. Kinetic characterization reveals the enzyme to be a classic mixed linkage β-1,4-β-1,3-glucanase that targets barley β-glucan and lichenan as its primary substrates and which exhibits no activity against β-1,4 or β-1,3 homopolymers of gluco- or manno-configured polysaccharides. The three-dimensional structure of the enzyme has also been solved in both native and inhibited forms at 1.5 and 1.6 Å, respectively. The structures shed light on the basis for substrate specificity and also hint that although GH26 enzymes active on manno-configured substrates use a $B_{2,5}$ transition state (Ducros *et al.*, 2002; Vincent *et al.*, 2004) this is unlikely to be the case for the GH26 enzymes active on gluco-configured substrates. This work adds to the emerging picture (Davies *et al.*, 2003) that the conformational itinerary of glycosidases is a subtle interplay of substrate configuration with enzyme three-dimensional structure, and thus, the conformational machinations are not conserved in different members of the same family.

4.2. EXPERIMENTAL PROCEDURES

4.2.1. Bacterial strains, plasmids and culture conditions

Escherichia coli strains XL1-Blue (Stratagene) and BL21 (DE3) (Novagen) were cultured at 37 ºC in LB unless otherwise stated. Media were supplemented with 100 mg/L ampicillin or 2 mg/L 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-gal) to select for *E. coli* transformants and recombinants, respectively. *E. coli* cells used to propagate bacteriophage were grown on LB supplemented with 10 mM $MgSO_4$ and 0.2% maltose and were plated out on NZYM top agar (0.7%). The phage library was screened for GH activity as described previously (Fontes *et al.*, 1995). The bacteriophage and plasmids employed in this work were λZAPII (Stratagene) and pET21a (Novagen).

4.2.2. Expression and purification of *Ct*Lic26A-Cel5E recombinant derivatives

DNA encoding *Ct*Lic26, *Ct*Cel5, *Ct*Lic26-Cel5, and *Ct*Lic26-Cel5E-CBM11 were amplified by PCR and cloned into *Nhe*I and *Xho*I restricted pET21a to generate pCF1, pCF2, pCF3 and pCF4, respectively (Table 4.1). The recombinant proteins encoded by

these four plasmids contain a C-terminal His_6-tag. To generate pCF1, QuickChange site-directed mutagenesis (primers described in Table 4.1) of pCF1 was used to introduce the authentic stop codon into *Lic26A*, and thus, the encoded protein (*Ct*Lic26As) does not contain a His_6-tag. QuickChange site-directed mutagenesis was also employed to generate E109A and E222A mutants of *Ct*Lic26A (primers in Table 4.1).

Table 4.1. Primers used for the cloning and mutagenesis of *Ct*Lic26A-Cel5E recombinant derivatives

Construct	Primers	Sequence (5'-3')
pCF1	Forward	CTC**GCTAGC**AATTACAACAGTGGT[a]
	Reverse	CAC**CTCGAG**GGCTCCTATTGCCTCCC
pCF2	Forward	CTC**GCTAGC**AAGGCTGTCGACCCCTT
	Reverse	CAC**CTCGAG**ATTAAATAGTGCATTGAGG
pCF3	Forward	CTC**GCTAGC**AATTACAACAGTGGT
	Reverse	CAC**CTCGAG**ATTAAATAGTGCATTGAGG
pCF4	Forward	CTC**GCTAGC**AATTACAACAGTGGT
	Reverse	CAC**CTCGAG**AGCACCAATCAGCTTGAT
E109A	Forward	TTAAGACCTCTTCATGCAGCCAACGGAGACTGG
	Reverse	CCAGTCTCCGTTGGCTGCATGAAGAGGTCTTAA
E222A	Forward	CCCATCATTATAGCAGCGTTTGCATCAGCTGAA
	Reverse	TTCAGCTGATGCAAACGCTGCTATAATGATGGG

[a]The *Nhe*I and *Xho*I restriction sites introduced in the forward and reverse primers, respectively, are shown in bold.

To express the clostridial proteins, *E. coli* BL21 (DE3) (Novagen) cells harbouring the appropriate recombinant plasmid were cultured in LB containing ampicillin at 37 ºC to mid-exponential phase ($OD_{550\ nm}$=0.6), at which point IPTG was added to a final concentration of 1 mM and the cultures were incubated for a further 5 h. The His_6-tagged recombinant proteins were purified from cell-free extracts by IMAC as described previously (Dias *et al.*, 2004). For crystallization, *Ct*Lic26A was further purified by size exclusion chromatography following the method of Dias *et al.* (2004). *Ct*Lic26As was purified from osmotic shock fractions by anion exchange chromatography (Bolam *et al.*, 1996) followed by size exclusion chromatography (Dias *et al.*, 2004). Preparation of SeMet *Ct*Lic26A followed standard procedures (Dias *et al.*, 2004), and the protein was purified using the same procedure employed for the native enzyme except that 10 mM mercaptoethanol was included in all buffers and the purified protein was exchanged into water containing 5 mM DTT.

4.2.3. Enzyme assays

The activity of *Ct*Lic26A-Cel5E-truncated derivatives against various polysaccharides was determined as described previously (Dias *et al.*, 2004) by detecting the release of reducing sugars using the Nelson (Nelson, 1944) and Somogyi (Somogyi, 1945) reagent. Standard assays for *Ct*Lic26A were carried out in 50 mM sodium phosphate buffer, pH 7.0, containing 1 mg/mL BSA at 37 ºC, whereas the kinetic parameters of Cel5E were determined in 50 mM sodium acetate buffer, pH 4.5, containing 1 mg/mL BSA also at 37 ºC. To explore the pH profile of both Cel5E and *Ct*Lic26A 50 mM sodium acetate, pH 4-6, sodium phosphate/citrate, pH 6-7.5, Tris-HCl, pH 7.5-8.5, and CAPS, pH 8.5-10, buffers were used in enzyme assays that employed 0.2% lichenan as the substrate. Kinetic parameters were determined using six substrate concentrations that straddled the K_m value. 4-Methylumbelliferyl (MU) glucosides were prepared and utilized as lichenase substrates as described previously (Malet and Planas, 1997). In the case of *Ct*Lic26A, to determine activity against MU glucosides, product formation was monitored at 365 nm and the concentration of 4-methylumbelliferone was calculated using a molar extinction coefficient at 365 nm and pH 7.0 of 3520 M^{-1} cm^{-1}. The K_i value for Glc-β1,3-isofagomine (Glc-isoF) was determined by measuring the apparent K_m for MU Glc-β1,4-Glc-β1,3-Glc-MU using inhibitor concentrations ranging from 200 nM to 6 μM.

4.2.4. Crystallization and data statistics

Pure proteins as judged by SDS-PAGE were concentrated to 10-20 mg/mL, and buffer was exchanged into water (Sigma) using a Vivaspin 10 kDa cut-off concentrator. *Ct*Lic26A was screened using the hanging drop vapour diffusion method together with the Hampton Crystal screen, Crystal screen 2 and Hampton PEG/Ion screen (Hampton research, Alison Viejo, CA). Drops containing 1 μL of protein were mixed together with 1 or 2 μL of the mother liquor. Initial crystals were found to grow in the CSS Crystal screen I, condition number 12. This condition was optimized further to improve crystal quality and the cryogenic properties, resulting in conditions of 0.2-1.2 M sodium formate, 0.1 M sodium cacodylate buffered at pH 6.5 and 5-20% PEG 4000. A cryo-protectant solution was produced by supplementing the mother liquor with an additional 20% glycerol. Crystals were harvested in rayon fibre loops then bathed in cryoprotectant solution before flash-freezing in liquid N_2. SeMet crystals were produced in cacodylate buffer and back-soaked into MES buffer at the same pH to remove the

arsenic from the crystals (and hence prevent the masking of the anomalous signal; arsenic and selenium have similar absorption spectra at energies at the Se K edge 11.8667 and 12.6578 keV, respectively). Data were collected at the ESRF from single crystals at -153 ºC with a Δφ of 0.5º. Non-derivative native data were collected at a wavelength of 0.9340 Å over an oscillation range of 135º on ID14-1 using an ADSC Q4R CCD charged-coupled device detector. Selenium derivative data were collected at a wavelength of 0.9791 Å over an oscillation range of 400º on ID29 using an ADSC Q210 2D charged-coupled device detector.

4.2.5. Structure solution and refinement

All data were processed and reduced with the HKL suite (Otwinowski and Minor, 1997) or MOSFLM (Leslie, 1992). All other computing was undertaken using the CCP4 suite (1994) unless otherwise stated. Native *Ct*Lic26A crystals were found to belong to the space group $P2_12_12_1$ with the approximate cell dimensions of *a*=49.8 Å, *b*=63.6 Å, *c*=100.0 Å. The structure solution of *Ct*Lic26A was solved using SAD. Heavy atom positions were determined using SHELXD, initial phases calculated using MLPHLARE, and ophase improvement performed with DM. 5% of the total reflections were flagged for cross-validation before refinement, and the behavior of these data was used to monitor the model at various stages of refinement for the weighting of geometrical and temperature factor restraints. REFMAC (Murshudov *et al.*, 1997), in conjunction with ARP/wARP (Perrakis *et al.*, 1999), were used to build the sequence into the electron density automatically. QUANTA and X-FIT (Accelrys, San Diego, CA, USA) were used to make manual corrections to the model. Solvent molecules were added using X-SOLVATE and checked manually. The structure was validated using PROCHECK (Laskowski *et al.*, 1993) before deposition.

4.2.6. Structure determination of a *Ct*Lic26A complex with Glc-isoF

*Ct*Lic26As (wild type without His_6-tag) was concentrated to approximately 22 mg/mL in a Vivaspin 10 kDa cut-off concentrator and washed into water. The protein was screened using the sitting drop vapour diffusion method together with the Hampton Crystal screen, Crystal screen 2 and Hampton PEG/Ion (Hampton research, Alison Viejo, CA), and the CSS Crystal screen I. 100-nL droplets were formed by the mosquito® (TTP LabTech Ltd, Royston, Herts, UK) liquid-handling robot, mixing 50 nL protein solution with 50 nL of well solution. Initial crystals were formed in Hampton

crystal screen 2, condition 26. Further optimisation of this condition in the hanging drop vapour diffusion method using 1 μL of protein mixed with 1 μL of well solution combined with serial streak seeding of the resultant crystalline precipitate led to the formation of diffraction quality, prismatic crystals in a mother liquor of 0.15 M ammonium sulphate and 30% PEG 5000 monomethylether buffered to pH 6.5 with 4-morpholine-ethanesulfonic acid (MES). The crystals were harvested in rayon fibre loops with the mother liquor providing cryoprotection before being flash-frozen in liquid N_2. Data were collected on station ID29 at the ESRF from single crystals at -153 ºC at a wavelength of 0.9792 Å; a high resolution sweep was collected with Δφ of 0.5º over an oscillation range of 134.5º. Low resolution data were collected with Δφ of 1.5º over an oscillation range of 135º.

The data were indexed and integrated with MOSFLM. The protein was found to crystallize in space group $P2_12_12_1$ with cell dimensions of *a*=49.27 Å, *b*=63.01 Å and *c*=78.18 Å. Data reduction and merging were performed in SCALA. Starting phases were obtained by molecular replacement using AMoRe (Navaza and Saludjian, 1997) and the previous native *Ct*Lic26A structure as a starting model. 5% of the data were set aside for determination of R_{free} to monitor the model during refinement. Refinement was undertaken in REFMAC, and solvent molecules were added using ARP/wARP and checked manually.

4.3. RESULTS AND DISCUSSION

4.3.1. Cloning and expression of *Ct*Lic26A-Cel5E

In a previous study (Yague *et al.*, 1990), it was shown that a *C. thermocellum* cellulosomal enzyme, EGH (now designated *Ct*Lic26A-Cel5E to reflect current nomenclature for carbohydrate-active enzymes by Henrissat *et al.*, 1998), contained a GH5 and a GH26 catalytic module and a C-terminal type I dockerin, the latter domain targeting the enzyme to the clostridial cellulosome. More recently a family 11 carbohydrate binding module (CBM11) in *Ct*Lic26A-Cel5E that is located between the GH5 domain and the dockerin, which binds to *β*-1,4- and mixed linked *β*-1,3-*β*-1,4-glucans, was also identified (Carvalho *et al.*, 2004), Figure 4.1.

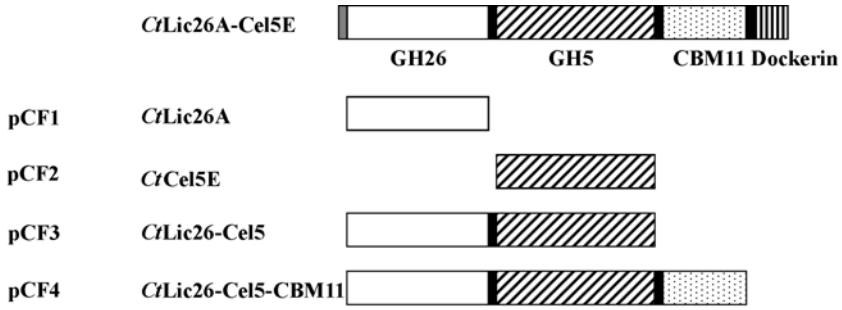

Figure 4.1. Schematic of the molecular architecture of *Ct*Lic26A-Cel5E.
The modules encoded by the defined recombinant plasmids are indicated with the grey and black boxes representing the signal peptide and linker sequences, respectively.

The GH5 module (Cel5E) was in preliminary data shown to display endoglucanase activity. To characterize the properties of Cel5E in more detail and to determine the substrate specificity of *Ct*Lic26A, the gene, designated *celH*, encoding *Ct*Lic26A-Cel5E was isolated from a genomic library of *C. thermocellum* constructed in λZAPII by screening for clones that displayed GH activity and sequencing the clostridial insert. The clone encoding *Ct*Lic26A-Cel5E also contained a second putative GH gene, *cel5M*, which is located 269 bp upstream to and transcribed in the same direction as *celH*. Cel5M, encoded by *cel5M,* is 561 amino acids. Although the protein lacks an obvious signal peptide and a dockerin sequence and is, thus, not a component of the *C. thermocellum* cellulosome, it contains a transmembrane domain that extends from Leu^{96} to Phe^{114}. The N-terminal 113 residues of Cel5M exhibit no homology to sequences in the SWISS-PROT data base, whereas the C-terminal approximately 450 residues display sequence similarity with GH5 endoglucanases, exhibiting 31% identity with a *Cryptococcus neoformans* protein (accession number AAW44253). Interestingly, the catalytic acid-base glutamate is replaced by glutamine (Gln^{318}), although the catalytic nucleophile (Glu^{443}) is retained, and thus, we predict that Cel5M is unable to hydrolyse β-glucans. It is possible that the GH5 protein displays activity against S-linked glucosides (in GH1 myrosinases the catalytic acid/base glutamate is similarly replaced by glutamine (Burmeister *et al.*, 1997)) or perhaps its true role in the bacterium is non-catalytic as is the case with many GH family members.

4.3.2. Biochemical properties of *Ct*Lic26A-Cel5E catalytic derivatives

Regions of *celH* encoding the various modules of *Ct*Lic26A-Cel5E, Figure 4.1, were amplified by PCR and cloned into *E. coli* expression vectors. The encoded

recombinant proteins, which were all expressed at high levels, were purified by IMAC to electrophoretic homogeneity. Interrogation of the biochemical properties of the *Ct*Lic26A-Cel5E derivatives shows that Cel5E displays typical endo-β-1,4-glucanase activity, hydrolysing cellulosic substrates such as CMC, HEC, and β-glucans, while exhibiting very low levels of xylanase activity and no activity against mannans and the β-1,3-glucan, laminarin. Typical of endoglucanases, Cel5E was considerably more active against lichenan, a highly accessible β-glucan, than the crystalline (and, thus, inaccessible) substrate Avicel, Table 4.2.

Table 4.2. Kinetic parameters of *Ct*Lic26A-Cel5E and its derivatives.

Enzymes	Substrate	k_{cat}	K_m	k_{cat}/K_m
		min^{-1}		
*Ct*Lic26A	Lichenan	776	0.020 mg mL^{-1}	3.9×10^{4} min^{-1} mg^{-1} mL
Cel5E	Lichenan	496	0.032 mg mL^{-1}	1.6×10^{4} min^{-1} mg^{-1} mL
*Ct*Lic26-Cel5	Lichenan	700	0.011 mg mL^{-1}	6.4×10^{4} min^{-1} mg^{-1} mL
*Ct*Lic26-Cel5-CBM11	Lichenan	679	0.009 mg mL^{-1}	7.5×10^{4} min^{-1} mg^{-1} mL
E109A *Ct*Lic26A	Lichenan	NA[a]	NA	NA
E222A *Ct*Lic26A	Lichenan	NA	NA	NA
*Ct*Lic26A	Avicel	NA	NA	NA
Cel5E	Avicel	1.38	0.190 mg mL^{-1}	7.3 min^{-1} mg^{-1} mL
*Ct*Lic26-Cel5	Avicel	1.64	0.187 mg mL^{-1}	8.7 min^{-1} mg^{-1} mL
*Ct*Lic26-Cel5-CBM11	Avicel	20.5	0.405 mg mL^{-1}	5.1×10^{1} min^{-1} mg^{-1} mL
*Ct*Lic26A	Glcβ3GlcβMU	2.8	5.3 mM	5.0×10^{-1} min^{-1} mM^{-1}
*Ct*Lic26A	Glcβ4Glcβ3GlcβMU	32	87 µM	3.7×10^{2} min^{-1} mM^{-1}
*Ct*Lic26A	Glcβ4Glcβ4Glcβ3GlcβMU	67	42 µM	1.6×10^{3} min^{-1} mM^{-1}
E109A *Ct*Lic26A	Glcβ4Glcβ3GlcβMU	0.7	3.2 µM	2.2×10^{2} min^{-1} mM^{-1}
E222A *Ct*Lic26A	Glcβ4Glcβ3GlcβMU	NA	NA	NA

[a]NA, no activity detected. The assay can detect activity that is $>10^{-5}$ of the wild type enzyme.

CBM11 potentiated the activity of Cel5E against the insoluble substrate Avicel but not against the soluble polysaccharide lichenan, a feature that is common to many cellulose-binding CBMs (Tomme *et al.*, 1988; Hall *et al.*, 1995; Boraston *et al.*, 2003; Boraston *et al.*, 2004). *Ct*Lic26A, in contrast to Cel5E, does not exhibit any activity against galactomannans, glucomannans, laminarin, Avicel or soluble cellulosic derivatives, but appears to be specific for β-1,3-1,4-glucans, such as lichenan and barley β-glucan, Table 4.2. Although CBM11 interacts with lichenan it does not enhance the activity of *Ct*Lic26A against this substrate, and similarly, there is no obvious synergy between the two catalytic modules. Although the temperature optimum of both catalytic modules in *Ct*Lic26A-Cel5E is approximately 60 ºC, *Ct*Lic26A and Cel5E display

maximal activity at pH 7.0 and 4.5, respectively. It is unusual for two catalytic modules in the same enzyme to display significantly different pH optima and the biological significance of this phenomenon is unclear. The relative activity of *Ct*Lic26A against aryl *β*-gluco-oligosaccharides is as follows: Glc-*β*-1,3-Glc-*β*-MU << Glc-*β*-1,4-Glc-*β*-1,3-Glc-*β*-MU < Glc-*β*-1,4-Glc-*β*-1,4-Glc-*β*-1,3-Glc-*β*-MU, Table 4.2. Because the lichenase displays no activity against laminaritriose (Glc-*β*-1,3-Glc-*β*-1,3-Glc) the natural bond hydrolysed by the enzyme is Glc-*β*-1,4-Glc. Thus, the glycone region of the substrate binding cleft of *Ct*Lic26A accommodates a substrate that comprises (-4 to -1) Glc-*β*-1,4-Glc-*β*-1,4-Glc-*β*-1,3-Glc, Figure 4.2.

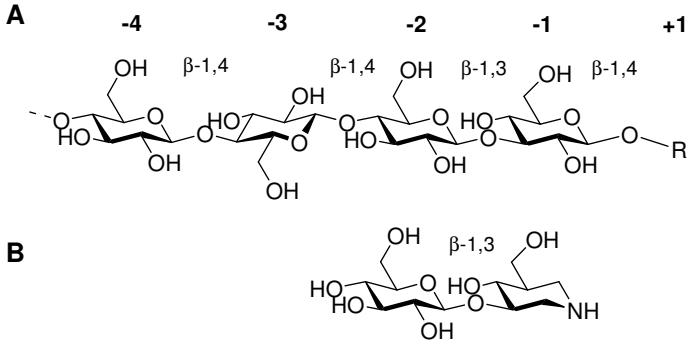

Figure 4.2. Specificity of *C. thermocellum Ct*Lic26A.
(A) Schematic diagram indicating the subsite specificity and (B) the disaccharide inhibitor laminaribiose-derived isofagomine (K_i 1.2 μM at pH 7.0) which is observed binding to the -2 and -1 subsites.

Such a subsite specificity reflects the structures of lichenan and barley *β*-glucan whose dominant repeating unit (along with Glc-*β*-1,4-Glc-*β*-1,3) is Glc-*β*-1,4-Glc-*β*-1,4-Glc-*β*-1,3 (Brett and Waldren, 1996). The relative increase in k_{cat}/K_m values, from the MU disaccharide to the tetrasaccharide, suggests that the -3 subsite contributes significantly to catalysis, as was observed previously with the structurally unrelated lichenases from family GH16 (Malet and Planas, 1997). Furthermore, the higher k_{cat} of *Ct*Lic26A against lichenan compared with an oligosaccharide that occupies subsites -4 to +1 indicates that the enzyme contains additional aglycone subsites, which only the polysaccharide can harness.

4.3.3. Three-dimensional structure of *Ct*Lic26A

The structure of *Ct*Lic26A was solved using single-wavelength anomalous dispersion techniques, optimised for the f'' signal of Se from a selenomethionyl-derived form of His-tagged *Ct*Lic26A. Subsequently, an isomorphous native form of *Ct*Lic26A was refined at 1.5 Å resolution. Details of data and structure quality are shown in Table 4.3.

Table 4.3. X-ray data and structure quality for *C. thermocellum Ct*Lic26A.
Outer resolution shell statistics are given in parentheses.

	***Ct*Lic26A SeMet**	***Ct*Lic26A native**	***Ct*Lic26A Glc-isoF**
Beam line	ID29	ID14.1	ID29
Wavelength (Å)	0.9791	0.9340	0.9792
Resolution of data (Å)	25-1.45 (1.5-1.45)	36-1.50 (1.5-1.50)	40-1.60 (1.64-1.60)
R_{merge}	0.054 (0.11)	0.048 (0.103)	0.060 (0.087)
Mean I/σI	27 (14)	37 (12)	26 (17)
Completeness %	96 (80)	96 (94)	99 (98)
Multiplicity	5.4 (5.0)	5.4 (5.3)	6.5 (5.3)
R_{cryst}		0.131	0.15
R_{free}		0.154	0.18
RMSD 1-2 bonds (Å)		0.014	0.008
RMSD 1-3 bonds (°)		1.446	1.149
Average main chain B (Å^2)		9	4
Average side chain B (Å^2)		15	6
Average substrate B (Å^2)		NA[a]	3
Average solvent B (Å^2)		26	19

[a]NA, no activity detected.

