Structural Studies of Cellulose and Chitin Active Enzymes

Wimal Ubhayasekera
Faculty of Natural Resources and Agricultural Sciences
Department of Molecular Biology
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2005

Acta Universitatis Agriculturae Sueciae
2005:18

ISSN 1652-6880
ISBN 91-576-7017-X
© 2005 Wimal Ubhayasekera, Uppsala
Tryck: SLU Service/Repro, Uppsala 2005

Abstract

Ubhayasekera, W., 2005, *Structural Studies of Cellulose and Chitin Active Enzymes*. Doctoral dissertation
ISSN: 1652-6880, ISBN: 91-576-7017-X
Serial publication number: 2005:18

Cellulose and chitin, the main ways of storing biological energy in nature, also play a vital role in the structures of many organisms. Cellulose is the main structural component in plants whereas chitin is found in invertebrates and fungi. Gaining a better understanding of the degradation of these polymers can have direct or indirect economic impact. This thesis summarizes the structural perspectives of the cellulose and chitin degradation machinery.

The white-rot fungus *Phanerochaete chrysosporium* has six cellobiohydrolases, which are expressed differentially with varying stimuli and time intervals. X-ray crystal structures of one of the six isozymes (Pc_Cel7D) suggested that it uses a retention mechanism and acts from the reducing end of cellulose chain. Homology modeling of the other enzymes supported the same sort of mechanism for all except one (Pc_Cel7B) and considerably different dynamic properties for two isozymes (Pc_Cel7A and Pc_Cel7B).

Piromyces sp. strain E2 Cel9A and Cel6A as well as *Piromyces equi* Cel6A are modular structures, which function as parts of the fungal cellulosome of the respective organisms. Homology modeling supported the conclusion that Cel9A is an endoglucanase having a wide active site cleft and a conserved calcium-binding site with an inverting catalytic mechanism, whereas the Cel6As are processive cellobiohydrolases that act via an inverting mechanism that releases cellobiose from the non-reducing end of the cellulose chain.

Brassica juncea endo acting chitinase is a pathogenesis-related protein that acts in defense of the plant. A homology model of the catalytic module was useful in designing mutants that helped us to understand the substrate binding and catalytic processes. X-ray crystal structures of the catalytic module and a mutant extended the knowledge of how the enzyme acts during the catalysis, with conformational changes opening and closing the enzyme.

The homology model of yam, *Dioscorea opposita*, class IV endochitinase suggests that this enzyme catalyzes chitin cleavage via an inverting mechanism. Deletions in class IV chitinases compared to class I/II cluster at the ends of the substrate-binding cleft, shortening it by one glycosyl unit at each end. The shorter cleft might be expected to recognize and grasp a small section of exposed chitin on a fungal hyphal wall, more effectively attacking it.

Keywords: cellobiohydrolase, cellulase, cellulose, chitin, chitinase, endoglucanase, homology modeling, pathogenesis-related protein, structural studies, X-ray crystallography

Author's address: Wimal Ubhayasekera, Department of Molecular Biology, Swedish University of Agricultural Sciences, Box 590, Biomedical Center, SE-751 24, Uppsala, Sweden.
E-mail: wimal@xray.bmc.uu.se

පිදුම
මගේ ආදරණීය මැණියන්ට හා පියාණන්ට

To my loving mother & father

Contents

1 Introduction 9
1.1 General Introduction 9
1.2 Aim of the thesis 10

2 Background 11
2.1 Glycoside hydrolase classification 11
2.2 Cellulose and cellulases 12
2.2.1 Cellulose 12
2.2.2 Cellulose degradation 13
2.3 Chitin and chitin active enzymes 16
2.3.1 Chitin 16
2.3.2 Chitin degradation 17
2.3.3 Plant chitinase genes and classes 18
2.4 Auxiliary modules 20
2.4.1 Cellulose binding 20
2.4.2 Cellulosome 20
2.4.3 Chitin binding 21
2.4.4 Linker 21
2.4.5 Vacuolar targeting 22
2.5 Glycoside hydrolase catalytic mechanisms 22

3 Results and Discussion 25
3.1 *Phanerochaete chrysosporium* cellobiohydrolases 25
3.1.1 Pc_Cel7D structure and function (Paper I) 25
3.1.2 Other cellobiohydrolases of P. chrysosporium *(Paper II)* 31
3.2 *Piromyces* sp. strain E2 Cel9A (Paper III) 33
3.2.1 Homology modeling 33
3.2.2 Results and discussion 33
3.2.3 Auxiliary modules in Cel9A 34
3.3 *Piromyces* sp strain E2 Cel6A (Paper IV) 36
3.3.1 Homology modeling 36
3.3.2 Cel6A Results and discussion 36
3.4 *Brassica juncea* chitinase 37
3.4.1 Model of the catalytic module (Paper V) 38
3.4.2 Brassica juncea *crystal structures (Paper VI)* 39
3.5 Yam chitinase (Paper VII) 43
3.5.1 Homology modeling 43
3.5.2 Overview of the model 43

4 Concluding remarks 45

5 Reference 46

6 Acknowledgements 51

Appendix

Appendix A

Table of data collection and refinement statistics for *Brassica juncea* chitinase strcutures discussed in section 3.4.2

Papers I-VII

This thesis is based on the following publications and manuscripts, which are referred to in the summary by their Roman numerals.

I. Wimal Ubhayasekera, Muñoz, I.G., Vasella, A., Ståhlberg, J. And Mowbray, S.L. (2004). Structures of *Phanerochaete chrysosporium* Cel7D in complex with product and inhibitors. Accepted for publication in *FEBS Journal*.

II. Inés G. Muñoz, Wimal Ubhayasekera, Hongbin Henriksson, Istvan Szabó, Göran Pettersson, Gunnar Johansson, Sherry L. Mowbray and Jerry Ståhlberg (2001). Family 7 Cellobiohydrolases from *Phanerochaete chrysosporium*: Crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 Å resolution and homology models of the isozymes. *J. Mol. Biol*. 314, 1097-1111.

III. Steenbakkers, P. J. M., Ubhayasekera, W., Goossen, H. J. A. M., van Lierop, E. M. H. M., van der Drift, C., Vogels, G. D., Mowbray, S. L. & Op den Camp , H. J. M. (2002). An intron-containing family 9 cellulase gene encodes the dominant 90 kDa component of the cellulosome of the anaerobic fungus *Piromyces* sp. strain E2. *Biochem J*. 365, 193-204.

IV. Harry R. Harhangi, Alexander C. J. Freelove, Wimal Ubhayasekera, Maarten van Dinther, Peter J. M. Steenbakkers, Anna Akhmanova, Chris van der Drift, Mike S.M. Jetten, Sherry L. Mowbray, Harry J. Gilbert and Huub J. M. Op den Camp (2003). Cel6A, a major exoglucanase from the cellulosome of the anaerobic fungi *Piromyces* sp. E2 and *Piromyces equi Biochem Biophys Acta*, 1628, 30-39.

V. Ce Mun Tang, Mee-Len Chye, Sathishkumar Ramalingam, Ouyang Shi-Wen, Kai-Jun Zhao, Wimal Ubhayasekera and Sherry Mowbray (2004). Functional analysis of the chitin-binding domains and the catalytic domain of *Brassica juncea* chitinase BjCHI1. *Plant Mol Biol*, 56, 285-298.

VI. Wimal Ubhayasekera, Ce Mun Tang, Terese Bergfors, Gunnar Berglund, Mee-Len Chye and Sherry L. Mowbray (2005). Crystal structure of a family 19 chitinase from *Brassica juncea* shows flexibility of binding cleft loops. *manuscript*.

VII. Takuji Mitsunaga, Minoru Iwase, Wimal Ubhayasekera, Sherry Lynn Mowbray and Daizo Koga (2004). Molecular cloning of a genomic DNA encoding yam class IV chitinase and homology modeling. *J. Bioscience, Biotechnology and Biochemistry*, 68(7), 1508-17.

Additional publications

i. Nina Grantcharova , Wimal Ubhayasekera , Sherry L. Mowbray, Joseph R.McCormick, and Klas Flärdh, (2003). A missense mutation in ftsZ differentially affects vegetative and developmentally controlled cell division in *Streptomyces coelicolor* A3(2). *Mol Microbiol*, 47, 645-656.

Reprints of the articles were made with permission from the publishers.

Abbreviations

Bj – *Brassica juncea*
CBM – cellulose binding module
CBH – cellobiohydrolase
CM – Catalytic module
CtBM – chitin binding module
EG - endoglucanase
GG-S-GG - methyl 4-S-β-cellobiosyl-4-thio-β-cellobioside $(Glc)_2$-S-$(Glc)_2$
GH – glycoside hydrolase
IBTG - *o*-iodobenzyl-thio-glucoside
MR – molecular replacement
Pc – *Phanerochaete chrysosporium*
Pc_Cel7D – cellobiohydrolase Cel7D from *Phanerochaete chrysosporium*
PDB – Protein Data Bank;
PR – pathogenesis related protein
r.m.s. – root mean square
SBM – substrate binding module
Tr_Cel7A – cellobiohydrolase Cel7A from *Trichoderma reesei*

1 Introduction

1.1 General Introduction

The carbon cycle is one of the most important pathways on earth, maintaining the flux of raw materials and energy through the many and varied members of earth's living systems. Plants using the energy from sunlight, together with water, fix a considerable fraction of the total carbon dioxide in the atmosphere. Microbes such as bacteria and fungi are able to recycle the fixed carbon with their specific extracellular enzymes. Carbon fixation produces glucose initially, and this glucose is the basis for producing a number of important polymers such as cellulose and chitin.