The structure of the native form of the enzyme could be traced with no breaks from Ser^7 to His^{290} with the final five traceable residues from the metal affinity tag. *Ct*Lic26A forms a $(\beta/\alpha)_8$ barrel structure, with the catalytic acid/base (Glu^{109}) and nucleophile (Glu^{222}) on strands β-4 and β-7, respectively (Figure 4.3a), as is typical for enzymes classified into clan GH-A (Henrissat *et al.*, 1995). The observation that E222A was inactive (<10^{-5} of wild type *Ct*Lic26A activity) whereas E109A displayed no activity against lichenan and a significantly reduced k_{cat} and K_m (compared to wild type *Ct*Lic26A) against Glc-β-1,4-Glc-β-1,3-Glc-β-MU, Table 4.2, is entirely consistent with Glu^{222} and Glu^{109} comprising the catalytic nucleophile and acid-base residues, respectively. Another notable feature of the catalytic apparatus of *Ct*Lic26A is that it bares a GH26 "fingerprint" in that the residue preceding the acid/base, which interacts with the 2-position of the substrate is histidine, unlike other non-family 26 members of Clan GH-A where the residue preceding the acid is asparagine. A DALI (Holm and

Sander, 1993) search reveals that *Ct*Lic26A is similar to other clan GH-A GHs with the closest similarity, not surprisingly, with the family GH26 mannanase, Man26A, with which *Ct*Lic26A shares approximately 17% sequence identity (DALI Z score 21.6 with a RMSD of 2.7 Å over 239 common Cα positions).

4.3.4. Structure of a *Ct*Lic26A-inhibitor complex: molecular basis for substrate specificity in family GH26

The structure of the C-terminal His-tagged *Ct*Lic26A revealed that the affinity tag of one molecule occluded the catalytic and substrate binding centre of a symmetry-equivalent molecule. This crystal form could not be used for crystal soaking experiments. Subsequently, "native" *Ct*Lic26A (*Ct*Lic26As), lacking the C-terminal tag, through the re-introduction of authentic stop codon, was prepared for ligand-binding studies. Crystals were grown in the presence of the disaccharide competitive inhibitor Glc-isoF (Macdonald *et al.*, 2002; Macdonald *et al.*, 2004), Figure 4.2. The disaccharide competively inhibits the lichenase at pH 7.0 with a K_i of 1.2 μM, which is >1000-fold lower than the K_m for the corresponding substrate Glc-β-1,3-Glc-β-MU. Isofagomine analogues generally bind much more tightly to GHs than the corresponding substrate, reflecting a positively charged "anomeric" nitrogen and a strongly favourable entropy of binding (Varrot *et al.*, 2003). The structure of the inhibited enzyme could be traced with no breaks from Ser^{7} to Leu^{283}. Comparison of the complex structure with that of the native enzyme reveals the only significant main-chain difference occurs in the loop between Glu^{258} and Trp^{264}, which is raised by approximately 3.5 Å at Glu^{261} to allow binding to the affinity tag in the C-terminal tagged protein used for the native, but not complex structure determination. Indeed, a subsequent structure of non-complexed, but His-tag-less protein (not shown), the mobile loop is indeed found to coincide with that of the inhibited enzyme.

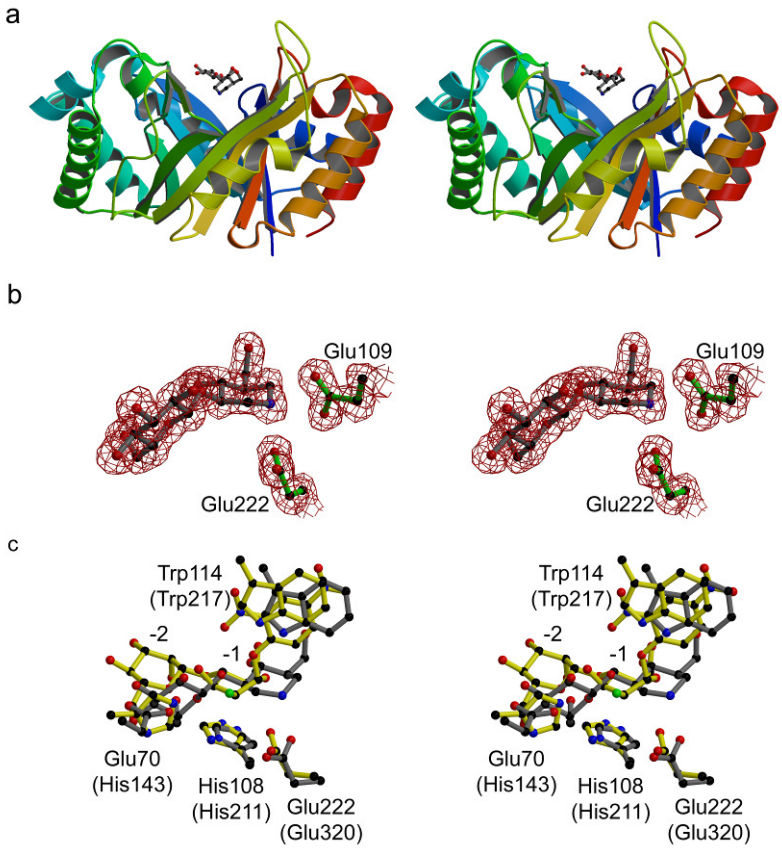

Figure 4.3. Three-dimensional structure of *C. thermocellum* Lic26A and its complex with the laminaribio-derived isofagomine.
(a) Protein schematic, color-ramped from N-terminus (blue) to C-terminus (red) with the Glc-isoF ligand in ball-and-stick. (b) Observed 2 F_{obs}-F_{calc} electron density (contoured at 1σ) for Glc-isoF and the catalytic acid/base and nucleophile, Glu^{109} and Glu^{222}, respectively. (c) The overlap of the active centres of the Glc-isoF complex of *Ct*Lic26 (this study, grey) and the Michaelis complex of unhydrolysed DNP 2Fmannobioside complex of *Cj*Man26 (Ducros *et al.*, 2002, yellow). *Ct*Lic26 residues discussed in the text are labelled with the equivalent residue of *Cj*Man26 in parentheses. This figure, in divergent (wall-eyed) stereo was drawn with MOLSCRIPT (Kraulis, 1991) and BOBSCRIPT (Esnouf, 1997).

Glc-isoF binds, as expected, in the -2 and -1 subsites. Extremely low level residual electron density in the +1 subsite suggesting very weak binding of a second Glc-isoF molecule, but this could not be modeled appropriately. In the -2 and -1 subsites, however, the electron density at 1.6 Å resolution is unambiguous (Figure 4.3b), and thus, the complex sheds light on the crucial interactions that define substrate specificity in GH26 (Figure 4.4). In both subsites the rings lie in $^4C_1$ (chair) conformations. Consistent with other work (Notenboom *et al.*, 2000; Mark *et al.*, 2001; Varrot *et al.*, 2003; Zechel *et al.*, 2003), the endocyclic nitrogen of the isofagomine moiety makes favourable interactions with both the nucleophile, Glu^{222}, and the acid-base Glu^{109}. Other interactions of the -1 subsite include recognition of both O4 and O6

by Glu[261], with Tyr[185] interacting with the nucleophile (and most-likely O5). Phe[256] provides an aromatic platform which, although it does not "stack" with a ring in $^4C_1$ conformation, may well favour binding of a distorted sugar at the transition state (see below). Isofagomine lacks a hydroxyl at C2, but simple modeling studies suggest that both His[108] and Tyr[115] would interact with the O2 of a glucoside in this subsite, although the relative contribution of these two side-chains will very much reflect the type and extent of substrate distortion during catalysis.

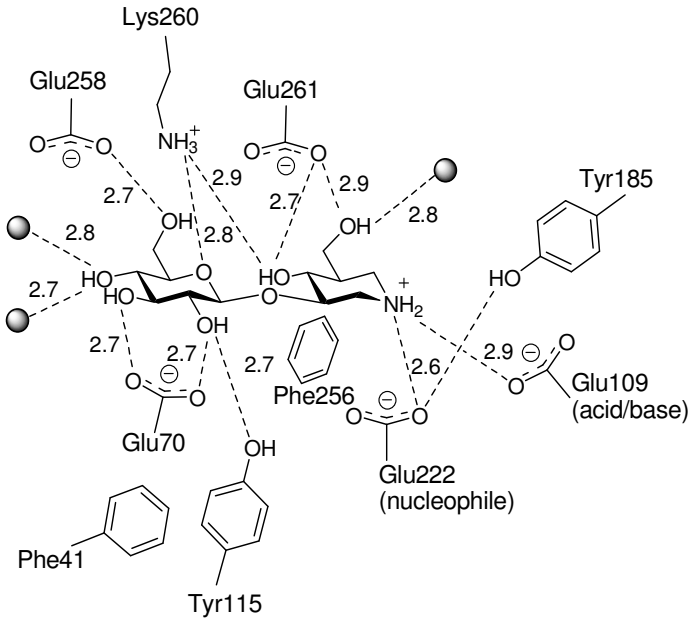

Figure 4.4. Schematic diagrams of the interactions of the *C. thermocellum Ct*Lic26A with Glc-isoF.

In the -2 subsite there is "tight" recognition of the sugar hydroxyls with Glu[258] and Lys[260] interacting with O6, Glu[70] with O3 and O2 and Tyr[115] with O2. Residues 258-261 are also positioned to prevent binding of a β-1,4 linked sugar in the -2 subsite. Phe[41] provides an aromatic platform making van der Waals contacts with the α face of the glucoside. This latter interaction presumably also contributes significantly to the recognition of laminaribiosyl moieties in subsites -2 and -1 as the relative twist of the two sugars is very different in 1,3- compared to 1,4-linked disaccharides. The -2 to -3 subsite boundary most likely selects for a β-1,4 linkage as no binding of the Glc-Glc-isoF trisaccharide was observed in crystal. The structure of *Ct*Lic26A becomes more open in this region, but the steric clashes provided by Glu[70], Trp[72] and the loop from

Asn^{42} to Trp^{44} would most likely all disfavour binding of a β-1,3 linked glucoside in subsite -3. In contrast, simple docking of a β-1,4 linked glucoside suggests that it would make no steric clashes.

4.3.5. Catalysis and substrate distortion

Although GH26 is dominated by endo-β-1,4 mannanases, two other activities have recently been described, lichenase (this work) and β-1,3-xylanase (Araki *et al.*, 2000; Okazaki *et al.*, 2002). The comparison of *Ct*Lic26A with β-mannanases from this family is, therefore, particularly important from a mechanistic standpoint. The enzymatic hydrolysis of glycosides features oxocarbenium ion-like transition states in which the anomeric centre becomes sp^2-hybridized, and partial positive charge accumulates primarily across the endocyclic O5-C1 bond (Davies *et al.*, 1997a; Vasella *et al.*, 2002). For pyranosides such a species demands planarity of C5, O5, C1 and C2 at or near the transition state, a situation found only in $^4H_3$ and $^3H_4$ (half-chair) conformations (or their closely related envelope forms) and $^{2,5}B$ and $B_{2,5}$ (boat conformations) (Davies *et al.*, 1997a). In the context of these conformational criteria, β-mannanases, exemplified by Man26A, are a chemically fascinating group of enzymes. The Michaelis complex of unhydrolysed substrate was trapped on Man26A (Ducros *et al.*, 2002). It revealed an unusual $^1S_5$ conformation which together with the $^OS_2$ conformation for the covalent intermediate, strongly suggested a novel conformational itinerary for these enzymes through a $B_{2,5}$ transition-state. Such a transition state would have the crucial facet that it would alleviate 1,2-*cis*-di-axial repulsions at the transition state by placing a manno-configured O2 pseudo-equatorial while still placing the glycosidic oxygen axial, allowing in-line nucleophilic attack. Support for these proposals comes both from the demonstration that retaining α-mannosidases use the reverse ($^OS_2$-$B_{2,5}$) conformational pathway (Numao *et al.*, 2003) and by otherwise counterintuitive observations that inhibitors with in-plane substituents at O2 are potent mannosidase inhibitors (Lillelund *et al.*, 2003; Vincent *et al.*, 2004). It has previously been proposed that a crucial structural feature that helps discriminate between these two conformational pathways is the recognition of O3, an atom whose position changes markedly between $^4H_3$ and $B_{2,5}$ transition-states (Ducros *et al.*, 2002; Vincent *et al.*, 2004). It may be significant that the two enzyme activities specific for xylo- and gluco-configured substrates in family GH26 are both β-1,3-glycosidases in which O3 is

tethered to the adjacent sugar which, in tandem with active-site topography, may legislate against the sugar attaining a $B_{2,5}$ transition-state.

Previous comparison of family GH1 and GH5 enzymes specific for glucosides, with closely related enzymes active on mannose-derived substrates, suggested that changes in loop conformations around the -1 subsite resulted in a different position for the O3-interacting side-chains and contributed to the gluco *versus* manno specificity. In GH26, such loop conformational changes are not obvious, but such a structural difference may not be as necessary because the 3-position in a β-1,3-linked polymer is necessarily tethered. Such a hypothesis is certainly consistent with the overlap of the catalytic centres of Man26A with *Ct*Lic26A (Figure 4.3c). In the -1 subsite, interactions around O2, C1, and O5 are invariant; in particular, the positions of catalytic acid (Glu^{109}) and nucleophile (Glu^{222}), the O2-interacting His^{108} and Tyr^{185} are all near identical between the manno- and gluco-specific enzymes. In the +1 subsite, Trp^{114} of *Ct*Lic26A lies in the same position as Trp^{217} of Man26A, the latter known to be involved in "stacking" against the leaving group moiety. What are indeed demonstrably different are the interactions around O3. In Man26A, O3 is coordinated by His^{143}. In *Ct*Lic26A O3 is involved in the glycosidic linkage to the -2 subsite glucoside and, thus, makes no hydrogen bonds (<3.1 Å) to protein. Instead, the *Ct*Lic26A -2 sugar lies both displaced 4-6 Å and rotated almost exactly 90º relative to the -2 sugar of Man26A. Glu^{70} (the residue most structurally analogous to His^{153} of Man26A) is then able to coordinate the O2 and O3 of the -2 subsite sugar.

4.4. SUMMARY

The biochemical properties of *Ct*Lic26A, thus, indicate that the enzyme is a typical lichenase. The members of GH26 predominantly display endo-β-1,4-mannanase activity, whereas two enzymes in this family hydrolyse β-1,3-xylan. *Ct*Lic26A is, thus, the first example of a GH26 enzyme that displays lichenase activity. Inspection of the SwissProt data base reveals a GH26 *Clostridium acetobutylicum* protein (accession number Q97G16) that displays 52% sequence identity with *Ct*Lic26A. Significantly, Glu^{70}, Phe^{41}, Tyr^{115}, Glu^{258} and Lys^{260}, which all play key roles in recognition of the β-1,3-linked gluco-configured sugar in the -2 subsite of Lic26A (see below), are

conserved in the *C. acetobutylicum* protein, indicating that this GH26 member is likely to display lichenase activity.

The 112 families of GHs provide an excellent system for studying enzyme specificity and particularly the conformation of sugars at and near the transition state. A defining feature of recent years has been the realization that different enzymes may use distinct transition state structures, all governed by the requirements for an oxocarbenium ion-like species. It is clear that transition-state structure is dictated both by the chemical structure of the glycon moiety itself and the three-dimensional environment of the active centre; neither facet is dominant. For example, xyloside hydrolysis in family GH11 is believed to go via $^{2,5}B$ transition states (Sabini *et al.*, 1999), whereas in the structurally dissimilar family GH10, a $^{4}H_{3}$ transition state seems more likely. Glucoside hydrolysis likely also takes these two different transition states exemplified by families GH5 (Davies *et al.*, 1998) and GH6 (Varrot *et al.*, 2005), respectively. Similarly, mannoside hydrolysis most likely occurs via a $B_{2,5}$ transition state in β-mannosidase families 5 (Dias *et al.*, 2004) and 26 (Ducros *et al.*, 2002) and α-mannosidase family GH38 (Numao *et al.*, 2003) but via a $^{3}H_{4}$ conformation in family GH47 (Vallée *et al.*, 2000). Furthermore, although many of these families are structurally quite distinct, others display a high degree of sequence and structural similarity. β-Glycosidases from GH5 and GH26 are particularly fascinating since these families contain enzymes specific for both manno- and gluco-configured substrates, yet they appear to have structural invariance around the 2-position of the glycone in the -1 subsite. So, although in peripheral subsites substrate specificity simply reflects recognition of the axial *versus* equatorial nature of $^{4}C_{1}$ (chair)-conformed glycoside moieties, in the catalytic -1 subsite similar recognition apparatus around O2 favours binding of transition states in which O2 lies pseudoequatorial, $^{4}H_{3}$ for glucosides and $B_{2,5}$ for mannosides. GH5 and GH26 present different evolutionary routes to solve the conformational aspects of catalysis. In mannosidases in these two families the O3-interacting residue favours binding of a pseudoaxial O3 at the $B_{2,5}$ transition state, and this residue is indeed displaced for the glucosidases from this family. Such features need not apply to the GH26 enzymes specific for gluco-configured substrates, exemplified by *Ct*Lic26A here, since all non-mannosidases in this family reported thus far are active on β-1,3-linked substrates in which O3 is bonded to the adjacent sugar. The different yet complementary strategies adopted throughout evolution by

glycosidases not only reveal how three-dimensional structure and substrate reactivity combine to allow catalysis on different substrates but also strongly point to new and powerful strategies for enzyme inhibition though conformational mimicry with ramifications far beyond plant cell wall degradation.

Acknowledgements

The authors thank the ESRF for provision of data collection facilities and Professor Eleanor Dodson for advice with phasing strategies. The authors acknowledge funding from the Biotechnology and Biological Sciences Research Council for funding. GD is a Royal Society University Research Fellow. AG and CG were funded by Fundação para a Ciência e Tecnologia (Portugal), through individual grants SFRH/BPD/5551/2001 and SFRH/BD/16731/2004, respectively. ET solved the *Ct*Lic26A structure.

CHAPTER 5 Role of a family 11 Carbohydrate Binding Module (CBM) in the function of a recombinant cellulase used to supplement a barley-based diet for broiler chicken

Catarina I.P.D. Guerreiro[1], Teresa Ribeiro[1], Patrícia I.P. Ponte[1], Maria M.S. Lordelo[2], Luísa Falcão[2], João P.B. Freire[2], Luís M.A. Ferreira[1], José A.M. Prates[1], and Carlos M.G.A. Fontes[1]

[1]CIISA, Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; [2]Instituto Superior de Agronomia, Tapada da Ajuda, 1349-017 Lisboa, Portugal

Adapted from *Br. Poult. Sci.* 49, 446-454

Abstract

Cellulases and xylanases display a modular architecture that comprises a catalytic module linked to one or more non-catalytic Carbohydrate-Binding Modules (CBMs). CBMs have been classified, based on primary structure similarity, into 51 different families. These non-catalytic modules mediate a prolonged and intimate contact of the enzyme with the target substrate eliciting efficient hydrolysis of the target polysaccharides. A study was undertaken to investigate the importance of a family 11 CBM, displaying high affinities for barley β-glucans, in the function of recombinant derivatives of cellulase *Ct*Lic26A-Cel5E of *Clostridium thermocellum* used to supplement a barley-based diet for broiler chicken. The data show that birds fed on diets containing the recombinant *Ct*Lic26A-Cel5E modular derivatives or the commercial enzyme mixture Rovabio™ Excel AP display improved performance when compared with birds fed on diets not supplemented to exogenous enzymes ($P<0.05$). It is suggested that the enzyme dosage used in this study (30 U/kg of basal diet) was probably too high for the efficacy of the family 11 CBM to be noticed. It remains to be established if the targeting effect resulting from the incorporation of CBMs in plant cell wall hydrolases could be effective at lower exogenous enzyme dosages.

5.1. INTRODUCTION

It is widely recognized that inclusion of exogenous β-1,3-1,4-glucanases in barley-based diets and β-1,4-xylanases in wheat and rye-based diets for poultry improves the efficiency of feed utilization, enhances growth and contributes for a better use of low cost feed ingredients (Chesson, 1993; Bedford, 2000). It is usually agreed that plant cell wall hydrolases improve the nutritive value of cereal-based diets rich in soluble NSPs through a variety of mechanisms. Cellulases and hemicellulases efficiently contribute to reduce the digesta viscosity that is associated with the intake of soluble NSPs, therefore improving the rate of diffusion of substrates, digestive enzymes and nutrients (White *et al*., 1981; Fengler and Marquardt, 1988; Bedford *et al*., 1991; Bedford and Classen, 1992), while increasing the velocity of feed passage (Van der Klis *et al*., 1993). In addition, exogenous polysaccharidases may promote the proliferation of beneficial microflora in the final compartments of simple-stomach animals' gastrointestinal (GI) tract, by increasing the quantity and/or the quality of the substrates available for fermentation (Bedford and Morgan 1996; Apajalahti and Bedford, 1999). Finally, plant cell wall hydrolases may mediate their effects by releasing endosperm plant cell wall trapped nutrients that were otherwise unavailable for digestion (Hesselman and Aman, 1986). The action of one or a combination of the above-mentioned effects may depend on the type of animal, diet and exogenous enzyme used.

GHs (EC 3.2.1) that degrade plant cell wall polysaccharides display a complex molecular architecture comprising both catalytic domains and non-catalytic CBMs. The close interaction established between CBMs and plant carbohydrates allows the appended catalytic domain to intimately contact its target substrate, therefore potentiating catalysis (Boraston *et al.,* 2004). This proximity and targeting role of CBMs is of extreme importance since the complex interactions established between polysaccharides within the plant cell wall restrict their accessibility to enzyme attack (Guerreiro *et al.*, 2006; Chapter 3). In addition, some CBMs seem to play a more direct role during polysaccharide degradation by actively contributing to disrupt the carbohydrate structure (Vaaje-Kolstad *et al.*, 2005). CBMs are currently grouped into 51 sequence-based families (Coutinho and Henrissat, 2003; http://afmb.cnrs-mrs.fr/CAZY; April 2008) and were shown to display a large range of different ligand specificities. Therefore, CBMs that recognize cellulose, β-1,3-1,4-glucans, xylan, mannan, galactan, xyloglucan, arabinan and laminarin have been identified and the

molecular determinants of binding specificity described. However, there is a paucity of information concerning the importance of non-catalytic CBMs in the function of exogenous β-1,3-1,4-glucanases used to supplement cereal-based diets for simple stomached animals. Recently, a family 6 xylan-binding domain was shown to improve the efficacy of a microbial recombinant xylanase, *in vivo*, when used to supplement wheat and rye-based diets for poultry (Fontes *et al.*, 2004). Animals supplemented with a bi-modular xylanase containing catalytic and xylan-binding domains grew significantly faster than animals fed on diets containing exclusively the xylanase catalytic domain. However, the role of β-glucan and cellulose-binding domains in the function of exogenous cellulases used to supplement barley-based diets for poultry remains to be established.

The possibility of using individual recombinant enzymes for feed supplementation allows for the identification of the most efficient enzyme molecular architectures and biocatalyst combinations for the supplementation of various simple-stomach animal diets. One of the objectives for this work was to compare the capacity of two truncated derivatives of *Clostridium thermocellum* bi-functional cellulase *Ct*Lic26A-Cel5E (Chapter 4), consisting on the enzyme two catalytic modules expressed as an individual entity or fused to its endogenous family 11 CBM, to enhance the performance of poultry fed on barley-based diets. It was previously shown that the family 11 CBM from *Ct*Lic26A-Cel5E is a β-glucan-binding domain that presents significant affinities for both β-1,4- and β-1,3-1,4-mixed linked glucans (Carvalho *et al*., 2004).

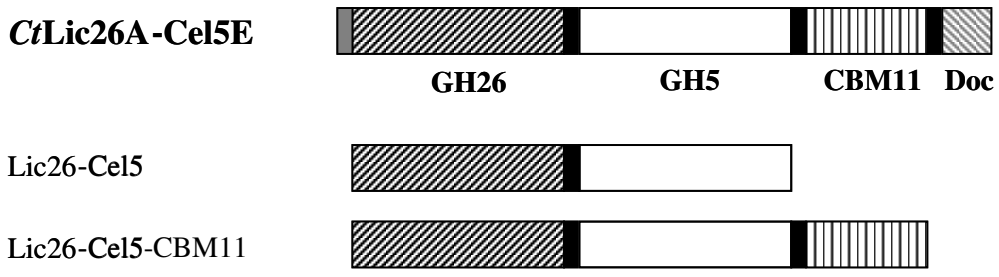

Figure 5.1. Domain organization of *Ct*Lic26A-Cel5A and its truncated derivatives Lic26-Cel5 and Lic26-Cel5-CBM11 used in this study.
The β-glucanase (GH26), cellulase (GH5), β-glucan binding domain (CBM11) and the dockerin (Doc) are indicated. The grey and the black boxes represent the linker sequences and the signal peptides, respectively.

5.2. EXPERIMENTAL PROCEDURES

5.2.1. Enzyme preparation

C. thermocellum *Ct*Lic26A-Cel5E is a thermostable bi-functional enzyme containing β-1,3-1,4-glucanase (GH26) and β-1,4-cellulase (GH5) catalytic domains and two non-catalytic modules. The molecular architecture of *Ct*Lic26A-Cel5E and its truncated recombinant derivatives used in this study are presented in Figure 5.1. The enzyme contains an N-terminal GH26, followed by a second GH5 catalytic module, a family 11 CBM (CBM11) and a C-terminal dockerin characteristic of other *C. thermocellum* cellulosomal enzymes (Taylor *et al*., 2005; Chapters 1 and 4). The *Ct*Lic26A-Cel5E truncated derivatives Lic26-Cel5E-CBM11 and Lic26-Cel5E were hyperexpressed in *Escherichia coli* following the protocols described in Chapter 4 (Taylor *et al*., 2005). The recombinant plasmids, containing the clostridial genes under the control of T7 promoters in the prokaryotic expression vector pET21a (Novagen), were used to transform BL21(DE3) *E. coli* cells (Novagen). Recombinant *E. coli* strains were grown on LB medium to mid-exponential phase (OD_{600nm}=0.5) and recombinant gene expression was induced by adding IPTG to a final concentration of 1 mM. Cells were collected after 5 h induction at 37 ºC and protein extracts prepared by ultrasonication followed by centrifugation. The recombinant proteins were purified by IMAC as described by Fontes *et al.* (2004). Both recombinant proteins, Lic26-Cel5E-CBM11 and Lic26-Cel5E, retain considerable catalytic activity at 40 ºC and are resistant to proteolytic degradation (Ponte *et al.*, 2008).

5.2.2. Animals, diets and management

The composition of the barley-based diet used in this study (Table 5.1) was formulated to contain adequate nutrient levels as defined by the NRC (1994), with an energy content slightly below the recommended values. The basal diet was supplemented with no enzyme (C0) or with 30 U/kg of Lic26-Cel5E or Lic26-Cel5E-CBM11. In addition, a fourth treatment corresponded to the supplementation of the basal diet with a calculated 15 U/kg of the commercial enzyme cocktail Rovabio™ Excel AP (Adisseo, France; treatment termed Rov for simplification), which corresponds to an incorporation ratio of 50 g of enzyme per ton of feed as recommended by the fabricant. One hundred and sixty day-old chicks (Ross 308) were divided into 40 battery brooders, with a capacity of 4 animals per pen, exposed to constant light for the

duration of the trial. Water and a barley-based feed were available *ad libitum* throughout the experiment and were provided from automatic drinking nipples and a hanging feeder, respectively. The brooders were located in an environmentally-controlled room, which was adjusted daily to the recommended temperatures according to standard brooding practice. Birds were individually weighed at the commencement of the experiment and were randomly assigned into one of the four treatments, with 10 replicates of four birds per treatment. Weekly, feed consumption and individual body weights were recorded. Gain to feed ratios were calculated by dividing the weight gain per pen, per week and at the end of the experiment, including the weight gain of any dead birds, by the total feed consumed during the respective period. Bird mortality was assigned to daily. At the end of the experiment, at day 28, one bird per pen was slaughtered by an intravenous injection of an aqueous isotonic solution of 125 mg Tiopental Braun (Braun, Barcelona, Spain). The size of the various GI compartments was measured and digesta samples were collected and stored at -20 ºC for later analysis. Levels of cellulase and hemicellulase activity in the GI tract were measured as described below.

Table 5.1. Ingredient composition and estimated analysis of the cereal-based feed.

Ingredients	**g/kg**
Barley	550.00
Soybean meal 47%	300.61
Soybean oil	50.73
Maize	50.45
Salt	2.50
Calcium carbonate	8.10
Dicalcium phosphate[1]	10.79
DL-Methionine	1.60
Mineral and vitamin premix[2]	2.00
Estimated nutrient content	
Energy (MJ ME/kg DM)	12
Crude Protein	208.0
Ether extract	73.3
Crude cellulose	48.7

[1]Contained 200 g/kg Ca and 180 g/kg P.

[2]Mineral-vitamin premix provided the following per kilogram of diet: biotin 0.5 mg, calcium pantothenate 10 mg, cholecalciferol 0.05 mg, cyanocobalamin 0.12 mg, folic acid 0.5 mg, menadione 2 mg, nicotinic acid 30 mg, pyridoxine 1.7 mg, retinol 2.7 mg, thiamin 1 mg, α-tocopherol 20 mg, riboflavin 4.2 mg, Co 0.2 mg, Cu 10 mg, Fe 80 mg, I 1 mg, Mn 100 mg, Se 0.3 mg, Zn 80 mg, monensin 0.1 g.

5.2.3. Analytical procedures

To standardize the number of enzyme units used to supplement the basal diet, the catalytic activity of the various exogenous enzymes was determined under identical experimental conditions. Catalytic activity was determined at 40 ºC by measuring reducing sugar released, following the method described by Taylor *et al*. (2005), using barley β-glucan (Megazyme®, Ireland) as the substrate. One unit of catalytic activity is defined as the amount of enzyme required to release one µmol of product per min. The extract containing Rovabio™ Excel AP (Rov) enzymes was prepared by ressuspending 75 mg of the enzyme mixture in 10 mL of 50 mM NaHEPES buffer, pH 7.5, which was followed by an incubation overnight at 4 ºC with gentle agitation and a centrifugation at 13,000*g* for 5 min.

Previously to detection of β-glucanase activity, digesta samples were centrifuged and the supernatant recovered for analysis. Initially, qualitative analysis of cellulase activity in the digesta samples recovered from the various GI compartments was assessed in agar plates, using barley β-glucan (Megazyme®, Ireland) at 0.1% (w/v) final concentration, in 10 mM Tris-HCl pH 7.0. Catalytic activity was detected after 16 hour incubation at 37 ºC through the Congo Red assay plate, as described in Ponte *et al.* (2004) and Mourão *et al*. (2006). Zymogram analysis was performed as described by Fontes *et al*. (2004). Briefly, digesta proteins were separated through SDS-PAGE in 10% 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 four 30 min washes in 100 mM sodium succinate, pH 6.3, containing 10 mM $CaCl_2$ and 1 mM DTT. The gel was incubated overnight 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, β-glucanase activity was detected using a 0.1% (w/v) Congo Red solution, for 15 min and washing with 1 M NaCl until excess dye was removed. Areas of catalytic activity appeared as colourless zones in a dark blue background after a quick wash in a 0.5% (v/v) solution of acetic acid. For measuring the viscosity of small intestine contents, samples were centrifuged for 10 min at 9000 rpm and the viscosity of the supernatant was measured using a Brookfield viscometer (Model LVDVCP-II, Brookfield Engineering Laboratories, Middleboro, MA) whose cup was maintained at 24 ºC.

5.2.4. Statistical analysis

Statistical analysis of data related to bird's performance was conducted by analysis of variance, using the General Linear Models procedure of SAS (SAS Inst. Inc., Cary, NC). Means with a significant F ratio were separated by the least significant difference test. The experimental unit was a cage of four birds. Unless otherwise stated, differences were considered significant when $P<0.05$.

5.3. RESULTS AND DISCUSSION

In order to evaluate the importance of a family 11 CBM in the function of a recombinant β-glucanase *in vivo*, truncated forms of the modular cellulase *Ct*Lic26A-Cel5E of *C. thermocellum*, with or without the enzyme's β-glucan-binding domain, *Ct*CBM11, were produced and used to supplement a barley-based diet for broiler chicks. The basal diets were prepared, supplemented with the required enzymes and used to feed broiler chicks *ad libitum* from days 1-28. During the experiment, the mortality rate was low (3.1%) and was not related to treatments.

5.3.1. Bird performance

Data concerning the body weight, weight gain, feed intake and gain:feed ratio throughout the experiment are summarized in Table 5.2. The data show that final body weight of birds fed on the barley-based diets supplemented with exogenous polysaccharidases was significantly higher than that of the control birds not supplemented with the microbial biocatalysts. Differences in body weight were visible as soon as day seven and remained significant throughout the duration of the experiment. Differences in body weight and weight gain in birds receiving the recombinant cellulases, with or without the family 11 CBM, or the commercial enzyme mixture were not significant. In addition, feed intake was higher in birds fed on diets supplemented with β-glucanases, especially in the first three weeks of the experiment. During the last week of the trial, broiler chicken fed on the basal diet not supplemented with the exogenous polysaccharidases had similar feed intakes when compared to birds from the other treatments. Since there were no differences between the gain:feed ratios of the four groups, the data suggest that an improvement in feed intake was responsible for the better performances displayed by birds fed on diets containing the microbial

cellulases. Taken together, the results suggest that, at these levels of enzyme incorporation, the family 11 CBM of *Ct*Lic26A-Cel5E of *C. thermocellum* did not improve the biological effectiveness of the associated GH catalytic modules, when the recombinant enzymes were used as a supplement for barley-based diets for poultry.

Table 5.2. Growth performance of broilers fed on a barley-based diet not supplemented (C0) or supplemented with a commercial cellulase mixture (Rov) or truncated derivatives of *C. thermocellum* *Ct*Lic26A-Cel5E *β*-glucanase containing (Lic26-Cel5E-CBM11) or not containing (Lic26-Cel5E) a family 11 CBM.

	C0	**Rov**	**Lic26-Cel5E**	**Lic26-Cel5E-CBM11**	**SEM[3]**	**p(*F*)**
Body Weight (g)						
0d	42.4	42.6	42.2	42.1	0.152	NS[2]
7d	142.8[b]	154.9[a]	154.1[a]	148.1[ab]	3.209	0.028
14d	331.7[b]	366.7[a]	362.6[a]	355.5[a]	8.106	0.014
21d	674.1[b]	739.7[a]	725.1[a]	732.6[a]	13.426	0.004
28d	1158.0[b]	1243.4[a]	1240.4[a]	1247.9[a]	21.313	0.009
Weight Gain (g)						
0-7d	100.5[b]	112.4[a]	111.8[a]	106.0[ab]	3.195	0.030
7-14d	188.9[b]	211.6[a]	208.6[a]	207.5[a]	5.745	0.023
14-21d	342.5[b]	373.0[a]	362.4[a]	377.1[a]	7.587	0.008
21-28d	484.0	503.8	515.4	512.4	10.544	NS
0-28d	1115.7[b]	1200.8[a]	1198.2[a]	1205.9[a]	21.288	0.009
Feed Intake[1] (g)						
0-7d	120.6[b]	138.6[a]	138.2[a]	128.8[ab]	3.81	0.003
7-14d	278.8[b]	316.3[a]	318.6[a]	309.2[a]	10.209	0.024
14-21d	552.4[b]	609.9[a]	587.8[ab]	604.5[a]	16.875	0.066
21-28d	851.6	866.9	888.8	895.7	23.698	NS
0-28d	1803.3[b]	1931.4[a]	1933.2[a]	1937.7[a]	42.093	0.063
Gain:Feed Ratio						
0-7d	0.834	0.810	0.809	0.831	0.0173	NS
7-14d	0.714	0.669	0.639	0.671	0.0337	NS
14-21d	0.610	0.620	0.621	0.624	0.0148	NS
21-28d	0.570	0.526	0.581	0.576	0.0239	NS
0-28d	0.618	0.602	0.619	0.623	0.0131	NS

[1]For feed intake the p(*F*) values and significant differences were presented when P<0.1
[2]NS, not significant
[3]SEM, standard error mean

The effect of including exogenous cellulases in barley-based diets in the relative weight and length of the birds' organs and GI tract compartments were evaluated (Table 5.3). Enzyme supplementation had no effect on crop, gizzard and liver relative weights or on the duodenum, jejunum and caecum lengths. In contrast, ileum relative length was significantly reduced (P<0.05) in birds receiving the commercial enzyme mixture. In addition, diet supplementation with exogenous *β*-glucanase activities significantly

contributed to decrease the viscosity of small intestine contents (Table 5.3). It is well established that moderate and high levels of soluble NSPs in diets for broiler chicken contribute to increase the size of specific compartments of the bird's GI tract (Brenes *et al.*, 1993; Petersen *et al.*, 1993). Solubilization of structural polysaccharides contributes to increase digesta viscosity, such as observed in birds not supplemented with the exogenous enzymes, and this may contribute to a decreased rate of feed passage. Therefore, the lower feed intakes presented by birds given the control treatment may result from a higher digesta viscosity and a decreased feed passage rate. In addition, the increased viscosity and the decreased digesta passage rate contribute to increase the digesta bulk in the GI tract. As a consequence of this phenomenon the length of the small intestine is increased, which constitutes a potential physiological adaptation allowing for an improvement in feed consumption and nutrient uptake. However, although generally the dietary inclusion of cellulases and hemicellulases contribute to decrease intestine length (Brenes *et al.*, 2002), it is presently unknown why the recombinant enzymes used in this study were not effective for decreasing ileum length such as it was observed for the small intestine of birds receiving the commercial mixture.

Table 5.3. Relative weight and length of birds' organs and GI tract and viscosity of digesta samples of broilers fed on a barley-based diet not supplemented (C0) or supplemented with a commercial cellulase mixture (Rov) or truncated derivatives of *C. thermocellum Ct*Lic26A-Cel5E *β*-glucanase containing (Lic26-Cel5E-CBM11) or not containing (Lic26-Cel5E) a family 11 CBM.

	C0	**Rov**	**Lic26-Cel5E**	**Lic26-Cel5E-CBM11**	**SEM[2]**	**p(*F*)**
Relative Weight (g/100 g BW)						
Crop	3.86	3.90	4.32	4.62	0.230	NS[1]
Gizzard	14.59	14.60	12.83	13.85	0.924	NS
Liver	30.35	30.09	31.42	28.93	1.002	NS
Relative Length (cm/kg BW)						
Duodenum	19.7	18.0	18.6	19.4	0.54	NS
Jejunum	55.7	50.9	51.8	52.2	1.51	NS
Ileum	55.9[a]	50.7[b]	54.5[a]	52.8[ab]	1.34	0.044
Caecum	13.7	13.8	13.4	14.5	0.32	NS
Content Viscosity (cps)						
Duodenum+Jejunum	11.31[a]	5.53[b]	7.10[b]	7.24[b]	1.404	0.039
Ileum	16.98[a]	8.95[b]	10.26[b]	9.61[b]	1.831	0.009

[1]NS, not significant; [2]SEM, standard error mean.

Data presented here highlight the capacity of single recombinant cellulases to enhance the nutritive value of barley-based diets for poultry, questioning the need for using enzyme mixtures containing a large array of different enzyme specificities for targeting the anti-nutritive factors present in those diets. This is not completely unexpected, since data previously reported by Philip *et al.* (1995) suggested that a recombinant single-domain cellulase, which originates also from the anaerobic bacterium *C. thermocellum*, was as efficient as a complex mixture of cellulases in improving the nutritive value of a barley-based diet for broiler chicken. It has been previously shown that one of the major demonstrated actions of feed cellulases is to decrease the degree of polymerization of soluble β-glucans, through the randomly cleavage of glycosydic bonds in the polysaccharide backbone. The reduction in carbohydrate chain length contributes to decrease the levels of digesta viscosity (Fengler and Marquardt, 1988; Bedford and Morgan, 1996). Therefore, the data presented in this report indicates that single purified recombinant cellulases or enzyme mixtures containing cellulases, can equally perform this action. However, since the recombinant enzymes were incorporated in the barley-based diet at twice the level used by the commercial enzyme mixture it is presently unknown if at similar incorporation rates the exogenous enzymes could fulfil their roles with similar efficiencies. Only in that case, the presence of non-cellulase activities, such as xylanase, mannanase, galactanase or pectinase, in enzyme mixtures for supplementation of barley-based diets could be considered redundant.

It is well established that CBMs contribute to enhance the activity of adjacent catalytic modules by increasing enzyme concentration on the substrate surface (Fernandes *et al.*, 1999; Gilbert *et al.*, 2002). This action is particularly important in plant cell wall hydrolases that need to be targeted to their specific substrates, which are usually less accessible in the complex organization of the plant cell wall. The family 11 CBM of *Ct*Lic26A-Cel5E binds both β-1,4- and β-1,3-1,4-mixed linked glucans (Carvalho *et al.*, 2004). It is also well known that CBMs are particularly important for the hydrolysis of insoluble substrates (Gilbert *et al.*, 2002). It has been previously shown that, although the molar activity of the recombinant cellulases Lic26-Cel5E-CBM11 and Lic26-Cel5E against barley β-glucan is similar, the presence of the family 11 CBM promotes the action of the modular cellulase against insoluble cellulose forms, such as Avicel (Taylor *et al.*, 2005; Chapter 4). Since barley-based diets are relatively poor in insoluble polysaccharides, we envisaged that the major contribution of the

family 11 CBM of *Ct*Lic26A-Cel5E would be related with the targeting of the associated catalytic domains to the anti-nutritive soluble β-glucans. However, data presented here revealed that the family 11 CBM was unable to affect the efficiency of the GH5 and GH26 catalytic domains *in vivo*, as reflected by the inability of the tri-modular enzyme to improve the nutritive value of a barley-based diet for poultry when compared with the recombinant bi-modular cellulase lacking a non-catalytic CBM. Nevertheless, it is possible that a lack of efficacy of the family 11 CBM is related with the relatively high dosage of recombinant enzyme used in this study. In fact, the level of enzyme used for feed supplementation could have been so high that there would be always enzyme available in the proximity of the anti-nutritive substrate. Therefore, it is possible that the ultimate effect of the family 11 CBM in directing the recombinant enzyme to its target β-glucans can only be revealed at lower levels of enzyme supplementation. In addition, it is also possible that a fraction of the recombinant enzyme containing the non-catalytic domain could be retained by the insoluble cellulose, therefore decreasing the effective concentration of the cellulase available for depolymerizing the anti-nutritive β-glucans. If these two factors, individually or in conjunction, contributed for the lack of effectiveness of the family 11 CBM in the particular conditions of the present experiment, remains to be evaluated.

5.3.2. Recombinant β-glucanase stability *in vivo*

To evaluate the stability of the exogenous GHs during passage through the GI tract, β-glucanase activity was qualitatively determined in digesta samples collected in the various digestive compartments of ten animals per treatment. The data, exemplified in Figure 5.2 and collectively presented in Table 5.4, demonstrated that cellulase activity could be detected along the entire digestive tract of most animals fed on diets supplemented with the plant cell wall hydrolases. In addition, while caeca samples collected from birds of the group not receiving exogenous enzymes were positive for cellulase activity, no β-glucan degrading properties were detected in the contents of the other GI compartments. However, two of the animals of this negative control group presented low but detectable β-glucanase activity in the crop. Therefore, taken together, the results suggest that a considerable percentage of the exogenous enzymes resist to the acidic and proteolytic conditions that are prevalent in some portions of the digestive tract.

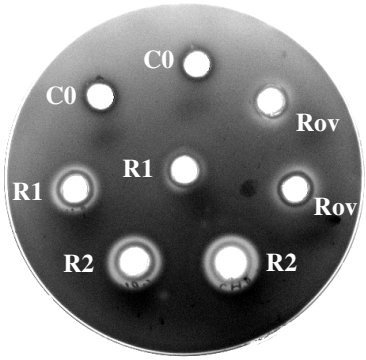

Figure 5.2. Detection of *β*-glucanase activity in the crop contents of broilers fed on barley-based diet not supplemented (C0) or supplemented with a commercial cellulase mixture (Rov) or truncated derivatives of *C. thermocellum Ct*Lic26A-Cel5E *β*-glucanase containing (R2, enzyme Lic26-Cel5E-CBM11) or not containing (R1, enzyme Lic26-Cel5E) a family 11 CBM.

To analyze potential changes in the molecular architecture of the recombinant cellulases during passage through the GI tract, digestive samples of birds of treatments receiving Lic26-Cel5E-CBM11 and Lic26-Cel5E were subjected to zymogram analysis. The data, displayed in Figure 5.3, indicate that both Lic26-Cel5E-CBM11 and Lic26-Cel5 are prone to proteolytic cleavage in the birds' GI tract, which occurs initially but moderately in the crop and then completely in the gizzard and in the following GI compartments. Therefore, it is suggested that both recombinant enzymes are proteolytically cleaved in the linker regions connecting the Lic26, the Cel5 and the CBM11 modules, which contributes to release the two 32-35 kDa catalytic domains that still retain significant catalytic activity.

Table 5.4. Number of birds, out of ten animals analyzed, fed on a barley-based diet not supplemented (C0) or supplemented with a commercial cellulase mixture (Rov) or truncated derivatives of *C. thermocellum Ct*Lic26A-Cel5E *β*-glucanase containing (Lic26-Cel5E-CBM11) or not containing (Lic26-Cel5E) a family 11 CBM presenting *β*-glucanase activity in digesta samples collected from various gastrointestinal compartments.