Cellulose is found in plants as micro-fibrils, where the micro-fibrils help form a cell wall that is structurally strong. Although cellulose is primarily a plant material, some bacteria also have the capacity to produce it. Cellulose occurs in essentially all parts of plants either in a pure form (*e.g.* a cotton fiber) or in combination with other materials, such as lignin and hemicellulose.

Chitin is the most abundant nitrogen-bearing biopolymer with a high turnover in nature, *i.e.* that most important in the carbon cycle. The abundance of chitin biomass is second only to cellulose, but when it comes to recycling, chitin degrades much faster. Chitin, as an insoluble polymer, is widely distributed among microorganisms and invertebrates as structural components in the cell wall and skeletal tissue. In plants, chitin-derived molecules play a significant role in growth regulation. Approximately ten gigatons (1×10^{13} kg) of chitin are synthesized and degraded each year in the biosphere (Muzzarelli, R.A.A., 1999; Stevens, W.F., 1996).

It is fascinating to uncover the secrets of nature that allow the breakdown of such hard materials as cellulose and chitin. Most microorganisms, such as bacteria and fungi, have the appropriate tools, *i.e.* enzymes like cellulases and chitinases, in their 'tool box' that can be used to degrade such polymers. In most cases, there are multiple copies of the same type of enzyme (isozymes) with relatively small differences. The reasons for the occurrences of multiple copies of the same enzyme have yet to be fully disclosed, although it is presumed that such advantages do exist.

The agents needed for polymer breakdown, *e.g.* cellulases and chitinases, have become even more useful due to our ability to chose or tailor them to suit our needs. Apart from the value of understanding the enzymology and carbon recycling, we gain information that has vast industrial and economic value by studying these enzymes. Cellulases are used in many industrial preparations of products such as food, beverages, textiles, pulp and paper, as well as in saccharification. Chitin-active enzymes are important in chitosan production, and in obtaining smaller chitin-based compounds, as well as in plant protection and the production of disease-resistant plants.

1.2 Aim of the thesis

The overall aim of this thesis work was to furnish the structural perspectives needed to interpret the existing biochemical data, and to get insights into the catalysis of glycoside hydrolase (GH) family 6, 7, 9 and 19 enzymes. This task was achieved through both X-ray crystal structures and homology models. Structures of *Phanerochaete chrysosporium* Cel7D catalytic module (CM) include the apo protein as well as product/inhibitor complexes; these, together with structures of *Brassica juncea* chitinase CMs and an inactive mutant, represent the crystallographic work in the thesis. Homology models of *P. chrysosporium* Cel7A, Cel7B, Cel7C, Cel7E, Cel7F, *Piromyces* sp. strain E2 Cel9A and Cel6A, *Piromyces equi* Cel6A, and *Brassica juncea* chitinase CM were employed as sources of 3D information in the absence of crystal structures.

2 Background

2.1 Glycoside hydrolase classification

Many carbohydrate-active enzymes have evolved in nature, in all sorts of organisms, *i.e.* archaebacteria, eubacteria and eukaryotes, to deal with the wide variety of carbohydrate chemistry needed. Like other enzymes, each of these can be classified according to the catalyzed reaction's type and substrate specificity. These features are used in the International Union of Biochemistry (IUB) enzyme nomenclature (1984) in which EC 3.2.1.x numbers were given to various glycoside hydrolases, where the first three digits indicate the hydrolysis of *O*-glycosyl linkages. The last number shows the substrate, and sometimes gives information concerning the reaction mechanism. The main limitation of this classification is that neither structural nor evolutionary information is included. An increasing number of protein structures in the databases display evidence of both divergent (changes in specificity and reaction type) and convergent (proteins with different folds catalyze the same reaction) evolution.

Therefore, based on sequence identities and similarities a new glycoside hydrolase (GH) classification was proposed (Henrissat, B., 1991). As significant sequence similarity indicates a structural similarity (Chothia, C. & Lesk, A.M., 1986), members in the same sequence family exhibit the same fold. This fact gives researchers the opportunity for homology modeling of the family members, if one or more sufficiently similar 3D structures is known (Henrissat, B., 1991).