	C0	**Rov**	**Lic26-Cel5E**	**Lic26-Cel5E-CBM11**
Crop	2	7	8	7
Gizzard	0	3	3	4
Duodenum	0	3	3	3
Jejunum	0	0	3	2
Caecum	10	10	10	10

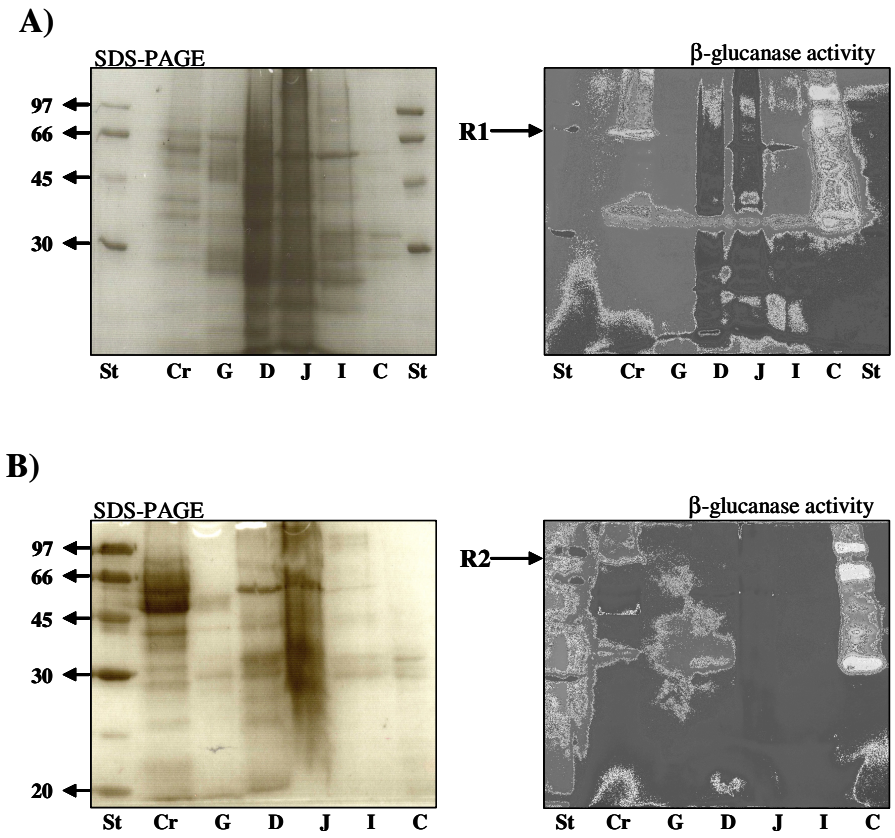

Figure 5.3. Zymogram analysis of digesta samples collected from various regions of the GI tract of birds fed on a barley-based diet supplemented with the recombinant β-glucanases Lic26-Cel5 (A) or Lic26-Cel5-CBM11 (B).
Proteins were fractionated through SDS-PAGE and stained for β-glucanase activity after enzyme renaturation. Abbreviations: St, low molecular mass protein standards (kDa); Cr, Crop; G, Gizzard; D, Duodenum; J, Jejunum; I, Ileum; C, Caecum. The location of the two exogenous recombinant enzymes is highlighted (R1, Lic26-Cel5 and R2, Lic26-Cel5-CBM11).

Experiments performed *in vitro*, using purified Lic26-Cel5E-CBM11 and Lic26-Cel5, demonstrated that incubation of both recombinant cellulases with a pancreatic extract of peptidases (EC 3.4) lead to a proteolytic cleavage of the recombinant enzymes at the linker sequences, as it was observed here *in vivo* (Ponte *et al.*, 2008). The data confirmed that this phenomenon does not affect the biological capacity of the resulting catalytic domains to degrade purified soluble β-glucans *in vitro*, as it was previously demonstrated (Ponte *et al.*, 2008). However, it is possible that the proteolytical removal of the β-glucan binding domain from the Lic26-Cel5E-CBM11 could have contributed to reduce its capacity to interact with the anti-nutritive soluble polysaccharide in the duodenum and the subsequent GI compartments. If the similar efficacy presented by the two recombinant enzymes analyzed in this study *in vivo*, results from the proteolytic transformation of both microbial recombinant enzymes in

similar polypeptides after the crop, remains to be established. Notwithstanding this possibility, it was recently shown that the proteolytic release of a xylan-binding domain from a recombinant xylanase in the duodenum of broiler chicks have no effect on the capacity of the recombinant enzyme to improve the nutritive value of a wheat-base diet, when compared with the xylanase derivative consisting only on the enzyme catalytic domain (Fontes *et al*., 2006).

5.4. CONCLUSIONS

Results presented here suggest that individual recombinant cellulases could be as effective as complex mixtures of GHs in attenuating the depreciative effects of soluble polysaccharides found in barley-based diets. In addition, the data suggest that when incorporated at high dosage rates (30 U/kg of basal diet) a tri-modular cellulase containing a family 11 CBM and its double-domain counterpart consisting on the enzyme's catalytic modules have equal capacities to improve the nutritive value of a barley-based diet. Furthermore, both recombinant enzymes were prone to peptidolysis in the birds' GI tract. It is suggested that this process, although not affecting the capacity of the resulting enzymes to degrade the anti-nutritive β-glucans *in vitro*, could have influenced the capacity of Lic26-Cel5E-CBM11 to act *in vivo*. The capacity of the non-catalytic family 11 CBM to elicit the function of *Ct*Lic26A-Cel5E recombinant derivatives, when incorporated at lower dosages rates in barley-based diets for poultry, is currently under investigation.

Ackowledgements

We thank Sociedade Agrícola da Quinta da Freiria SA for supplying the 0 day old chicks and Reagro for providing the feed ingredients used in these experiments. This work was supported by Fundação para a Ciência e a Tecnologia (POCI/CVT/69329/2006). CG, TR and PP were supported by Fundação para a Ciência e a Tecnologia through grants SFRH/BD/16731/2004, SFRH/BD/32321/2006 and SFRH/BD/17969/2004, respectively.

CHAPTER 6 *Escherichia coli* expression and purification of four antimicrobial peptides fused to a family 3 Carbohydrate-Binding Module (CBM) from *Clostridium thermocellum*

Catarina I.P.D. Guerreiro[1], Carlos M.G.A. Fontes[1], Miguel Gama[2] and Lucília Domingues[2]

[1]CIISA, Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; [2]IBB-Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, 4710-057 Braga, Portugal

Adapted from *Protein Expr. Purif.* 59, 161-168

Abstract

Antimicrobial peptides (AMPs) are molecules that act in a wide range of physiological defensive mechanisms developed to counteract bacteria, fungi, parasites and viruses. Several hundreds of AMPs have been identified and characterized. These molecules are presently gaining increasing importance, as a consequence of their remarkable resistance to microorganism adaptation. CBMs are non-catalytic domains that anchor GHs into complex carbohydrates. *Clostridium thermocellum* produces a multi-enzyme complex of cellulases and hemicellulases, termed the cellulosome, which is organized by the scaffoldin protein CipA. Binding of the cellulosome to the plant cell wall results from the action of CipA family 3 CBM (CBM3), which presents a high affinity for crystalline cellulose. Here CipA family 3 CBM was fused to four different AMPs using recombinant DNA technology and the fusion recombinant proteins were expressed at high levels in *Escherichia coli* cells. CBM3 does not present antibacterial activity and does not bind to the bacterial surface. However, the four recombinant proteins retained the ability to bind cellulose, suggesting that CBM3 is a good candidate polypeptide to direct the binding of AMPs into cellulosic supports. A comprehensive characterization of the antimicrobial activity of the recombinant fusion proteins is currently under evaluation.

6.1. INTRODUCTION

AMPs are cationic molecules with a wide range of antimicrobial activities, which vary from bacteria to fungi, parasites and viruses. These molecules have been recognized as an important component of the non-specific host defense system and innate immunity of all animal classes, including insects, amphibians and mammals (Kang *et al.*, 1999). Several hundreds of AMPs have already been identified and characterized (http://www.bbcm.units.it/~tossi; April 2008). These molecules represent very appealing novel tools for fighting infections, since microorganisms do not easily develop resistance mechanisms against their mode of action as it has been described for other types of antimicrobial compounds. This is probably due to the particular capacity of AMPs' to bind and disrupt cellular membranes. It is well recognized that mutations in the genetic information encoding the overall plasma membrane structure and which may counteract the action of AMPs are less likely to occur than variations in the enzyme physiology within a host cell (Lee *et al.*, 2002).

In early studies involving AMPs, these molecules were obtained through purification from their original hosts or by chemical synthesis. This, of course, presented many restrictions to AMPs research mainly in what concerned the production of detectable quantities of the peptides. In recent years, many small cationic AMPs have been successfully produced using recombinant DNA methods in heterologous hosts (Peng *et al.*, 2004; Yang *et al.*, 2004; Čipáková *et al.*, 2006; Xu *et al.*, 2006), allowing the purification of larger quantities of the recombinant proteins. It is presently recognized that producing cationic AMPs using recombinant techniques will accelerate research on their pharmaceutical potential and potential clinical application (Xu *et al.*, 2006). Many host cells have been selected for expression of AMPs but *Escherichia coli* has been established as one of the most popular recombinant bioreactors due to its fast growth rate and well-established expression systems (Xu *et al.*, 2006). Since AMPs are usually very small to be produced effectively individually, many authors aimed to express them in fusion with other molecules (Peng *et al.*, 2004; Moon *et al.*, 2006; Xu *et al.*, 2006; Xu *et al.*, 2007). Problems concerning low levels of expression, poor solubility, toxicity for the host cell, product proteolysis and poor recovery yields are usually overcome by the implementation of the fusion technology (Wei *et al.*, 2005).

CBMs are non-catalytic domains present in GHs which target the associated catalytic modules to their substrates, therefore potentiating enzyme activity. Based on primary structure similarity, CBMs have been grouped into 51 sequence-based families

(http://afmb.cnrs-mrs.fr/CAZY; April 2008). Structural studies on representatives of the majority of CBM families demonstrate that the topology of CBM ligand-binding sites complements the conformation of the target polysaccharide. Thus, in Type A modules, which interact with the flat surfaces of crystalline polysaccharides, such as cellulose, the binding site comprises a planar hydrophobic platform that contains three exposed aromatic amino acids (Boraston *et al.*, 2004). In contrast, Type B CBMs, which bind to single polysaccharide chains, accommodate the ligands within extended clefts of varying depths (Boraston *et al.*, 2004). In general, the capacity of CBMs to specifically recognize a range of polysaccharides has been shown to be pivotal in a range of biotechnological applications (Tomme *et al.*, 2003).

The development of mechanisms for targeting bio-molecules to a diversity of matrixes is an emerging theme in biochemistry and biotechnology. In this work four different AMPs were fused to a family 3 CBM of CipA, a non-hydrolytic structural protein responsible for the assembly of *Clostridium thermocellum* multi-enzyme complex of cellulases and hemicellulases, termed the cellulosome (Tormo *et al.*, 1996). The major goal of this approach was not only to promote the levels of expression of the selected AMPs in *E. coli* but to develop a biotechnological tool that would allow fixing a variety of polypeptides, such as AMPs, onto cellulosic surfaces. Therefore, the immobilization of these fusion proteins in cellulosic supports will allow the generation of novel bio-products possessing antimicrobial properties. Cellulose is a safe and inert macromolecule and has excellent physical properties. In addition, it is cheap and commercially available in many different forms and has been approved for many pharmaceutical and human uses (Tomme *et al.*, 2003).

6.2. EXPERIMENTAL PROCEDURES

6.2.1. Cloning of the DNA sequence of CBM3 with a N-terminal linker

The gene encoding CBM3 fused to the endogenous CipA N-terminal linker sequence (LK) was isolated through PCR. The inclusion of the linker region in the encoding DNA sequence was envisaged to improve flexibility in the resulting fusion proteins. The DNA fragment encoding CBM3 and its N-terminal LK from *Clostridium thermocellum* was amplified by PCR from *C. thermocellum* YS genomic DNA with the thermostable DNA polymerase *Pfu* Turbo (Stratagene). The primers used were 5'-ACACCGACCAAGGGAGCAACA-3' (forward primer) and 5'-

TTCTTTACCCCATACAAGAAC-3' (reverse primer). PCR was performed as follows: preheating at 95 ºC for 3 min, 25 cycles at 95 ºC for 20 s, 55 ºC for 30 s and 72 ºC for 30 s, followed by a final elongation period at 72 ºC for 10 min. The amplified product was recovered from a 1% agarose gel using the Gel Band Purification Kit (GE Health) and ligated into pMOS*Blue* (GE Health) to generate pCG1. The DNA insert of pCG1 was sequenced to ensure that no mutations had occurred during the PCR. This plasmid was used as template for the subsequent amplifications aiming at generating the fusion genes.

6.2.2. Cloning of the DNA sequences encoding the fusion proteins

For these studies four cationic and amphiphilic peptides were selected: peptide-1 (LKLLKKL), peptide-2 (LKLLKKLLKLLKKLGGGK), peptide-3 (LKKLLKKLKKLLKK) and porcine myeloid antibacterial peptide-23 (PMAP-23) (RIIDLLWRVRRPQKPKFVTVWVR) (Zanetti *et al.*, 1994), as depicted in Table 6.1. The nucleotide sequence of each peptide was engineered to allow codon optimization in *E. coli*. Sequences encoding the four peptides were included in the primers and fusion genes were generated through PCR, using the DNA polymerase FideliTaq (GE Health), utilizing plasmid pCG1 as template. The sequences of the primers used in this study for this particular aim are shown in Table 6.2. Primers included *Nhe*I and *Xho*I restriction sites. To allow the sub-cloning of the gene encoding CBM3 fused to the linker sequence (LK-CBM3) into the expression vector, the DNA sequence encoding this truncated version of CipA was also amplified by PCR (Table 6.2). The PCR reactions were performed as follows: preheating at 95 ºC for 2 min, 25 cycles at 95 ºC for 30 s, 55 ºC for 30 s and 68 ºC for 1 min, followed by a final elongation period of 68 ºC for 5 min. Amplified fragments were recovered from a 1% agarose gel using the Qiagen Gel Extraction Kit (Qiagen), ligated into plasmid pMOS*Blue* and sequenced to ensure that no mutations had occurred during the PCR. The generated recombinant plasmids were digested with *Nhe*I and *Xho*I and the excised products were cloned into expression vector pET21a (Novagen), previously digested with the same restriction enzymes. This vector carries a T7*lac* promotor and the resulting recombinant proteins contained a C-terminal His_6-tag to facilitate purification.

Table 6.1. Primary sequence of antimicrobial peptides used for fusion with a family 3 CBM.

Antimicrobial peptide	No. of residues	Primary sequence	References
peptide-1	7	LKLLKKL	
peptide-2	18	LKLLKKLLKLLKKLGGGK	
peptide-3	14	LKKLLKKLKKLLKK	(Blondelle and Houghten, 1992)
PMAP-23	23	RIIDLLWRVRRPQKPKFVTVWVR	(Zanetti *et al.*, 1994)

Table 6.2. Primers used for cloning the DNA sequences encoding various AMPs in fusion with the gene of a family 3 CBM.

Construct	Primers	Sequence (5'-3')
Peptide-1-LK-CBM3	Forward	CTC**GCTAGC**CTGAAACTGCTGAAAAAACTGACACCGACCAAG GGAGCA[a]
	Reverse	CAC**CTCGAG**TTCTTTACCCCATACAAGAAC
Peptide-2-LK-CBM3	Forward	CTC**GCTAGC**CTGAAACTGCTGAAAAAACTGCTGAAACTGCTG AAAAAACTGGGTGGTGGTAAAACACCGACCAAGGGAGCA
	Reverse	CAC**CTCGAG**TTCTTTACCCCATACAAGAAC
Peptide-3-LK-CBM3	Forward	CTC**GCTAGC**CTGAAAAAACTGCTGAAAAAACTGAAAAAACT GCTGAAAAAAACACCGACCAAGGGAGCA
	Reverse	CAC**CTCGAG**TTCTTTACCCCATACAAGAAC
PMAP-23-LK-CBM3	Forward	CTC**GCTAGC**CGTATTATTGATCTGCTGTGGCGTGTGCGTCGTC CGCAGAAACCGAAATTTGTGACCGTGTGGGTGCGTACACCGA CCAAGGGAGCA
	Reverse	CAC**CTCGAG**TTCTTTACCCCATACAAGAAC
LK-CBM3	Forward	CTC**GCTAGC**CACACCGACCAAGGGAGCA
	Reverse	CAC**CTCGAG**TTCTTTACCCCATACAAGAAC

[a]The *Nhe*I and *Xho*I restriction sites introduced in the forward and reverse primers, respectively, are shown in bold.

6.2.3. Expression of recombinants LK-CBM and fused proteins

To optimize expression of the various recombinant proteins, different growing and induction conditions were tested for a variety of recombinant *E. coli* strains. Below there is a description of the optimized growing conditions obtained that allowed the expression of each specific protein. To express peptide-1 fused to LK-CBM3 (peptide-1-LK-CBM3) and peptide-3 fused to LK-CBM3 (peptide-3-LK-CBM3), *E. coli* Tuner (DE3) (Novagen) cells harboring the appropriate recombinant plasmids were cultured in LB containing 100 μg/mL ampicillin at 37 ºC to mid-exponential phase ($OD_{595\ nm}$=0.6). At this point IPTG was added to a final concentration of 0.2 mM and the cultures were incubated at 18 ºC and 30 ºC for 16 hours, respectively. The proteins LK-CBM3 and PMAP-23-LK-CBM3, which results from the fusion of LK-CBM3 with PMAP-23, were expressed in *E. coli* BL21 (DE3) (Novagen) cells harboring the appropriate recombinant plasmids and cultured in LB containing 100 μg/mL ampicillin at 37 ºC to mid-exponential phase ($OD_{595\ nm}$=0.6). At this point IPTG was added to a final

concentration of 1 mM and the cultures were incubated at 37 ºC for further 4 hours. Peptide-2 in fusion with LK-CBM3 (peptide-2-LK-CBM3) was found to be insoluble when expressed by different *E. coli* strains under a range of induction conditions, which included inducing temperatures varying from 16 to 37 ºC and IPTG concentrations ranging from 0.01 to 1 mM.

6.2.4. Purification of recombinant proteins

Recombinant *E. coli* cells were harvested by centrifugation at 7,000*g* at 4 ºC for 15 min, ressuspended in 20 mM Tris-HCl pH 7.0, 20 mM NaCl, 5 mM $CaCl_2$ ($2H_2O$) (Buffer C) and disrupted by sonication on ice for 12 min. The resulting cell-free extracts were collected by centrifugation at 20,440*g* at 4 ºC for 30 min and the His_6-tagged recombinant proteins purified by IMAC, using 5 mL Niquel Hi-Trap Columns (GE Health). The column charged with 0.1 M $NiSO_4$ was equilibrated with 20 mM Na_3PO_4, 500 mM NaCl, 40 mM Imidazole, pH 7.4. The cell extracts were loaded into the column, which was washed with the equilibration buffer. Finally, the recombinant proteins were eluted with 20 mM Na_3PO_4, 500 mM NaCl, 300 mM Imidazole, pH 7.4. After purification, proteins were buffer-exchanged into Buffer C, using PD-10 Sephadex G-25M gel-filtration columns (GE Health). Purified proteins were analyzed by mass spectrometry and SDS-PAGE using 12% (w/v) acrylamide gels. Predicted sizes for the recombinant proteins were 22.1 kDa for LK-CBM3, 23.0 kDa for peptide-1-LK-CBM3, 24.1 kDa for peptide-2-LK-CBM3, 23.8 kDa for peptide-3-LK-CBM3 and 25.1 kDa for PMAP-23-LK-CBM3.

6.2.5. Binding assays

Qualitative assessment of the binding of the recombinant proteins to Avicel (Sigma) was developed as follows: 100 µg of LK-CBM3, 50 µg of peptide-1-LK-CBM3, 50 µg of peptide-3-LK-CBM3 and 50 µg of PMAP-23-LK-CBM3, all in Buffer C, with 1% (v/v) Tween 20, were mixed with 4 mg of Avicel in a final reaction volume of 200 µL. The mixture was incubated for 1 h at 30 ºC with vigorous shaking. The insoluble ligand was collected by centrifugation at 13,000*g* at room temperature for 10 min. The supernatant, containing the unbound fraction, was recovered and the resulting pellet was washed four times with 400 µL Buffer C and 20 µL 1% (v/v) Tween 20. Finally, the carbohydrate pellet was ressuspended in 100 µL 10% (w/v) SDS and 25 µL

5× SDS Sample Buffer. Bound and unbound fractions were analyzed by SDS-PAGE using a 12% (w/v) acrylamide gel.

6.2.6. Antimicrobial assay

An antimicrobial assay for assessing the antibacterial properties of LK-CBM3 was implemented based on the procedures described by Morassuti *et al.* (2005). Briefly, 100 mL of LB were inoculated with colonies from an overnight culture plate of *E. coli* XL10 Gold (Stratagene) and grown at 37 ºC until $OD_{600\ nm}$=0.5. The cells (1 mL) were recovered by centrifugation at 3,000*g* at 10 ºC for 10 min and the supernatant discarded. Cells were washed twice with 1 mL of Buffer C. Finally, bacteria were ressuspended in 1 mL of the same buffer and serially diluted until 10^{-4}. Precisely 150 µL of *E. coli* XL10 Gold 10^{-4} were incubated with 150 µL LK-CBM3 at three different protein concentrations (0.1 mg/mL, 1 mg/mL and 2 mg/mL, all diluted in Buffer C). Several controls were incorporated in the experiment, including the incubation of the cells solely in Buffer C or with BSA at the same protein concentrations used for LK-CBM3. Cells were incubated under the various conditions at 210 rpm in an orbital incubator at 37 ºC for 2 h. At the end of the incubation period, 50 µL of each sample were plated onto LB agar plates and the number of bacterial colonies formed was counted after a 16 h incubation at 37 ºC. The remaining of the non-plated mixture was centrifuged at 13,000*g* at 4 ºC for 5 min and the supernatants used for protein quantification using Bradford reagent (Bradford, 1976) and analyzed by SDS-PAGE using a 14% (w/v) acrylamide gel.

6.3. Results and Discussion

6.3.1. Expression and purification of recombinant LK-CBM and fused proteins

Four peptides, termed peptide-1, peptide-2, peptide-3 and PMAP-23 (Zanetti *et al.*, 1994), which details are presented in Table 6.1, were fused with the family 3 CBM of CipA from *Clostridium thermocellum*. The recombinant fusion proteins were engineered to contain an N-terminal AMP domain and a C-terminal His_6-tag. Peptide-1, peptide-2 and peptide-3 are synthetic amphiphilic cationic peptides, which design was based on the studies provided by other authors (Blondelle and Houghten, 1992; Haynie *et al.*, 1995). Similar molecules have already proved to have potent antimicrobial activity against Gram negative and Gram positive bacteria (Blondelle and Houghten,

1992). It has also been previously shown that peptides composed of both hydrophobic and hydrophilic residues, such as leucine (L) and lysine (K), present antimicrobial activity against *Escherichia coli* and *Staphylococcus aureus*, both in the soluble form or immobilized in a water-insoluble non-degradable polymer support (Haynie *et al.*, 1995). These apparently simple cationic peptides with proven detectable antimicrobial activity were selected to fuse with *Ct*CBM3. In addition, a different peptide with antimicrobial activity, termed PMAP-23, was also selected for fusion with the CBM (Zanetti *et al.*, 1994). This peptide derives from porcine myeloid cells, is highly cationic and also tends to form an amphipathic structure, typical of other antibacterial peptides (Kang *et al.*, 1999). Synthetic PMAP-23 showed antibacterial activity against Gram positive and negative strains but no lytic activity against human erythrocytes, which suggested a certain degree of selectivity in disrupting cell membranes (Zanetti *et al.*, 1994; Kang *et al.*, 1999). It has been suggested that PMAP-23 antibacterial activity results from the interaction of the peptide with the bacteria cell membrane, followed by bacteria membrane alteration (Zanetti *et al.*, 1994; Kang *et al.*, 1999).

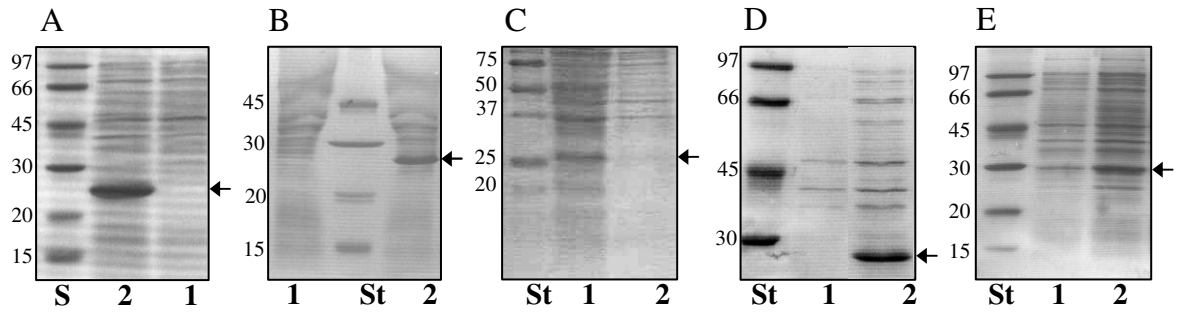

Figure 6.1. Expression of the recombinant proteins LK-CBM3 (A), peptide-1-LK-CBM3 (B), peptide-2-LK-CBM3 (C), peptide-3-LK-CBM3 (D) and PMAP-23-LK-CBM3 (E).
Lanes: 1, uninduced cells; 2, cells induced with IPTG; St, molecular mass protein standards (kDa). Protein extracts of lanes 1 and 2 result from the ressuspention of the bacterial cells in Buffer C. Arrows highlight the position of each one of the recombinant proteins.

Data presented in Figure 6.1, confirms that all four recombinant fusion proteins were expressed in *E. coli*. However, only peptide-1-LK-CBM3, peptide-3-LK-CBM3, PMAP-23-LK-CBM3 and LK-CBM3 were expressed in the soluble fraction of *E. coli*, although as a result of several optimization experiments. In contrast, peptide-2-LK-CBM3 formed inclusion bodies in *E. coli* and protein precipitation *in vivo* was persistent while using different *E. coli* expression strains and different growth and

induction temperatures (Figure 6.2 and Figure 6.3). To overcome the insolubility of some AMPs, as well as their citotoxic properties, some authors expressed these peptides in *E. coli* in fusion with different protein partners, utilizing different expression systems. Therefore, SMAP-29 was expressed using the intein-mediated system (Morassuti *et al.*, 2005), the human *β*-defensin 2 with the thioredoxin expression system (Peng *et al.*, 2004; Xu *et al.*, 2006), Hal18 by fusion with baculoviral polyhedron protein (Wei *et al.*, 2005) and LL-37 utilizing the glutathione S-transferase fusion system (Moon *et al.*, 2006), amongst many other examples. In addition, enterocins A and B, which are bacteriocins from *Enterococcus faecium*, were produced in fusion with a microbial CBM, although the experiment aimed at developing an easier purification system for the recombinant proteins (Klocke *et al.*, 2005). In general, the aim of these experiments was to develop a mechanism to produce at high levels the fusion protein in order to enhance the production of the AMP itself. Therefore, after purification of the fused proteins, a proteolytic digestion step using a specific protease is included in order to release the AMP in the purified form. In contrast with the above mentioned experiments, the aim of the work reported here was to obtain the AMPs fused with *Ct*CBM3 as the final product. More than a mere fusion partner, CBM3 was used to target the AMP for cellulosic surfaces. Notwithstanding this major goal of the experiments, it can be anticipated that the expression and purification of the AMP fusion recombinant proteins was probably facilitated by the presence of CBM3 and the C-terminal His_6-tag.

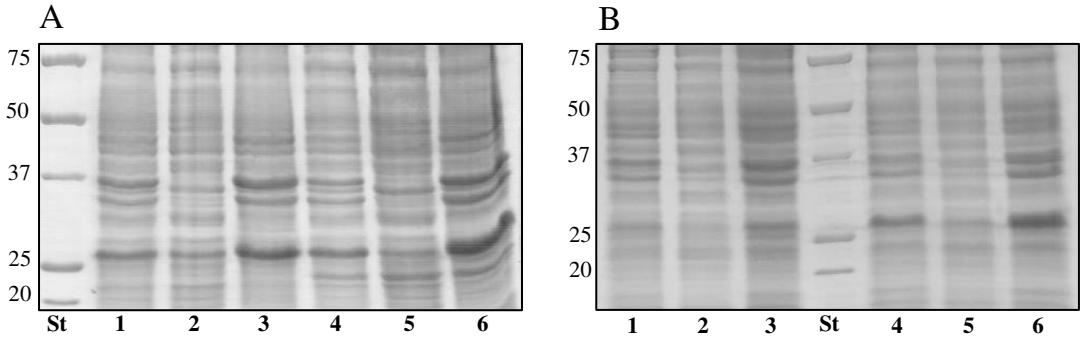

Figure 6.2. Expression of recombinant peptide-2-LK-CBM3 in different *E. coli* strains grown under different growth and induction temperatures.
(A) *E. coli* Tuner (lanes 1-3) and BL21 (DE3)pLysS (lanes 4-6) grown at 30 ºC, induced at 30 ºC. (B) BL21 (DE3)pLysS grown at 37 ºC, induced at 18 ºC (lanes 1-3) and at 30 ºC (lanes 4-6). Lanes 1 and 4, cells induced with IPTG; lanes 2 and 5, soluble cell-free extract; lanes 3 and 6, insoluble cell pellet; St, molecular mass protein standards (kDa).

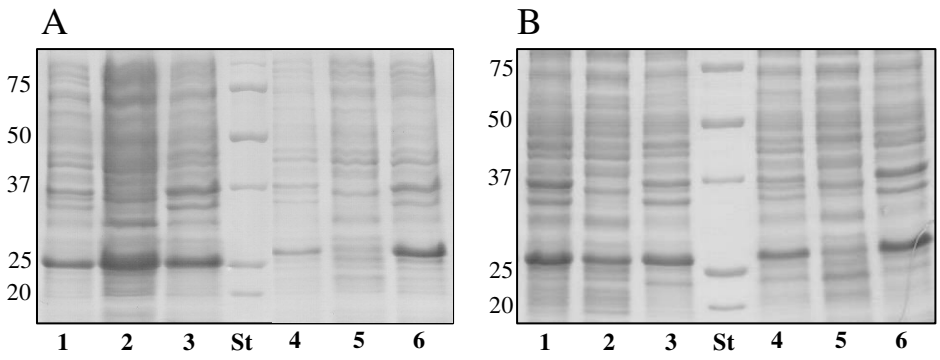

Figure 6.3. Comparing the expression and solubility of recombinant peptide-2-LK-CBM3 with two other recombinant proteins expressed in *E. coli* Tuner.
(A) peptide-1-LK-CBM3 (lanes 1-3) and peptide-2-LK-CBM3 (lanes 4-6) induced at 18 ºC. (B) peptide-3-LK-CBM3 (lanes 1-3) and peptide-2-LK-CBM3 (lanes 4-6) induced at 30 ºC. Lanes 1 and 4, cells induced with IPTG; lanes 2 and 5, soluble cell-free extract; lanes 3 and 6, insoluble cell pellet; St, molecular mass protein standards (kDa).

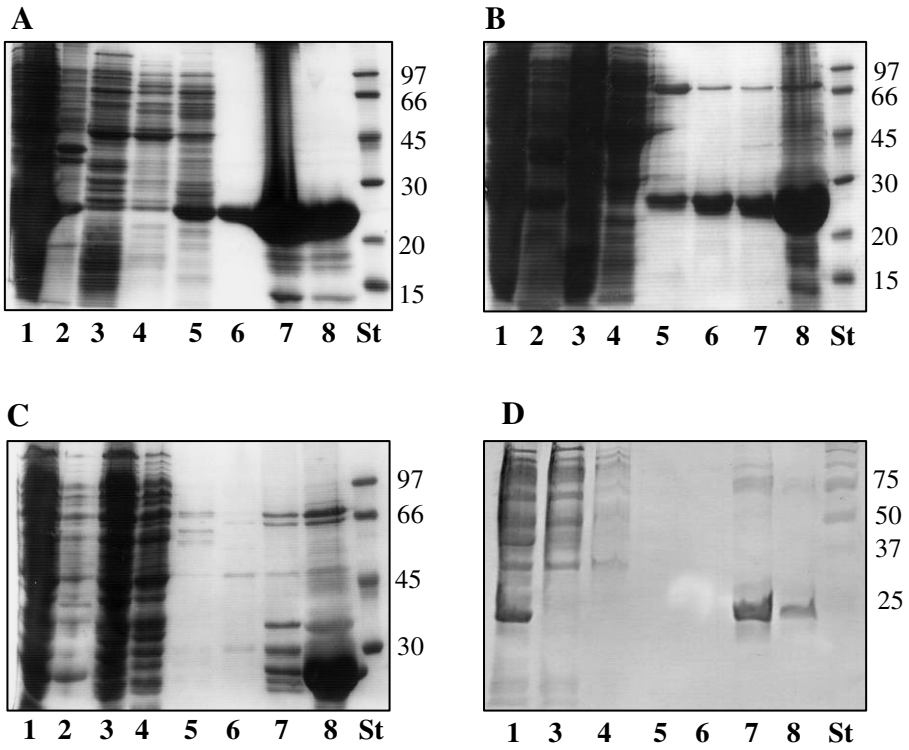

Figure 6.4. Purification of the recombinant proteins LK-CBM3 (A), peptide-1-LK-CBM3 (B), peptide-3-LK-CBM3 (C) and PMAP-23-LK-CBM3 (D) by affinity chromatography.
Lanes: 1, soluble cell-free extract; 2, insoluble cell pellet; 3, column filtrate; 4, 5 and 6, column washes; 7 and 8, pure protein; St, molecular mass protein standards (kDa).

After optimizing expression in *E. coli*, the recombinant proteins were purified through IMAC. The data, presented in Figure 6.4, demonstrate that all proteins were purified from most *E. coli* contaminants and were obtained at high concentrations.

Following the purification step, the proteins were buffer exchanged into Buffer C and their concentration calculated. Purified proteins presented final concentrations ranging from 0,5 to 2,5 mg/mL. Mass spectrometry analysis confirmed the expected sizes and, therefore, the integrity of the four recombinant proteins, suggesting that the N-terminal AMP domain was not subjected to proteolysis in *E. coli* and during the purification steps.

6.3.2. Binding assays

*Ct*CBM3 from CipA is a very well characterized and studied CBM (Tormo *et al.*, 1996; Tomme *et al.*, 2003) and its three-dimensional structure has already been elucidated (Tormo *et al.*, 1996). It is a Type A CBM, which means that it possesses a planar surface that interacts tightly with crystalline cellulose. *Ct*CBM3 equilibrium affinity constant (K_a) for Avicel was reported to be 7.7×10^6 M^{-1} (Tomme *et al.*, 2003). To investigate whether the fusion of each of the four AMPs to *Ct*CBM3 had some effect on its capacity to interact with crystalline cellulose, the capacity of the recombinant proteins to bind Avicel was evaluated. The data, presented in Figure 6.5, confirmed that all AMPs-LK-CBM3 retained capacity to bind Avicel, as revealed by the minor quantity of protein present in the unbound fraction and the resulting association with the carbohydrate fraction.

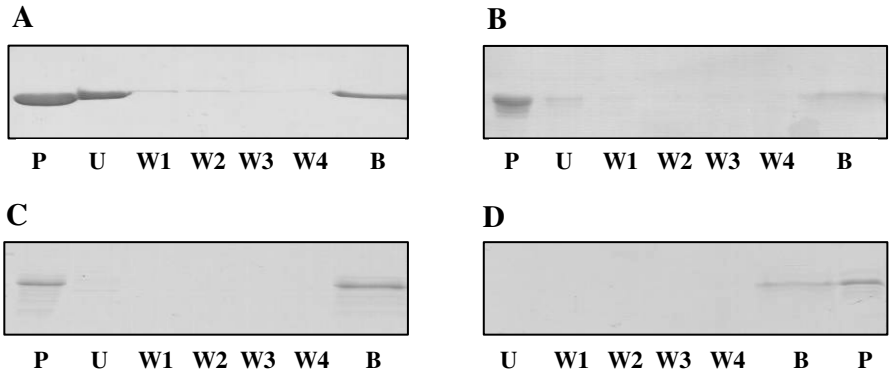

Figure 6.5. Binding of LK-CBM3 (A), peptide-1-LK-CBM3 (B), peptide-3-LK-CBM3 (C) and PMAP-23-LK-CBM3 (D) to Avicel.
Lanes: P, purified protein used in the experiment; U, unbound protein; W1, W2, W3 and W4, wash fractions; B, bound protein.

Although a small quantity of LK-CBM3 was shown not to bind the structural carbohydrate (Panel A, lane U), this phenomenon is certainly due to a relative excess of protein concentration in relation to the Avicel. In addition, the data also suggested that incorporation of His_6-tag in the recombinant proteins had no effect in the capacity of the protein to bind cellulose. Taken together the results suggest that fusion of CBM3 with AMPs had no affect in cellulose-binding activity of the CBM, which makes the recombinant proteins potential candidates to target AMPs to cellulosic supports.

6.3.3. Evaluating the antimicrobial activity of CBM3

It has been reported that C-type lectins, another kind of carbohydrate-binding molecules, may present antimicrobial activity (Cash *et al.*, 2006; Sun *et al.*, 2008). Like CBMs, lectins are proteins that recognize carbohydrates and fulfil an important role in the immune system, which has been extensively studied over the last years. Lectins are well described but the understanding of the mechanisms by which they influence the immune responses is not well defined. However, it is believed that lectins are involved in the recognition of carbohydrates at cell surfaces (McGreal *et al.*, 2004). To find out whether the recombinant LK-CBM3 itself would present antibacterial activity, an assay was performed against *E. coli* XL10 Gold cells (Stratagene). BSA was incorporated in this experiment as a control protein which does not possess recognizable antimicrobial activity. Both proteins were used at three different concentrations: 0.1 mg/mL, 1 mg/mL and 2 mg/mL. The data, presented in Table 6.3, demonstrated that the numbers of CFU/mL are in the same range at all protein concentrations (LK-CBM3 and BSA) and had no variation when compared with the numbers of CFU following incubation with Buffer C. Therefore, the results suggest that LK-CBM3 did not affect the number of existing bacteria, confirming that this CBM does not present an antibacterial activity, in contrast to what has been described for C-type lectins (Cash *et al.*, 2006; Sun *et al.*, 2008). This observation is not particularly surprising in light of our present knowledge on the structure of Type A CBMs. It is well established that *Ct*CBM3 has a ligand binding specificity restricted for crystalline cellulose. In addition, its planar binding surface is not effective for interacting with individual carbohydrate chains that are usually present on the microorganisms' surface glycoproteins, as lectins do. Therefore, structural constrains would preclude the interaction of the family 3 CBM with the bacterial cell surface.

Table 6.3. Antimicrobial activity of LK-CBM3.
E. coli XL10 Gold was incubated with LK-CBM3 and BSA at three different concentrations and then plated onto LB agar plates. CFU/mL were calculated after a 16 h incubation at 37 ºC. The number of CFU/mL for the experiment using exclusively Buffer C was 8×10^2. The values are average of three replicates.

Protein concentration (mg/mL)	Proteins	
	LK-CBM3 ($\times10^2$ CFU/mL)	BSA ($\times10^2$ CFU/mL)
0.1	7	4
1	7	9
2	3	9

To confirm the inability of CBM3 to interact with the bacterial surface, LK-CBM3 and BSA were incubated with *E. coli* cells and bound protein was detected following cell centrifugation through the measuring of protein concentration in the supernatant. The data, presented in Table 6.4, confirms that the two proteins have no capacity to attach to the bacteria surface since the concentration of LK-CBM3 recovered by centrifugation after incubation with bacteria equals the one recovered after the same incubation with Buffer C, for all initial protein concentrations (0.1 mg/mL, 1 mg/mL and 2 mg/mL). A similar result was observed for BSA incubation with *E. coli* XL10 Gold and Buffer C (columns C and D from Table 6.4). Although a small quantity of protein was lost from the initial concentration (for the protein concentrations of 1 mg/mL and 2 mg/mL), this was probably due to protein precipitation during the assays as it was also observed for the incubation with Buffer C. The supernatants used for protein quantifications were also analyzed by SDS-PAGE. The data, presented in Figure 6.6, confirmed that lack of binding was not associated with LK-CBM3 (lanes 1, 2 and 3) and BSA (lanes 4, 5 and 6) proteolytic degradation.

Table 6.4. Interaction of LK-CBM3 and BSA with *E. coli* XL10 Gold.
LK-CBM3 and BSA were incubated with *E. coli* XL10 Gold and cell bound protein evaluated by measuring protein concentration on the supernatant. A, *E. coli* with LK-CBM3; B, Buffer C with LK-CBM3; C, *E. coli* with BSA and D, Buffer C with BSA.

Initial protein (mg/mL)	Unbound protein (mg/mL)			
	A	B	C	D
0.1	0.1	0.1	0.1	0.1
1	0.8	0.8	0.7	0.7
2	1.2	1.2	1.2	1.2

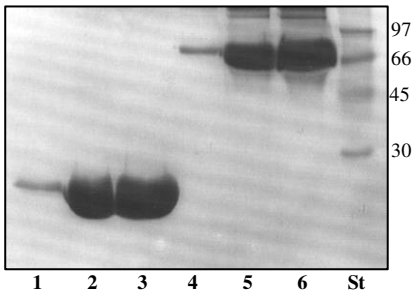

Figure 6.6. Binding of LK-CBM3 and BSA into *E. coli* XL10 Gold as judged by SDS 14% acrylamide gel analysis.
Lanes: 1, LK-CBM3 0.1 mg/mL; 2, LK-CBM3 1 mg/mL; 3, LK-CBM3 2 mg/mL; 4, BSA 0.1 mg/mL; 5, BSA 1 mg/mL; 6, BSA 2 mg/mL; St, molecular mass protein standards (kDa).

6.4. CONCLUSIONS

Here four AMPs were cloned and expressed in fusion with a family 3 CBM of *Clostridium thermocellum* cellulosome. Expression occurred in the soluble form for three recombinants and protocols for purifying the recombinant proteins by IMAC were developed. *Ct*CBM3 does not present detectable antimicrobial activity against *E. coli* XL10 Gold. Therefore, CBM3 is a good candidate polypeptide to direct binding of AMPs to cellulosic supports. Data presented here confirm that the fusion recombinants are now ready for further studies to measure and characterize their antimicrobial properties. Here it was shown that CBM3 may be attached to a non-related protein retaining its intrinsic biological activity (Tomme *et al.*, 2003). It has also been confirmed that AMPs produced in *Escherichia coli* fused to a CBM maintained their properties after purification and separation from the fusion partner (Klocke *et al.*, 2005). Therefore, the perspectives that the fusion proteins obtained in this work can display intrinsic antimicrobial activity are good and it is anticipated that their antimicrobial potency should be comparable to the AMPs expressed individually. Further studies will proceed in the future aiming at quantifying the antimicrobial properties of the proteins reported here, both in solution or fixed into insoluble surfaces.

Acknowledgments

This work was supported by the individual grant SFRH/BD/16731/2004 (to CG) from Fundação para a Ciência e a Tecnologia (Portugal). Authors wish to acknowledge the Protein Characterization and Mass Spectrometry Laboratory at Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal, for providing mass spectrometry data on the recombinant proteins.

CHAPTER 7 General Discussion and Future Perspectives

Plant cell walls are composed of an intricate network of polysaccharides, joined together by several molecular forces. Its degradation is one of the most important mechanisms for carbon recycling in nature. Glycoside hydrolases (GHs), as well as their associated carbohydrate-binding modules (CBMs), are involved in plant cell wall degradation and this work aims to elucidate some unclarified issues concerning their mechanisms of action.

Xyloglucan is the major hemicellulose in the plant cell wall of dicots. The importance of studying xyloglucan degradation is related with its role in plant cell wall morphogenesis and with its emerging biotechnological applications. Xyloglucan importance in food products, in novel pharmaceutical deliveries and in biofuel production has been increasing over the last years. In this work, structural and functional aspects of a GH and of two CBMs involved in xyloglucan hydrolysis and binding, respectively, were elucidated: xyloglucanase Xgh74A (Chapter 2) and *Ct*CBM30 and *Ct*CBM44 from *Ct*Cel9D-Cel44A (Chapter 3), all from the anaerobic thermophilic soil bacterium *Clostridium thermocellum*.

Data presented in Chapter 2 suggest that Xgh74A is a true xyloglucanase, since it possesses greater catalytic efficiency on xyloglucan than on any other substrate analyzed. The three-dimensional structure of Xgh74A was solved allowing the catalysis process to be further elucidated. The substrate-binding region lies in an open cleft and amino acid residues Asp^{70} and Asp^{480} seem to have a determinant role in substrate catalysis. Xgh74A apparently processes the xyloglucan chain in an endo way, unlike other member of the GH74 family, OXG-RCBH, which is an exoglucanase that releases two glycosyl residues from the reducing end of the xyloglucan polymer. The studies on Xgh74A presented in this work provide the first insights into the recognition and hydrolysis of xyloglucan by an endo-glycosidase and will support further developments in the field in the future.

Although xyloglucan is a major component of plant cell walls, CBMs that bind to this polysaccharide had not been identified previous to this work. In Chapter 3, two xyloglucan-interacting CBMs were characterized: *Ct*CBM30 and *Ct*CBM44, which bind with equal affinity to cellulose and xyloglucan. The unknown C-terminal module of *Ct*Cel9D-Cel44A (Ahsan *et al.*, 1996; Arai *et al.*, 2003) was revealed to constitute a CBM belonging to a novel CBM family (CBM44), with affinity for β-1,4-glucose polysaccharides, such as xyloglucan, CMC and Avicel. *Ct*CBM44 was the founder member of family 44. Crystal structures of *Ct*CBM30 and *Ct*CBM44 were determined revealing that these CBMs belong to the β-sandwich superfamily with Type B binding site topologies. In both cases, there is a platform of three tryptophans in the ligand-binding site that plays a pivotal role in accommodating the polysaccharide backbones. In addition, the orientation of the side-chains of the interacting tryptophans reflects the inherent helical structure of the polysaccharide allowing the decorations of xyloglucan to be exposed to the solvent. In these CBMs, the polar amino acid residues seem to play a less important role in ligand recognition than they usually do. Mutagenic studies confirmed the ligand recognition sites in the binding cleft for both *Ct*CBM30 and *Ct*CBM44 and also that the aromatic residues located on the edge of the cleft are pivotal to ligand recognition.

Some previous studies on CBMs that recognize the branched hemicelluloses xylan and galactomannan, indicated that the side-chains of decorated polysaccharides are usually exposed to the solvent and do not restrict ligand binding nor represent interacting specificity determinants. It is the case of *Cc*CBM17 from *Clostridium cellulovorans* (Notenboom *et al.*, 2001a), *Pe*CBM29-1 and *Pe*CBM29-2 from *Piromyces equi* (Charnock *et al.*, 2002; Flint *et al.*, 2005) and *Cm*CBM6-2 from *Cellvibrio mixtus* (Pires *et al.*, 2004). Therefore, it is suggested that Type B CBMs, due to their particular topology of the binding sites, possess, in general, the capacity to interact with similar affinities with decorated and undecorated polysaccharides. For molecules containing a backbone of β-1,4-linked undecorated glucose residues (cellulose) or decorated with extensive side-chains (xyloglucan) this was firstly confirmed with this work.

Like other GHs in family 44, Cel44A from *Ct*Cel9D-Cel44A displays mixed endo-β-1,4-glucanase/xylanase activity, therefore hydrolysing cellulosic substrates such as CMC and mixed linked β-1,4-β-1,3-glucans, while exhibiting reduced but still significant activity against oat spelt xylan. Data reported in this work show that Cel44A

is indeed a cellulase/xyloglucanase that displays some xylanase activity but it is not a true xylanase, as reported previously (Ahsan *et al.*, 1996). *Ct*CBM44 was found to potentiate Cel44A through a targeting effect, which means that *Ct*CBM44 potentiate the activity of Cel44A in the presence of the polysaccharides laminarin and pustulan. Therefore, the polysaccharides that do not act as substrates for Cel44A, can presumably make non-productive interactions with the active site of the enzyme.

*Ct*Lic26A-Cel5E (previously known as CelH) is a cellulosomal enzyme from *Clostridium thermocellum* that was described in Chapter 4. Several derivative recombinants were made and their catalytic properties assembled. A comprehensive kinetic characterization of both *Ct*Lic26A and *Ct*Cel5E was performed. *Ct*Lic26A revealed to be a classic mixed linkage β-1,4-β-1,3-glucanase that targets barley β-glucan and lichenan as its primary substrates. The enzyme showed no activity against β-1,4 or β-1,3 homopolymers of gluco- or manno-configured polysaccharides, like galactomannans, glucomannans, laminarin, Avicel or soluble cellulosic derivatives. Cel5E displays typical endo-β-1,4-glucanase activity, hydrolysing cellulosic substrates such as CMC, HEC and β-glucans. The family 11 CBM that is also present in this cellulosomal enzyme seems to potentiate the activity of Cel5E against the insoluble substrate Avicel. The three-dimensional structure of *Ct*Lic26A was solved in both native and inhibited forms, which contributed to understand the mechanism by which the catalytic process occurs in GHs family 26.

Biotechnological applications for CBMs were also further explored in this work. In Chapter 1, a review about current applications for CBMs was made and two novel applications were proposed. Chapter 5 described an experiment where the importance of a family 11 CBM as a supplement to a barley-based diet for broiler chicken was investigated. Diet supplementation with microbial cellulases and hemicellulases is well known to contribute to improve the nutritive value of cereals rich in soluble NSPs for poultry. However, there is a paucity of information concerning the importance of CBMs in potentiating the effects of cellulases and hemicellulases *in vivo*. Fontes *et al.* (2004) showed that a family 6 CBM from a *Clostridium thermocellum* xylanase improved the efficacy of the microbial recombinant xylanase, *in vivo*, when used to supplement wheat and rye-based diets for poultry. In that trial, animals supplemented with a bi-modular xylanase containing both a catalytic and a family 6 xylan-binding domain grew

significantly faster than animals fed on diets containing exclusively the xylanase catalytic domain.

The role of CBMs in the function of exogenous cellulases used to supplement barley-based diets for poultry remained, however, to be established. In Chapter 5, the family 11 CBM from *Clostridium thermocellum*'s *Ct*Lic26A-Cel5E, which displays high affinity for barley β-glucans, was evaluated. Two recombinant proteins, one containing the catalytic unit alone and the second containing the catalytic domain fused with the β-glucan specific CBM, were produced and used to supplement a barley-based diet for broiler chicken. The data showed that animals fed on diets supplemented with the recombinant *Ct*Lic26A-Cel5E modular derivatives or a commercial enzyme mixture displayed improved performance when compared with animals fed on diets not supplemented to exogenous enzymes. Together, the data suggest that individual recombinant cellulases could be as effective as complex mixtures of GHs in attenuating the depreciative effects of NSPs found in barley-based diets. There was, however, no significant difference between the recombinant enzymes containing or not the barley β-glucan specific CBM. It is possible that the enzyme dosage used in this study (30 U/kg of basal diet) was probably too high for the efficacy of *Ct*CBM11 to be noticed. Therefore, it remains to be established if the targeting effect resulting from the incorporation of CBMs in plant cell wall hydrolases could be effective at lower exogenous enzyme dosages and that is a challenge for future work. The capacity of the non-catalytic family 11 CBM to elicit the function of *Ct*Lic26A-Cel5E recombinant derivatives when incorporated at lower dosages rates in barley-based diets for poultry is currently under investigation in our research group.

In Chapter 6, a new approach for fixing bioactive molecules to cellulosic supports through their fusion to a CBM was presented. AMPs are molecules that possess the ability to fight infections and have been recognized as important components of the non-specific host defence system and innate immunity of all animal classes, including insects, amphibians and mammals. The mechanism of action of AMPs is currently a subject of extensive research since they may constitute an alternative to classic antimicrobial molecules, such as antibiotics. It has been suggested that AMPs may be more able to resist microorganisms' adaptation. The development of mechanisms for targeting bioactive molecules to a diversity of matrixes is an emerging theme in biochemistry and biotechnology. In Chapter 6, four different AMPs were fused to CipA CBM3, a non-hydrolytic structural protein responsible for the assembly of *Clostridium*

thermocellum's cellulosome. The major goal was to develop a biotechnological tool that would allow fixing a variety of polypeptides, such as AMPs, onto cellulosic surfaces, generating novel bio-products possessing antimicrobial properties. The recombinant fusion-proteins were successfully cloned and expressed in *Escherichia coli* and maintained the CBM3 high affinity for crystalline cellulose. Also the CBM3 itself was proved not to possess antimicrobial properties against Gram-negative bacteria. Testing the antimicrobial properties of the resulting recombinants proteins on several species of microorganisms in both soluble and fixed forms will allow the complete elucidation of the feasibility of the hypothesis advanced in this study.

CHAPTER 8 References

Abou-Hachem, M., Karlsson, E.N., Simpson, P.J., Linse, S., Sellers, P., Williamson, M.P., Jamieson, S.J., Gilbert, H.J., Bolam, D.N., and Holst, O. (2002) Calcium binding and thermostability of carbohydrate binding module CBM4-2 of Xyn10A from *Rhodothermus marinus*. *Biochemistry* 41, 5720-5729.

Ahsan, M.M., Kimura, T., Karita, S., Sakka, K., and Ohmiya K. (1996) Cloning, DNA sequencing, and expression of the gene encoding *Clostridium thermocellum* cellulase CelJ, the largest catalytic component of the cellulosome. *J. Bacteriol.* 178, 5732-5740.

Ali, E., Zhao, G., Sakka, M., Kimura, T., Ohmiya, K., and Sakka, K. (2005) Functions of family-22 carbohydrate-binding module in *Clostridium thermocellum* Xyn10C. *Biosci. Biotechnol. Biochem.* 69, 160-165.

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

Apajalahti, J., and Bedford, M.R. (1999) Improve bird performance by feeding its microflora. *W. Poult. Sci.* 15, 20-23.

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

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

Araki, T., Hashikawa, S., and Morishita, T. (2000) Cloning, sequencing, and expression in *Escherichia coli* of the new gene encoding β-1,3-xylanase from a marine bacterium, *Vibrio* sp strain XY-214. *Appl. Environ. Microbiol.* 66, 1741-1743.

Austin, S.C., Wiseman, J., and Chesson, A. (1999) Influence of non-starch polysaccharides structure on the metabolizable energy of UK wheat fed poultry. *J. Cereal Sci.* 29, 77-88.

Bauer, S., Vasu, P., Mort, A.J., and Somerville, C.R. (2005) Cloning, expression, and characterization of an oligoxyloglucan reducing end-specific xyloglucanobiohydrolase from *Aspergillus nidulans*. *Carbohydr. Res.* 340, 2590-2597.

Bayer, E.A., Kenig, R., and Lamed, R. (1983) Adherence of *Clostridium thermocellum* to cellulose. *J. Bacteriol.* 156, 818-827.

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

Bayer, E.A., and Lamed, R. (1986) Ultrastructure of the cell surface cellulose of *Clostridium thermocellum* and its interactions with cellulose. *J. Bacteriol.* 167, 828-836.

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

Bayer, E.A., Shimon, L.J., Shoham, Y., and Lamed, R. (1998) Cellulosomes-structure and ultrastructure. *J. Struct. Bio.* 124, 221-234.

Bedford, M.R. (2000) Exogenous enzymes in monogastric nutrition - their current value and future benefits. *Anim. Feed Sci. Technol.* 86, 1-13.

Bedford, M.R., Campbell, G.L., and Classen, H.L. (1991) The effect of pelleting, salt and pentosanase on the viscosity of intestinal contents and the performance of broiler fed rye. *Poult. Sci.* 70, 1571-1577.

Bedford, M.R., and Classen, A. (1992) Reduction of intestinal viscosity through manipulation of dietary rye and pentosanase concentration is affected throughout changes in carbohydrate composition of the intestinal aqueous phase and results in improved growth rate and food conversion efficiency of broiler chicks. *J. Nutr.* 122, 560-569.

Bedford, M.R., and Morgan, A.J. (1996) The use of enzymes in poultry diets. *W. Poult. Sci.* 52, 61-68.

Béguin, P., and Lemaire, M. (1996) The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. *Crit. Rev. Biochem. Mol. Biol.* 31, 201-236.

Beskrovnaya, O.I., Bukaniv, N.O., Donohue, L.C., Dackowski, W.R., Klinger, K.W., and Landes, G.M. (2000) Strong homophilic interactions of the Ig-like domains of polycystin-1, the protein product of an autosomal dominant polycystic kidney disease gene, PKD1. *Human Mol. Genet*. 9, 1641-1649.

Blondelle, S.E., and Houghten, R.A. (1992) Design of model amphipathic peptides having potent antimicrobial activities. *Biochemistry* 31, 12688-12694.

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

Bolam, D.N., Hughes, N., Virden, R., Lakey, J.H., Hazlewood, G.P., Henrissat, B., Braithwaite, K.L., and Gilbert, H.J. (1996) Mannanase A from *Pseudomonas fluorescens* spp cellulosa is a retaining glycosyl hydrolase in which E212 and E320 are the putative catalytic residues. *Biochemistry* 35, 16195-16204.

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

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

Boraston, A., McLean, B.W., Guarna, M.M., Amandaron-Akow, E., and Kilburn, D.G. (2001) A family 2a carbohydrate-binding module suitable as an affinity tag for proteins produced in *Pichia pastoris*. *Protein Expr. Purif.* 21, 417-423.

Boraston, A.B., Bolam, D.N., Gilbert, H.J., and Davies, G.J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* 382, 769-781.

Boraston, A.B., Chiu, P., Warren, R.A., and Kilburn, D.G. (2000a) Specificity and affinity of substrate binding by a family 17 carbohydrate-binding module from *Clostridium cellulovorans* cellulase 5A. *Biochemistry* 39, 11129-11136.

Boraston, A.B., Ghaffari, M., Warren, R.A., and Kilburn, D.G. (2002a) Identification and glucan-binding properties of a new carbohydrate-binding module family. *Biochem. J.* 361, 35-40.

Boraston, A.B., Kwan, E., Chiu, P., Warren, R.A., and Kilburn, D.G. (2003a) Recognition and hydrolysis of noncrystalline cellulose. *J. Biol. Chem.* 278, 6120-6127.

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

Boraston, A.B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D.R., Kilburn, D.G., and Davies, G.J. (2002c) Differential oligosaccharide recognition by evolutionarily-related β-1,4 and β-1,3 glucan-binding modules. *J. Mol. Biol.* 319, 1143-56.

Boraston, A.B., Revett, T.J., Boraston, C.M., Nurizzo, D., and Davies, G.J. (2003b) Structural and thermodynamic dissection of specific mannan recognition by a carbohydrate-binding module, *Tm*CBM27. *Structure* 11, 665-675.

Boraston, A.B., Tomme, P., Amandoron, E.A., and Kilburn, D.G. (2000b) A novel mechanism of xylan binding by a lectin-like module from *Streptomyces lividans* xylanase 10A. *Biochem. J.* 350, 933-941.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of protein utilizing the principles of protein-dye binding. *Anal. Biochem.* 72, 248-254.

Brenes, A., Marquardt, R.R., Guenter, W., and Viveros, A. (2002) Effect of enzyme addition on the performance and gastrointestinal tract size of chicks fed lupin seed and their fractions. *Poult. Sci.* 81, 670-8.

Brenes, A.M.S., Guener, W., and Marquardt, R.R. (1993) Effect of enzyme supplementation on the performance and digestive tract sue of broiler chickens fed wheat- and barley-based diets. *Poult. Sci.* 72, 1731-1739.

Brett, C.T., and Waldren, K. (1996) The molecular components of the wall. In Black, M., and Charlwood, B. (Eds.) *Physiology and Biochemistry of Plant Cell Walls* (pp.4-43) Chapman & Hall, London, UK.

Brogden, K.A. (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Rev. Microbiol.* 3, 238-250.

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

Brown, K.L., and Hancock, R.E.W. (2006) Cationic host defense (antimicrobial) peptides. *Curr. Opin. Immun.* 18, 24-30.

Brumer, H., Zhou, Q., Baumann, M.J., Carlsson, K., and Teeri, T.T. (2004) Activation of crystalline cellulose surfaces through the chemoenzymatic modification of xyloglucan. *J. Am. Chem. Soc.* 126, 5715-5721.

Brun, E., Moriaud, F., Gans, P., Blackledge, M.J., Barras, F., and Marion, D. (1997) Solution structure of the cellulose-binding domain of the endoglucanase Z secreted by *Erwinia chrysanthemi*. *Biochemistry* 36, 16074-16086.

Bueren, A.L., Morland, C., Gilbert, H.J., and Boraston, A.B. (2005) Family 6 carbohydrate-binding modules recognize the non-reducing end of beta-1,3-linked glucans by presenting a unique ligand binding surface. *J. Biol. Chem.* 280, 530-537.

Burmeister, W.P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997) The crystal structures of *Sinapis alba* myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. *Structure* 5, 663-675.

Bycroft, M., Bateman, A., Clarke, J., Hamill, S.J., Sandford, R., Thomas, R.L., and Chothia, C. (1999) The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease. *EMBO J.* 18, 297-305.

Carpita, N., and McCann, M. (2000) The cell wall. In Buchanan, B., Gruissem, W., and Jones, R. (Eds.), *Biochemistry and Molecular Biology of Plants*, John Wiley & Sons, Inc., Somerset, NJ.

Carvalho, A.L., Dias, F.M.V., Prates, J.A.M., Nagy, T., Gilbert, H.J., Davies, G.J., Ferreira, L.M.A., Romão, M.J., and Fontes, C.M.G.A. (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proc. Natl. Acad. Sci. USA* 100, 13809-13814.

Carvalho, A.L., Goyal, A., Prates, J.A., Bolam, D.N., Gilbert, H.J., Pires, V.M., Ferreira, L.M., Planas, A., Romão, M.J., and Fontes, C.M. (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. *J. Biol. Chem.* 279, 34785-34793.

Cash, H.L., Whitham, C.V., Behrendt, C.L., and Hooper, L.V. (2006) Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* 313, 1126-1130.

Chan, Y.R., and Gallo, R.L. (1998) PR-39, a syndecan-inducing antimicrobial peptide, binds and affects p130(cas). *J. Biol. Chem.* 273, 28978-28985.

Charnock, S.J., Bolam, D.N., Nurizzo, D., Szabo, L., McKie, V.A., Gilbert, H.J., and Davies, G.J. (2002) Promiscuity in ligand-binding: the three-dimensional structure of a *Piromyces* carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose. *Proc. Natl. Acad. Sci. USA* 99, 14077-14082.

Charnock, S.J., Bolam, D.N., Turkenburg, J.P., Gilbert, H.J., Ferreira, L.M., Davies, G.J., and Fontes, C.M. (2000) The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain. *Biochemistry* 39, 5013-5021.

Chatterjee, S., Chatterjee, D.K., Jani, R.H., Blumbach, J., Ganguli, B.N., Klesel, N., Limbert, M., and Seibert, G. (1992a) Mersacidin, a new antibiotic from *Bacillus*. *In vitro* and *in vivo* antibacterial activity. *J. Antibiot. (Tokyo)* 45, 839-845.

Chatterjee, S., Chatterjee, S., Lad, S.J., Phansalkar, M.S., Rupp, R.H., Ganguli, B.N., Fehlhaber, H.W., and Kogler, H. (1992b) Mersacidin, a new antibiotic from *Bacillus*. Fermentation, isolation, purification and chemical characterization. *J. Antibiot. (Tokyo)* 45, 832-838.

Chesson, A. (1993) Feed enzymes. *Anim. Feed Sci. Technol.* 45, 65-69.

Chhabra, S.R., and Kelly, R.M. (2002) Biochemical characterization of *Thermotoga maritima* endoglucanase Cel74 with and without a carbohydrate-binding module (CBM). *FEBS Lett.* 531, 375-380.

Choct, M. (1997) Feed non-starch polysaccharides: chemical structures and nutritional significance. *Feed Milling International*, June Issue, 13-26.

Čipáková, I., Gašperík, J., and Hostinová, E. (2006) Expression and purification of human antimicrobial peptide, dermcidin, in *Escherichia coli*. *Protein Expr. Purif.* 45, 269-274.

Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* D50, 760-763.

Cosgrove, D.J. (2001) Wall structure and wall loosening. A look backwards and forwards. *Plant Physiol.* 125, 131-134.

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

Coughlan, M.P. (1985) The properties of fungal and bacterial cellulases with comment of their production and application. In Russel, G.E. (Ed.), *Biotechnology and Genetic Engineering Reviews* (Vol. 3, pp.39-109) Intercept, UK.

Coutinho, P.M., and Henrissat, B. (1999) Carbohydrate-active enzymes: an integrated database approach. In Gilbert, H.J., Davies, G.J., Svensson, B., and Henrissat, B. (Eds.), *Recent Advances in Carbohydrate Engineering* (pp.3-12) Royal Society of Chemistry, Cambridge, UK.

Coutinho, P.M., and Henrissat B. (2008) Subject: CAZy – Carbohydrate-Active EnZYmes. http://afmb.cnrs-mrs.fr/CAZY.

Creagh, A.L., Ong, E., Jervis, E., Kilburn, D.G., and Haynes, C.A. (1996) Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. *Proc. Natl. Acad. Sci. USA* 93, 12229-12234.

Czjzek, M., Bolam, D.N., Mosbah, A., Allouch, J., Fontes, C.M., Ferreira, L.M., Bornet, O., Zamboni, V., Darbon, H., Smith, N.L., Black, G.W., Henrissat, B., and 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. *J. Biol. Chem.* 276, 48580-48587.

Davies, G.J., Ducros, V.M.A., Varrot, A., and Zechel, D.L. (2003) Mapping the conformational itinerary of beta-glycosidases by X-ray crystallography. *Biochem. Soc. Trans.* 31, 523-527.

Davies, G.J., Gloster, T.M., and Henrissat, B. (2005) Recent structural insights into the expanding world of carbohydrate-active enzymes. *Curr. Op. Struct. Biol.* 15, 637-645.

Davies, G.J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A.M., Schulein, M., and Withers, S.G. (1998) Snapshots along an enzymatic reaction coordinate: analysis of a retaining beta-glycoside hydrolase. *Biochemistry* 37, 11707-11713.

Davies, G.J., Sinnott, M.L., and Withers, S.G. (1997a) Glycosyl Transfer. In Sinnott, M.L. (Ed.), *Comprehensive Biological Catalysis* (Vol. 1, pp.119-209) Academic Press, London, UK.

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

Demain, A.L., Newcomb, M., and Wu, J.H. (2005) Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Rev.* 69, 124-154.

Dias, F.M.V., Vincent, F., Pell, G., Prates, J.A.M., Centeno, M.S.J., Tailford, L.E., Ferreira, L.M.A., Fontes, C.M.G.A., Davies, G.J., and Gilbert, H.J. (2004) Insights into the molecular determinants of substrate specificity in glycoside hydrolase family 5 by the crystal structure and kinetics of *Cellvibrio mixtus* mannosidase 5A. *J. Biol. Chem.* 279, 25517-25526.

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

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

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

Ding, S.-Y., Rincon, M.T., Lamed, R., Martin, J.C., McCrae, S.I., Aurilia, V., Shoham, Y., Bayer, E.A., and Flint, H.J. (2001) Cellulosomal scaffolding-like proteins from *Ruminococcus flavefaciens*. *J. Bacteriol.* 183, 1945-1953.

Doheny, J.G., Jervis, E.J., Guarna, M.M., Humphries, R.K., Warren, R.A., and Kilburn, D.G. (1999) Cellulose as an inert matrix for presenting cytokines to target cells: production and properties of a stem cell factor-cellulose-binding domain fusion protein. *Biochem. J.* 339, 429-343.

Doi, R.H., and Kosugi, A. (2004) Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nature Rev. Microbiol.* 2, 541-551.

Doi, R.H., Kosugi, A., Murashima, K., Tamaru, Y., and Han, S.O. (2003) Cellulosomes from mesophilic bacteria. *J. Bacteriol.* 185, 5907-5914.

Ducros, V., Zechel, D.L., Murshudov, G.N., Gilbert, H.J., Szabo, L., Stoll, D., Withers, S.G., and Davies, G.J. (2002) Substrate distortion by a beta-mannanase: snapshots of the Michaelis and covalent-intermediate complexes suggest a B(2,5) conformation for the transition state. *Angew. Chem. Int. Ed.* 41, 2824-2827.

Elsbach, P. (2003) What is the real role of antimicrobial polypeptides that can mediate several other inflammatory responses? *J. Clin. Investig.* 111, 1643-1645.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr.* D60, 2126-2132.

Esnouf, R.M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics Model* 15, 132-134.

Falla, T.J., Karunaratne, D.N., and Hancock, R.E.W. (1996) Mode of action of the antimicrobial peptide indolicidin. *J. Biol. Chem.* 271, 19298-19303.

Fengler, A.I., and Marquardt, R.R. (1988) Water-soluble arabinoxylans from rye. II. Effects of rate of dialysis and on the retention of nutrient by the chick. *Cereal Chem.* 65, 298-302.

Fernandes, A.C.G., Fontes, C.M.G.A., Gilbert, H.J., Hazelwood, G.P., Fernandes, T.H., and Ferreira, L.M.A. (1999) Homologous xylanases from *Clostridium thermocellum*: evidence for bi-funcional activity, synergism between xylanases catalytic modules and the presence of xylan-binding domains in enzyme complexes. *Biochem. J.* 342, 105-110.

Finnegan, P.M., Brumbley, S.M., O'Shea, M.G., Nevalainen, H., and Bergquist, P.L. (2005) Diverse dextranase genes from *Paenibacillus* species. *Arch. Microbiol.* 183, 140-147.

Flint, J., Bolam, D.N., Nurizzo, D., Taylor, E.J., Williamson, M.P., Walters, C., Davies, G.J., and Gilbert, H.J. (2005) Probing the mechanism of ligand recognition in family 29 carbohydrate-binding modules. *J. Biol. Chem.* 280, 23718-23726.

Flint, J., Nurizzo, D., Harding, S.E., Longman, E., Davies, G.J., Gilbert, H.J., and Bolam, D.N. (2004) Ligand-mediated dimerization of a carbohydrate-binding molecule reveals a novel mechanism for protein-carbohydrate recognition. *J. Mol. Biol.* 337, 417–426.

Fontes, C.M., Hazlewood, G.P., Morag, E., Hall, J., Hirst, B.H., and Gilbert, H.J. (1995) Evidence for a general role for non-catalytic thermostabilizing domains in xylanases from thermophilic bacteria. *Biochem. J.* 307, 151-158.

Fontes, C.M.G.A., Ponte, P.I.P., Reis, T.C., Soares, M.C., Gama, L.T., Dias, F.M.V., and 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. *Br. Poult. Sci.* 45, 648-656.

Fry, S.C. (1989) Cellulases, hemicelluloses and auxin-stimulated growth: a possible relationship. *Physiol. Plant.* 75, 532-536.

Fry, S.C., York, W.S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J.P., Kato, Y., Lorences, E.P., Maclachlan, G.A., McNeil, M., Mort, A.J., Reid, J.S.G., Seitz, H.U., Selvendran, R.R., Voragen, A.G.J., and White, A.R. (1993) An unambiguous nomenclature for xyloglucan-derived oligosaccharides. *Physiol. Plant.* 89, 1-3.

Garcia-Martinez, D.V., Shinmyo, A., Madia, A., and Demain, A.L. (1980) Studies on cellulase production by *Clostridium thermocellum*. *Eur. J. Appl. Microbiol. Biotechnol.* 9, 189-197.

Garcia-Olmedo, F., Molina, A., Alamillo, J.M., and Rodriguez-Palenzuela, P. (1998) Plant defense peptides. *Biopolymers* 47, 479-491.

Gaskell, A., Crennell, S., and Taylor, G. (1995) The three domains of a bacterial sialidase: a β-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* 3, 1197-1205.

Gilbert, H.J., Bolam, D.N., Szabo, L., Xie, H., Williamson, M.P., Simpson, P.J., Jamal, S., Boraston, A.B., Kilburn, D.G., and Warren, R.A.J. (2002) An update on Carbohydrate-Binding Modules, in Teeri, T.T, Svensson, B., Gilbert, H.J., and Feizi, T. (eds.) *Carbohydrate bioengineering*, pp. 89-98, Royal Society of Chemistry, London, UK.

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

Gill, J., Rixon, J.E., Bolam, D.N., McQueen-Mason, S., Simpson, P.J., Williamson, M.P., Hazlewood, G.P., and Gilbert, H.J. (1999) The type II and X cellulose-binding domains of *Pseudomonas* xylanase A potentiate catalytic activity against complex substrates by a common mechanism. *Biochem. J.* 342, 473-480.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15, 305-308.

Guerreiro, C.I., Najmudin, S., Carvalho, A.L., Prates, J.A., Correia, M.A., Alves, V.D., Ferreira, L.M., Romão, M.J., Gilbert, H.J., Bolam, D.N., and Fontes, C.M. (2006) Xyloglucan is recognized by carbohydrate-binding modules that interact with beta-glucan chains. *J. Biol. Chem.* 281, 8815-28.

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

Hall, J., Black, G.W., Ferreira, L.M., Millward-Sadler, S.J., Ali, B.R., Hazlewood, G.P., and Gilbert, H.J. (1995) The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas fluorescens* subsp *cellulosa* is important for the efficient hydrolysis of Avicel. *Biochem. J.* 309, 749-756.

Hancock, R.E.W. (1997) Peptide antibiotics. *Lancet* 349, 418-422.

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

Hatada, Y., Hidaka, Y., Nogi, Y., Uchimura, K., Katayama, K., Li, Z., Akita, M., Ohta, Y., Goda, S., Ito, H., Matsui, H., Ito, S., and Horikoshi, K. (2004) Hyper-production of an isomalto-dextranase of an *Arthrobacter* sp by a proteases-deficient *Bacillus subtilis*: sequencing, properties, and crystallization of the recombinant enzyme. *Appl. Microbiol. Biotechnol.* 65, 583-592.

Hayashi, T. (1989) Xyloglucans in the primary cell wall. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 40, 139-168.

Haynie, S.L., Crum, G.A., and Doele, B.A. (1995) Antimicrobial activities of amphiphilic peptides covalently bonded to a water-insoluble resin. *Antimicrob. Agents Chemother.* 39, 301-307.

Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280, 309-316.

Henrissat, B. (1998) Glycosidase families. *Biochem. Soc. Trans.* 2, 153-156.

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

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P., and Davies, G. (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. USA*. 92, 7090-7094.

Henrissat, B., Teeri, T.T., and Warren, R.A. (1998) A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. *FEBS Lett.* 425, 352-354.

Henshaw, J.L., Bolam, D.N., Pires, V.M., Czjzek, M., Henrissat, B., Ferreira, L.M.A., Fontes, C.M.G.A., and Gilbert, H.J. (2004) The family 6 carbohydrate-binding module *Cm*CBM6-2 contains two ligand-binding sites with distinct specificities. *J. Biol. Chem.* 279, 21552-21559.

Hespell, R., and O'Bryan, P.J. (1992) Purification and characterization of an α-L-arabinofuranosidase from *Butyrivibrio fibrisolvens* GS113. *Appl. Environ. Microbiol.* 58, 1082-1088.

Hesselman, K., and Aman, P. (1986) The effect of β-glucanase on the utilisation of starch and nitrogen by broiler chickens fed on barley of low and high viscosity. *Anim. Feed Sci. Technol.* 15, 83-93.

Hogg, D., Woo, E.-J., Bolam, D.N., McKie, V.A., Gilbert, H.J., and Pickersgill, R.W. (2001) Crystal structure of mannanase 26A from *Pseudomonas cellulosa* and analysis of residues involved in substrate binding. *J. Biol. Chem.* 276, 31186-31192.

Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233, 123-138.

Hwang, S., Ahn, J., Lee, S., Lee, T.G., Haam, S., Lee, K., Ahn, I.-S., and Jung, J.-K. (2004) Evaluation of cellulose-binding domain fused to a lipase for the lipase immobilization. *Biotechnol. Lett.* 26, 603-605.

Irwin, D.C., Cheng, M., Xiang, B., Rose, J.K., and Wilson, D.B. (2003) Cloning, expression and characterization of a family-74 xyloglucanase from *Thermobifida fusca*. *Eur. J. Biochem.* 270, 3083-3091.

IUB (1984) Enzyme Nomenclature: Recommendations of the Nomenclature Committee of the International Union of Biochemistry on the Nomenclature and Classification of Enzyme-Catalysed Reactions, Academic Press, London and New York.

Jamal, S., Nurizzo, D., Boraston, A.B. and Davies, G.J. (2004) X-ray crystal structure of a non-crystalline cellulose-specific carbohydrate-binding module: CBM28. *J. Mol. Biol.* 339, 253-258.

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

Jancarik, J., and Kim, S.H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* 24, 409-411.

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

Jenssen, H., Hamill, P., and Hancock, R.E.W. (2006) Peptide antimicrobial agents. *Clin. Microbiol. Rev.* 19, 491-511.

Jervis, E.J., Haynes, C.A., and Kilburn, D.G. (1997) Surface diffusion of cellulases and their isolated binding domains on cellulose. *J. Biol. Chem.* 272, 24016-24023.

Johnson, P.E., Tomme, P., Joshi, M.D., and McIntosh L.P. (1996) Interaction of soluble cellooligosaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC2. NMR and ultraviolet absorption spectroscopy. *Biochemistry* 35, 13895-13906.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* A47, 110-119.

Józefiak, D., Rutkowski, A., Jensen, B.B., and Engberg, R.M. (2006) The effect of *β*-glucanase supplementation of barley- and oat-based diets on growth performance and fermentation in broiler chicken gastrointestinal tract. *Br. Poultry Sci.* 47, 57-64.

Kabsch, W. (1988a) Automatic indexing of rotation diffraction patterns. *J. Appl. Cryst*. 21, 67-72.

Kabsch, W. (1988b) Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl.Cryst.* 21, 916-924.

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

Kang, J.H., Shin, S.Y., Jang, S.Y., Kim, K.L., and Hahm, K.-S. (1999) Effects of tryptophan residues of porcine myeloid antibacterial peptide PMAP-23 on antibiotic activity. *Biochem. Biophys. Res. Commun.* 264, 281-286.

Kato, Y., Matsushita, J., Kubodera, T., and Matsuda, K. (1985) A novel enzyme producing isoprimeverose from oligoxyloglucans of *Aspergillus oryzae*. *J. Biochem.* 97, 801-810.

Kauffmann, C., Shoseyov, O., Shpigel, E., Bayer, E.A., Lamed, R., Shoham, Y., and Mandelbaum, R.T. (2000) A novel methodology for enzymatic removal of atrazine from water by CBD-fusion protein immobilized on cellulose. *Environ. Sci. Technol.* 34, 1292-1296.

Kavoosi, M., Meijer, J., Kwan, E., Creagh, A.L., Kilburn, D.G., and Haynes, C.A. (2004) Inexpensive one-step purification of polypeptides expressed in *Escherichia coli* as fusions with the family 9 carbohydrate-binding module of xylanase 10A from *T. maritima*. *J. Chromatogr. B* 807, 87-94.

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

Kim, H., Goto, M., Jeong, H.J., Jung, K.H., Kwon, I., and Furukawa, K. (1998) Functional analysis of a hybrid endoglucanase of bacterial origin having a cellulose-binding domain from a fungal exoglucanase. *Appl. Biochem. Biotechnol.* 75, 193-204.

Klaenhammer, T.R. (1988) Bacteriocins of lactic acid bacteria. *Biochimie* 70, 337-349.

Klocke, M., Mundt, K., Idler, F., Jung, S., and Backhausen, J.E. (2005) Heterologous expression of enterocin A, a bacteriocin from *Enterococcus faecium*, fused to a cellulose-binding domain in *Escherichia coli* results in a functional protein with inhibitory activity against *Listeria*. *Appl. Microbiol. Biotechnol.* 67, 532-538.

Knudsen, K.E.B. (1997) Carbohydrate and lignin contents of plant materials used in animal feeding. *Anim. Feed Sci. Technol.* 67, 319-338.

Koczulla, R., von Degenfeld, G., Kupatt, C., Krotz, F., Zahler, S., Gloe, T., Issbrucker, K., Unterberger, P., Zaiou, M., Lebherz, C., Karl, A., Raake, P., Pfosser, A., Boekstegers, P., Welsch, U., Hiemstra, P.S., Vogelmeier, C., Gallo, R.L., Clauss, M., and Bals, R. (2003) An angiogenic role for the human peptide antibiotic LL-37/hCAP-18. *J. Clin. Investig.* 111, 1665-1672.

Kormos, J., Johnson, P.E., Brun, E., Tomme, P., McIntosh L.P., Haynes, C.A., and Kilburn, D.G. (2000) Binding site analysis of cellulose-binding domain CBD(N1) from endoglucanase C of *Cellulomonas fimi* by site-directed mutagenesis. *Biochemistry* 39, 8844-8852.

Kraulis, J., Clore, G.M., Nilges, M., Jones, T.A., Petterson, G., Knowles, J., and Gronenborn, A.M. (1989) Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* 28, 7241-7257.

Kraulis, P.J. (1991) *MOLSCRIPT*: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* 24, 946-950.

Kroon, P.A., Williamson, G., Fish, N.M., Archer, D.B., and Belshaw, N.J. (2000) A modular esterase from *Penicillium funiculosum* which releases ferulic acid from plant cell walls and binds crystalline cellulose contains a carbohydrate-binding module. *Eur. J. Biochem.* 267, 6740-6752.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

Lamed, R., Naimark, J., Morgenstern, E., and Bayer, E.A. (1987) Specialized cell surface structures in cellulolytic bacteria. *J. Bacteriol.* 169, 3792-3800.

Lamed, R., and Zeikus, J.G. (1980) Ethanol production by thermophilic bacteria: relationship between fermentation product yields of and catabolic enzyme activities in *Clostridium thermocellum* and *Thermoanaerobium brockii*. *J. Bacteriol.* 144, 569-578.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993) *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283-291.

Lee, D.G., Kim, P.I., Park, Y., Woo, E.-R., Choi, J.S., Choi, C.-H., and Hahm, K.-S. (2002) Design of novel peptide analogs with potent fungicidal activity, based on PMAP-23 antimicrobial peptide isolated from porcine myeloid. *Biochem. Biophys. Res. Commun.* 293, 231-238.

Lee, I., Evans, B.R., and Woodward, J. (2000) The mechanism of cellulase action on cotton fibers: evidence from atomic force microscopy. *Ultramicroscopy* 82, 213-221.

Lehrer, R.I., and Ganz, T. (2002) Cathelicidins: a family of endogenous antimicrobial peptides. *Curr. Opin. Hematol.* 9, 18-22.

Lehtiö, J. (2001) Functional studies and engineering of family 1 carbohydrate-binding modules. PhD-thesis. Stockholm, Sweden: Royal Institute of Technology (http://www.lib.kth.se/Sammanfattningar/lehtio010914.pdf).

Lehtiö, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., and Teeri, T.T. (2003) The binding specificity and affinity determinants of family 1 and family 3 cellulose-binding domains. *Proc. Natl. Acad. Sci. USA* 100, 484-489.

Leslie, A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data, in *Joint CCP4 and ESF-EACMB newsletter on protein crystallography* Vol. 26, pp. 27-33. Daresbury Laboratory, Warrington, UK.

Levy, I., and Shoseyov, O. (2002) Cellulose-binding domains. Biotechnological applications. *Biotechnol. Adv.* 20, 191-213.

Lewis, W., Keshavarz-Moore, E., Windust, J., Bushell, D., and Parry, N. (2006) Construction and evaluation of novel fusion proteins for targeted delivery of micro particles to cellulose surfaces. *Biotechnol. Bioeng.* 94, 625-632.

Lillelund, V.H., Liu, H.Z., Liang, X.F., Søhoel, H., and Bols, M. (2003) Isofagomine lactams, synthesis and enzyme inhibition. *Org. Biomol. Chem.* 1, 282-287.

Linder, M., and Teeri, T.T. (1997) The roles and function of cellulose-binding domains. *J. Biotechnol.* 57, 15-28.

Lovell, S.C., Davis, I.W., Adrendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Prot. Struct. Funct. Genet.* 50, 437-450.

Lymar, E.S., Li, B., and Renganathan, V. (1995) Purification and characterization of a cellulose-binding β-glucosidase from cellulose-degrading cultures of *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* 61, 2976-2980.

Macdonald, J.M., Stick, R.V., Tilbrook, D.M.G., and Withers, S.G. (2002) Synthesis with glycosynthases: cello-oligomers of isofagomine and a tetrahydrooxazine as cellulase inhibitors. *Aust. J. Chem.* 55, 747-752.

Macdonald, J.M., Hrmova, M., Fincher, G.B., and Stick, R.V. (2004) The synthesis of 3-O-(β-D-glucopyranosyl)- and 3-O-(β-laminaribiosyl)-isofagomines, potent inhibitors of a 1,3-β-D-glucan endo-hydrolase. *Aust. J. Chem.* 57, 187-191.

Malet, C., and Planas, A. (1997) Mechanism of *Bacillus* 1,3-1,4-β-D-glucan 4-glucanohydrolases: kinetics and pH studies with 4-methylumbelliferyl β-D-glucan oligosaccharides. *Biochemistry* 36, 13838-13848.

Margolles-Clark, E., Tenkanen, M., Söderlund, H., and Penttilä, M. (1996) Acetyl xylan esterase from *Trichoderma reesei* contains an active-site serine residue and a cellulose-binding domain. *Eur. J. Biochem.* 237, 553-560.

Mark, B.L., Vocadlo, D.J., Zhao, D.L., Knapp, S., Withers, S.G., and James, M.N.G. (2001) Biochemical and structural assessment of the 1-N-azasugar GalNAc-isofagomine as a potent family 20 beta-N-acetylhexosaminidase inhibitor. *J. Biol. Chem.* 276, 42131-42137.

Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* 33, 491-497.

Maurice, S., Dekel, M., Shoseyov, O., and Gertler, A. (2003) Cellulose beads bound to cellulose-binding domain-fused recombinant proteins; an adjuvant system for parenteral vaccination of fish. *Vaccine* 21, 3200-3207.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr.* D61, 458-464.

McGreal, E.P., Martinez-Pomares, L., and Gordon S. (2004) Divergent roles for C-type lectins expressed by cells of the innate immune system. *Molec. Immun.* 41, 1109-1121.

McLean, B.W., Bray, M.R., Boraston, A.B., Gilkes, N.R., Haynes, C.A., and Kilburn, D.G. (2000) Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. *Protein Eng.* 13, 801-809.

Meinke, A., Damude, H.G., Tomme, P., Kwan, E., Kilburn, D.G., Miller, R.C., Jr., Warren, R.A.J., and Gilkes, N.R. (1995) Enhancement of the endo-β-1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop. *J. Biol. Chem.* 270, 4383-4386.

Merritt, E.A., and Bacon, D.J. (1997) Raster3D: Photorealistic molecular graphics. *Methods Enzymol*. 277, 505-524.

Milkowski, K., Clark, J.H., and Doi, S. (2004) New materials based on renewable resources: chemically modified highly porous starches and their composites with synthetic monomers. *Green Chem*. 6, 189-190.

Miron, J., Ben-Ghedalia, D., and Morrison, M. (2001) Adhesion mechanisms of rumen cellulolytic bacteria. *J. Dairy Sci.* 84, 1294-1309.

Miyanaga, A., Koseki, T., Matsuzawa, H., Wakagi, T., Shoun, H., and Fushinobu, S. (2004) Crystal structure of a family 54 α-L-arabinofuranosidase reveals a novel carbohydrate-binding module that can bind arabinose. *J. Biol. Chem.* 279, 44907-44914.

Miyazaki, S., Suisha, F., Kawasaki, N., Shirakawa, M., Yamatoya, K., and Attwood, D. (1998) Thermally reversible xyloglucan gels as vehicles for rectal drug delivery. *J. Control. Release.* 56, 75-83.

Mohand-Oussaid, O., Payot, S., Guedon E., Gelhaye, E., Youyou, A., and Petitdemange, H. (1999) The extracellular xylan degradative system in *Clostridium cellulolyticum* cultivated on xylan: evidence for cell-free cellulosome production. *J. Bacteriol.* 181, 4035-4040.

Moon, J.-Y., Henzler-Wildman, K.A., and Ramamoorthy, A. (2006) Expression and purification of a recombinant LL-37 from *Escherichia coli*. *Biochim. Biophys. Acta* 1711, 1351-1358.

Morag, E., Bayer, E.A., Lamed, R. (1990) Relationship of cellulosomal and noncellulosomal xylanases of *Clostridium thermocellum* to cellulose-degrading enzymes. *J. Bacteriol.* 172, 6098-6105.

Morassutti, C., De Amicis, F., Bandiera, A., and Marchetti, S. (2005) Expression of SMAP-23 cathelicidin-like peptide in bacterial cells by intein-mediated system. *Protein Expr. Purif.* 39, 160-168.

Mourão, J.L., Ponte, P.I.P., Prates, J.A.M., Centeno, M.S.J., Ferreira, L.M.A., Soares, M.A.C., and Fontes, C.M.G.A. (2006) Use of β-glucanases and β-1,4-xylanases to supplement diets containing alfalfa and rye for laying hens: effects on bird performance and egg quality. *J. Appl. Poult. Res.* 15, 256-265.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* D53, 240-255.

Najmudin, S., Guerreiro, C.I., Ferreira, L.M.A., Romão, M.J., Fontes, C.M.G.A., and Prates, J.A.M. (2005) Overexpression, purification and crystallization of the two C-terminal domains of the bifunctional cellulase *Ct*Cel9D-Cel44A from *Clostridium thermocellum*. *Acta Crystallogr.* F61, 1043-1045.

National Research Council (1994) Nutrient requirements of poultry. 9th rev., ed. Natl. Acad. Press, Washington, DC.

Navaza, J., and Saludjian, P. (1997) AMoRe: An automated molecular replacement program package. *Methods Enzymol.* A276, 581-594.

Nelson, N. (1944) A photometric adaptation of the Somogyi method for the determination of glucose. *J. Biol. Chem.* 153, 375-380.

Notenboom, V., Boraston, A.B., Chiu, P., Freelove, A.C., Kilburn, D.G., and Rose, D.R. (2001a) Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. *J. Mol. Biol.* 314, 797-806.

Notenboom, V., Boraston, A.B., Kilburn, D.G., and Rose, D.R. (2001b) Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. *Biochemistry* 40, 6248-6256.

Notenboom, V., Boraston, A.B., Williams, S.J., Kilburn, D.G., and Rose, D.R. (2002) High-resolution crystal structures of the lectin-like xylan-binding domain from *Streptomyces lividans* xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry* 41, 4246-4254.

Notenboom, V., Williams, S.J., Hoos, R., Withers, S.G., and Rose, D.R. (2000) Detailed structural analysis of glycosidase/inhibitor interactions: complexes of Cex from *Cellulomonas fimi* with xylobiose-derived aza-sugars. *Biochemistry* 39, 11553-11563.

Numao, S., Kuntz, D.A., Withers, S.G., and Rose, D.R. (2003) Insights into the mechanism of *Drosophila melanogaster* Golgi alpha-mannosidase II through the structural analysis of covalent reaction intermediates. *J. Biol. Chem.* 278, 48074-48083.

Ohara, H., Karita, S., Kimura, T., Sakka, K., and Ohmiya, K. (2000) Characterization of the cellulolytic complex (cellulosome) from *Ruminococcus albus*. *Biosci. Biotechnol. Biochem.* 64, 254-260.

Okazaki, F., Tamaru, Y., Hashikawa, S., Li, Y.T., and Araki, T. (2002) Novel carbohydrate-binding module of β-1,3-xylanase from a marine bacterium, *Alcaligenes* sp Strain XY-234. *J. Bacteriol.* 184, 2399-2403.

Orikoshi, H., Nakayama, S., Hanato, C., Miyamoto, K., and Tsujibo H. (2005) Role of the N-terminal polycystic kidney disease domain in chitin degradation by chitinase A from a marine bacterium, *Alteromonas* sp strain O-7. *J. Appl. Microbiol*. 99, 551-557.

O´Sullivan, A.C. (1997) Cellulose: the structure slowly unravels. *Cellulose* 4, 173-207.

Otomo, T., Teruya, K., Uegaki, K., Yamazaki, T., and Kyogoku, Y. (1999) Improved segmental isotope labelling of proteins and application to a larger protein. *J. Biomol. NMR* 14, 105-114.

Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, in Carter, C. Jr., and Sweet, R. (eds.) M*acromolecular Crystallography,* part A, vol. 276, pp. 307-326, Academic Press Inc., San Diego, USA.

Pagès, S., Bélaïch, A., Bélaïch, J.-P., Morag, E., Lamed, R., Shoham, Y., and 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, 517-527.

Pagès, S., Gal, L., Bélaïch, A., Gaudin, C., Tardif, C., and Bélaïch, J.-P. (1997) Role of scaffolding protein CipC of *Clostridium cellulolyticum* in cellulose degradation. *J. Bacteriol.* 179, 2810-2816.

Pala, H., Lemos, M.A., Mota, M., and Gama, F.M. (2001) Enzymatic upgrade of old paperboard containers. *Enzyme Microb. Technol.* 29, 274-279.

Pell, G., Williamson, M.P., Walters, C., Du, H., Gilbert, H.J., and Bolam, D.N. (2003) Importance of hydrophobic and polar residues in ligand binding in the family 15 carbohydrate-binding module from *Cellvibrio japonicus* Xyn10C. *Biochemistry* 42, 9316-9323.

Peng, L., Xu, Z., Fang, X., Wang, F., and Cen, P. (2004) High-level expression of soluble human β-defensin-2 in *Escherichia coli*. *Process Biochem.* 39, 2199-2205.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* 6, 458-463.

Petersen, S.T., Wiseman, J., and Bedford, M.R. (1993) The effect of age and diet on the viscosity of intestinal contents in broiler chicks. *Proceedings of the 10th Meeting of the British Society of Animal Production*, p. 434, Scarborough, UK.

Philip, J.S., Gilbert, H.J., and Smithard, R.R. (1995) Growth, viscosity and beta-glucanase activity of intestinal fluid in broiler chickens fed on barley-based diets with or without exogenous beta-glucanase. *Br. Poult. Sci.* 36, 599-603.

Pires, V.M., Henshaw, J.L., Prates, J.A., Bolam, D.N., Ferreira, L.M., Fontes, C.M., Henrissat, B., Planas, A., Gilbert, H.J., and Czjzek, M. (2004) The crystal structure of the family 6 carbohydrate-binding module from *Cellvibrio mixtus* endoglucanase 5a in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. *J. Biol. Chem.* 279, 21560-21568.

Pohlschröder, M., Canale-Parola, E., and Leschine, S.B. (1995) Ultrastructure diversity of the cellulase complexes of *Clostridium papyrosolvens* C7. *J. Bacteriol.* 177, 6625-6629.

Ponte, P.I.P., Ferreira, L.M.A., Soares, M.A.C, Gama, L.T., and Fontes, C.M.G.A. (2004) Xylanase inhibitors affect the action of exogenous enzymes used to supplement *Triticum durum*-based diets for broiler chicks. *J. Appl. Poult. Res.* 13, 660-666.

Ponte, P.I.P., Lordelo, M.M.S., Guerreiro, C.I.P.D., Soares, M.C., Mourão, J.L., Crespo, J.P., Crespo, D.G., Prates, J.A.M., Ferreira, L.M.A., and Fontes, C.M.G.A. (2008) Crop beta-glucanase activity limits the effectiveness of a recombinant cellulase used to supplement a barley-based feed for free-range broilers. *Br. Poult. Sci.* Accepted for publication.

Raghothama, S., Simpson, P.J., Szabo, L., Nagy, T., Gilbert, H.J., and Williamson, M.P. (2000) Solution structure of the CBM10 cellulose-binding module from *Pseudomonas* xylanase A. *Biochemistry* 39, 978-984.

Rechter, M., Lider, L., Cahalon, L., Baharav, E., Dekel, M., Seigel, D., Vlodavsky, I., Aingorn, H., Cohen, I.R., and Shoseyov, O. (1999) A cellulose-binding domain-fused recombinant human T-cell connective tissue-activating peptide-III manifests heparanase activity. *Biochem. Biophys. Res. Commun.* 255, 657-662.

Reiter, W.D. (2002) Biosynthesis and properties of the plant cell wall. *Curr. Opin. Plant Biol.* 5, 536-542.

Richins, R.D., Mulchandani, A., and Chen, W. (2000) Expression, immobilization, and enzymatic characterization of cellulose-binding domain-organophosphorus hydrolase fusion enzymes. *Biotechnol. Bioeng.* 69, 591-596.

Riley, M.A. (1998) Molecular mechanisms of bacteriocin evolution. *Annu. Rev. Genet.* 32, 255-278.

Rinaldi, A.C. (2002) Antimicrobial peptides from amphibian skin: an expanding scenario. *Curr. Opin. Chem. Biol.* 6, 799-804.

Rincon, M.T., Ding, S.-Y., McCrae, S.I., Martin, J.C., Aurilia, V., Lamed, R., Shoham, Y., Bayer, E.A., and Flint, H.J. (2003) Novel organization and divergent dockerin specificities in the cellulosome system of *Ruminococcus flavefaciens*. *J. Bacteriol.* 185, 703-713.

Roussel, A., and Cambillau, C. (1991). TURBO-FRODO. Silicon Graphics Geometry Partners Directory, p. 86, Silicon Graphics, Mountain View, California, USA.

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

SAS Institute (2004) SAS User´s Guide: Statistics. Version 8 Edition, SAS Institute Inc., Cary NC.

Saul, F.A., Rovira, P., Boulot, G., Damme, E.J., Peumans, W.J., Truffa-Bachi, P., and Bentley, G.A. (2000) Crystal structure of *Urtica dioica* agglutinin, a superantigen presented by MHC molecules of class I and class II. *Structure Fold. Des.* 8, 593-603.

Schärpf, M., Connely, G.P., Lee, G.M., Boraston, A.B., Warren, R.A.J., and McIntosh, L.P. (2002) Site-specific characterization of the association of xylooligosaccharides with the CBM13 lectin-like xylan-binding domain from *Streptomyces lividans* xylanase 10A by NMR spectroscopy. *Biochemistry* 41, 4255-4263.

Scheible, W.R., and Pauly, M. (2004) Glycosyltransferases and cell wall biosynthesis: novel players and insights. *Curr. Opin. Plant Biol.* 7, 285-295.

Selsted, M.E., Novotny, M.J., Morris, W.L., Tang, Y.-Q., Smith, W., and Cullor, J.S. (1992) Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils. *J. Biol. Chem.* 267, 4292-4295.

Selsted, M.E., Tang, Y.-Q., Morris, W.L., McGuire, P.A., Novotny, M.J., Smith, W., Henschen, A.H., and Cullor, J.S. (1993) Purification, primary structures, and antibacterial activities of β-defensins, a new family of antimicrobial peptides from bovine neutrophils. *J. Biol. Chem.* 268, 6641-6648.

Sheldrick, G., and Schneider, T. (1997) SHELXL: high-resolution refinement. *Methods Enzymol*. 277, 319-343 and SHELX-97/D/E manuals.

Shoseyov, O., Shani, Z., and Shpigel, E. (2001) Transgenic plants of altered morphology. US patent 6,184,440.

Shoseyov, O., Shani, Z., and Levy, I. (2006) Carbohydrate-binding modules: biochemical properties and novel applications. *Microbiol. Mol. Biol. Rev.* 70, 283-295.

Shoseyov, O., Takagi, M., Goldstein, M.A., and Doi, R.H. (1992) Primary sequence analysis of *Clostridium cellulovorans* cellulose-binding protein A. *Proc. Natl. Acad. Sci. USA* 89, 3483-3487.

Simpson, P.J., Bolam, D.N., Cooper, A., Ciruela, A., Hazlewood, G.P., Gilbert, H.J., and Williamson, M.P. (1999) A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity. *Structure* 7, 853-864.

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

Sleat, R., Mah, R.A., and Robinson, R. (1984) Isolation and characterization of an anaerobic, cellulolytic bacterium, *Clostridium cellulovorans* sp nov. *Appl. Environ. Microbiol.* 48, 88-93.

Smits, C.H.M., and Annison, G. (1996) Non-starch plant polysaccharides in broiler nutrition - towards a physiological valid approach to their determination. *W. Poultry Sci. J.* 52, 203-221.

Somogyi, M. (1945) A new reagent for the determination of sugars. *J. Biol. Chem.* 160, 69-73.

Somogyi, M. (1952) Notes on sugar determination. *J. Biol. Chem.* 195, 19-23.

Spezio, M., Wilson, D.B., and Karplus, P.A. (1993) Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry* 32, 9906-9916.

Stålbrand, H., Saloheimo, A., Vehmaanperä, J., Henrissat, B., and Penttilä, M. (1995) Cloning and expression in *Saccharomyces cerevisae* of a *Trichoderma reesei β*-mannanase gene containing a cellulose-binding domain. *Appl. Environ. Microbiol.* 61, 1090-1097.

Suetake, T., Tsuda, S., Kawabata, S., Miura, K., Iwanaga, S., Hikichi, K., Nitta, K., and Kawano, K. (2000) Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding structural motif. *J. Biol. Chem.* 275, 17929-17932.

Sun, Y.-D., Fu, L.-D., Jia, Y.-P., Du, X.-J., Wang, Q., Wang, Y.-H., Zhao, X.-F., Yu, X.-Q., and Wang, J.-X. (2008) A hepatopancreas-specific C-type lectin from the Chinese shrimp *Fenneropenaeus chinensis* exhibits antimicrobial activity. *Molec. Imunn.* 45, 348-361.

Suurnäkki, A., Tenkanen, M., Siika-Aho, M., Niku-Paavola, M.-L., Viikari, L., and Buchert, J. (2000) *Trichoderma reesei* cellulases and their core domains in the hydrolysis and modification of chemical pulp. *Cellulose* 7, 189-209.

Szabo, L., Jamal, S., Xie, H., Charnock, S.J., Bolam, D.N., Gilbert, H.J., and Davies, G.J. (2001) Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. *J. Biol Chem*. 276, 49061-49065.

Takeo, K. (1984) Affinity electrophoresis: principles and applications. *Electrophoresis* 5, 187-195.

Talbott, L.D., and Ray, P.M. (1992) Molecular size and separability features of pea cell wall polysaccharides. *Plant Physiol.* 98, 357-368.

Taylor, E.J, Goyal, A., Guerreiro, C.I.P.D., Prates, J.A.M., Money, V.A., Ferry, N., Fontes, C.M.G.A., and Davies, G.J. (2005) How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides: structure and activity of a *Clostridium thermocellum* lichenase, *Ct*Lic26A. *J. Biol. Chem.* 280, 32761-32767.

Terwilliger, T.C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr.* D55, 849-861.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22, 4673-4680.

Thompson, D.S. (2005) How do cell walls regulate plant growth? *J. Exp. Bot.* 56, 2275-2285.

Tomme, P., Boraston, A., Kormos, J.M., Warren, A.J., and Kilburn, D.G. (2000) Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules. *Enz. Microb. Technol.* 27, 453-458.

Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A.L., Sturch, K., Gilkes, N.R., Haynes, C.A., Warren, R.A., and Kilburn, D.G. (1998) Characterization and affinity applications of cellulose-binding domains. *J. Chromatogr. B.* 715, 283-296.

Tomme, P., Driver, D.P., Amandoron, E.A., Miller, J., Robert, C., Warren, R.A.J., and Kilburn, D.G. (1995) Comparison of a fungal (family I) and bacterial (family II) cellulose-binding domain. *J. Bacteriol.* 177, 4356-4363.

Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., and Claeyssens, M. (1988) Studies of the cellulolytic system of *Trichoderma reesei* QM 9414: analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur. J. Biochem.* 170, 575-581.

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

Travis, S.M., Anderson, N.N., Forsyth, W.R., Espiritu, C., Conway, B.D., Greenberg, E.P., McCray, Jr., P.B., Lehrer, R.I., Welsh, M.J., and Tack, B.F. (2000) Bactericidal activity of mammalian cathelicidin-derived peptides. *Infect. Immun.* 68, 2748-2755.

Vaaje-Kolstad, G., Horn, S.J., van Aalten, D.M., Synstad, B., and Eijsink, V.G. (2005) The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *J. Biol. Chem.* 280, 28492-28497.

Vagin, A. and Teplyakov, A. (1997) *MOLREP*: an Automated Program for Molecular Replacement. *J. Appl. Cryst.* 30, 1022-1025.

Vallée, F., Karaveg, K., Herscovics, A., Moremen, K.W., and Howell, P.L. (2000) Structural basis for catalysis and inhibition of N-glycan processing class I alpha 1,2-mannosidases. *J. Biol. Chem.* 275, 41287-41298.

Van der Klis, J.D., Van Voorst, A., and Van Cruyningen, C. (1993) Effect of a soluble polysaccharide (carboxymethyl cellulose) on the physico-chemical conditions in the gastrointestinal tract of broilers. *Br. Poult. Sci.* 34, 971-983.

Van Tilbeurgh, H., Tomme, P., Claeyssens, Bhikhabhai, R., and Pettersson, G. (1986) Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. *FEBS Letters.* 204, 223-227.

Varrot, A., Tarling, C.A., Macdonald, J., Stick, R.V., Zechel, D., Withers, S.G., and Davies, G.J. (2003) Direct observation of the protonation state of an imino sugar glycosidase inhibitor upon binding. *J. Am. Chem. Soc.* 125, 7496-7497.

Varrot, A., Leydier, S., Pell, G., Macdonald, J.M., Stick, R.V., Gilbert, H.J., and Davies, G.J. (2005) *Mycobacterium tuberculosis* strains possess functional cellulases. *J. Biol. Chem.* 280, 20181-20184.

Vasella, A., Davies, G., and Böhm, M. (2002) Glycosidase mechanisms. *Curr. Op. Chem. Biol.* 6, 619-629.

Vincent, F., Gloster, T.M., Macdonald, J., Morland, C., Stick, R.V., Dias, F.M.V., Prates, J.A.M., Fontes, C.M.G.A., Gilbert, H.J., and Davies, G.J. (2004) Common inhibition of both beta-glucosidases and beta-mannosidases by isofagomine lactam reflects different conformational itineraries for pyranoside hydrolysis. *Chem. Bio. Chem.* 5, 1596-1599.

Vincken, J.P., Dekeizer, A., Beldman, G., and Voragen, A.G.J. (1995) Fractionation of xyloglucan fragments and their interaction with cellulose. *Plant Physiol.* 108, 1579-1585.

Vincken, J.P., York, W.S., Beldman, G., and Voragen, A.G.J. (1997) Two general branching patterns of xyloglucan, XXXG and XXGG. *Plant Physiol.* 114, 9-13.

Wang, A.A., Mulchandani, A., and Chen, W. (2002) Specific adhesion to cellulose and hydrolysis of organophosphate nerve agents by a genetically engineered *Escherichia coli* strain with a surface-expressed cellulose-binding domain and organophosphorus hydrolase. *Appl. Environ. Microbiol.* 68, 1684-1689.

Wei, Q., Kim, Y.S., Seo, J.H., Jang, W.S., Lee, I.H., and Cha H.J. (2005) Facilitation of expression and purification of an antimicrobial peptide by fusion with baculoviral polyhedrin in *Escherichia coli*. *Appl. Environ. Microbiol.* 71, 5038-5043.

Weyer, E.R., and Rettger, L.F. (1927) A comparative study of six different strains of the organism commonly concerned in large-scale production of butyl alcohol and acetone by the biological process. *J. Bacteriol.* 14, 399-424.

White, W.B., Bird, H.R., Sunde, M.L., Burger, W.C., and Marlett, J.A. (1981) The viscosity interaction of barley beta-glucan with *Trichoderma viride* cellulase in the chick intestine. *Poult. Sci.* 60, 1043-1048.

Wood, T.M. (1988) Preparation of crystalline, amorphous, and dyed cellulase substrates. *Methods Enzymol.* 160, 19-25.

Xiao, Z.Z., Gao, P.J., Qu, Y.B., and Wang, T.H. (2001) Cellulose-binding domain of endoglucanase III from *Trichoderma reesei* disrupting the structure of cellulose. *Biotechnol. Lett.* 23, 711-715.

Xie, H., Bolam, D.N., Nagy, T., Szabo, L., Cooper, A., Simpson, P.J., Lakey, J.H., Williamson, M.P., and Gilbert, H.J. (2001a) Role of hydrogen bonding in the interaction between a xylan-binding module and xylan. *Biochemistry* 40, 5700-5707.

Xie, H., Gilbert, H.J., Charnock, S.J., Davies, G.J., Williamson, M.P., Simpson, P.J., Raghothama, S., Fontes, C.M.G.A., Dias, F.M., Ferreira, L.M.A., and Bolam D.N. (2001b) *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. *Biochemistry* 40, 9167-76.

Xu, G.Y., Ong, E., Golkes, N.R., Kilburn, D.G., Muhandiram, D.R., Harris-Brandts, M., Carver, J.P., Kay, L.E., and Harvey, T.S. (1995) Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry* 34, 6993-7009.

Xu, Q., Bayer, E.A., Goldman, M., Kenig, R., Shoham, Y., and Lamed, R. (2004) Architecture of the *Bacteroides cellulosolvens* cellulosome: description of a cell surface-anchoring scaffolding and a family 48 cellulase. *J. Bacteriol.* 186, 968-977.

Xu, Q., Gao, W., Ding, S.-Y., Kenig, R., Shoham, Y., Bayer, E.A., and Lamed, R. (2003) The cellulosome system of *Acetivibrio cellulolyticus* includes a novel type of adaptor protein and a cell-surface anchoring. *J. Bacteriol.* 185, 4548-4557.

Xu, X., Jin, F., Yu, X., Ji, S., Wang, J., Cheng, H., Wang, C., and Zhang, W. (2007) Expression and purification of a recombinant antibacterial peptide, cecropin, from *Escherichia coli*. *Protein Expr. Purif.* 53, 293-301.

Xu, Z., Bae, W., Mulchandani, A., Mehra, R.K., and Chen, W. (2002) Heavy metal removal by novel CBD-EC20 sorbents immobilized on cellulose. *Biomacromolecules* 3, 462-465.

Xu, Z., Peng, L., Zhong, Z., Fang, X., and Cen, P. (2006) High-level expression of a soluble functional antimicrobial peptide, human β-defensin 2, in *Escherichia coli*. *Biotechnol. Prog.* 22, 382-386.

Yague, E., Beguin, P., and Aubert, J.P. (1990) Nucleotide sequence and deletion analysis of the cellulase-encoding gene celH of *Clostridium thermocellum*. *Gene* 89, 61-67.

Yamatoya, K., and Shirakawa, M. (2003) Xyloglucan: structure, theological properties, biological functions and enzymatic modification. *Curr. Trends Polym. Sci.* 8, 27-72.

Yang, D., Biragyn, A., Kwak, L.W., and Oppenheim, J.J. (2002) Mammalian defensins in immunity: more than just microbicidal. *Trends Immunol.* 23, 291-296.

Yang, Y.-H., Zheng, G.-G., Li, G., Zhang, X.-J., Cao, Z.-Y., Rao, Q., and Wu, K.-F. (2004) Expression of bioactive recombinant GSLL-39, a variant of human antimicrobial peptide LL-37, in *Escherichia coli*. *Protein Expr. Purif.* 37, 229-235.

Yaoi, K., and Mitsuishi, Y. (2002) Purification, characterization, cloning, and expression of a novel xyloglucan-specific glycosidase, oligoxyloglucan reducing end-specific cellobiohydrolase. *J. Biol. Chem.* 277, 48276-48281.

Yaoi, K., and Mitsuishi, Y. (2004) Purification, characterization, cDNA cloning, and expression of a xyloglucan endoglucanase from *Geotrichum* sp M128. *FEBS Lett.* 560, 45-50.

Yaoi, K., Kondo, H., Noro, N., Suzuki, M., Tsuda, S., and Mitsuishi, Y. (2004) Tandem repeat of a seven-bladed beta-propeller domain in oligoxyloglucan reducing-end-specific cellobiohydrolase. *Structure* 12, 1209-1217.

Yaoi, K., Nakai, T., Kameda, Y., Hiyoshi, A., and Mitsuishi, Y. (2005) Cloning and characterization of two xyloglucanases from *Paenibacillus* sp strain KM21. *Appl. Environ. Microbiol*. 71, 7670-7678.

Yokoyama, R., Rose, J.K.C., and Nishitani, K. (2004) A surprising diversity and abundance of xyloglucan endotransglucosylase/hydrolases in rice. Classification and expression analysis. *Plant Physiol.* 134, 1088-1099.

Zanetti, M., Gennaro, R., and Romeo, D. (1995) Cathelicidins: a novel protein family with a common proregion and a variable C-terminal antimicrobial domain. *FEBS Lett.* 374, 1-5.

Zanetti, M., Storici, P., Tossi, A., Scocchi, M., and Gennaro, R. (1994) Molecular cloning and chemical synthesis of a novel antibacterial peptide derived from pig myeloid cells. *J. Biol. Chem.* 269, 7855-7858.

Zasloff, M. (1987) Magainins, a class of antimicrobial peptides from *Xenopus* skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc. Natl. Acad. Sci. USA* 84, 5449-5453.

Zechel, D.L., Boraston, A.B., Gloster, T.M., Boraston, C.M., Macdonald, J.M., Tilbrook, M.G., Stick, R.V., and Davies, G.J. (2003) Iminosugar glycosidase inhibitors: structural and thermodynamic dissection of the binding of isofagomine and 1-deoxynojirimycin to beta-glucosidases. *J. Am. Chem. Soc.* 125, 14313-14323.

Zhou, Q., Greffe, L., Baumann, M.J., Malmström, E., Teeri, T.T., and Brumer III, H. (2005) Use of xyloglucan as a molecular anchor for the elaboration of polymers from cellulose surfaces: a general route for the design of biocomposites. *Macromolecules* 38, 3547-3549.

Zhou, Q., Baumann, M.J., Piispanen, P.S., Teeri, T.T., and Brumer, H. (2006) Xyloglucan and xyloglucan endo-transglycosylases (XET): tools for *ex vivo* cellulose surface modification. *Biocatal. Biotransform.* 24, 107-120.

Zverlov, V.V., Kellermann, J., and Schwarz, W.H. (2005a) 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, 3646-3653.

Zverlov, V.V., Schantz, N., Schmitt-Kopplin, P., and Schwarz, W.H. (2005b) Two new major subunits in the cellulosome of *Clostridium thermocellum*: xyloglucanase Xgh74A and endoxylanase Xyn10D. *Microbiology* 151, 3395-3401.

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