There are 99 GH families that have been distinguished to date. Among them, members of families 5 through 10, 12, 26, 44, 45, 48, 51, 61, and 74 are cellulases (cellobiohydrolases and endoglucanases) whereas families 18 and 19 include chitinases. The inverting cellobiohydrolases of family 6 are believed to act processively from the non-reducing end whereas the family 7 retaining exocellulases (CBH) act from the reducing end of the cellulose chain producing cellobiose. Family 9 has inverting endoglucanases (EGs), cellobiohydrolases (CBH) and β-glucosidases. These EGs produce varying sized products. Family 18 is comprised of retaining exochitinases whereas family 19 has inverting endochitinases. GH catalytic mechanisms are described further in the section 2.5.

It has been shown that there exists a direct relationship between sequence and folding. Therefore, homologous proteins possess core regions with sequence similarities and a common fold, although some other parts of the proteins may differ. Homology models are thus an extremely useful way of viewing sequence information in the context of the available structural data. The success of homology modeling of a protein depends on the amino-acid sequence identity of the protein of interest and the homologous structure (template). To get a homology model of high quality, sequence identity has to be >50%. If this value drops down to 20%, there are likely to be large structural differences, making structure prediction impossible in these cases at present. However, even with such a low similarity, the active site may be conserved, allowing some modeling of the

region (Chothia, C. *et al.*, 1986). Many real cases lie in-between these two extremes, and appropriate caution must be taken in the interpretation of the available data.

2.2 Cellulose and cellulases

2.2.1 Cellulose

Cellulose (Figure 1), the most abundant structural polymer on earth, is one of the main energy storage reservoirs in nature. Anselme Payen, a French chemist, in 1837 separated wood material into its constituents, in the course of which he recognized the existence of cellulose in plant cell walls; the term 'cellulose', however, was not coined until later.

This unbranched polymer consists of 1000 to 1 million D-glucose units linked together by β-1,4 glycosidic bonds. These β-linked glucopyranose residues are stabilized in the chair structure due to the fully equatorial conformation ($^4C_1$) of the attached hydroxyl groups.

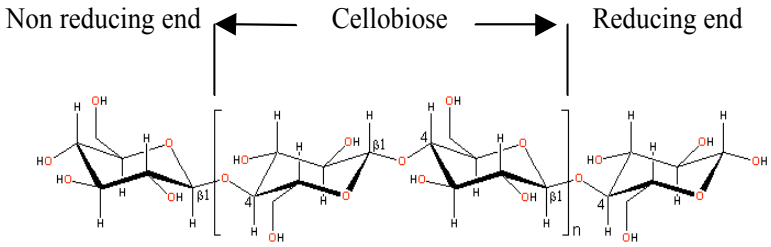

Figure 1. Chemical structure of cellulose, the polymer built from units of cellobiose

Sugar units are linked via single oxygen atoms (acetal linkages) between the C-1 of one pyranose ring and the C-4 of the next ring. This acetal is formed by reacting an alcohol with a hemiacetal, and removing a water molecule. Therefore, the glucose units in cellulose are referred to as anhydroglucose units. The stereochemistry or the spatial arrangement of the acetal linkages is very important, because the pyranose rings of cellulose have groups larger than hydrogen in the equatorial positions. Stereochemistry of the glucose at carbons 2, 3, 4 and 5 is fixed, but two different arrangements are possible for the hydroxyl at C4 in its bond to the carbon at C-1. When both the C-1 hydroxyl group and the C-6 carbon are on the same side of the ring, it is said to be in the α-configuration. In the case of cellulose, the C-1 oxygen is placed at the opposite side, giving the β-configuration. This β-configuration and the placement of all the functional groups in equatorial positions combine to make the cellulose chain linear, so making it a good fiber-forming polymer. The equatorial positions of the hydroxyls on the cellulose chain make them available for intra-molecular and intra-strand hydrogen bonding. These hydrogen bonds fix the cellulose chains in a highly ordered pattern, creating crystal-like, rigid, stable and insoluble microfibril structures. As well as such crystalline regions, there are also more disordered regions, where other molecules can make hydrogen bonds. These amorphous regions can absorb

large quantities of water, resulting in swelling but not dissolution. At the molecular level, cellulose has the same structure regardless of the source. However, it differs in the crystallinity and binding by other biological chemicals. Each sugar residue is rotated 180° with respect to the next, with the chain synthesized two residues at a time. Cellobiose is thus the repeating unit of cellulose (Clarke, A.J., 1997; Fibersource, 2005).

2.2.2 Cellulose degradation

All biologically and technologically important processes involving cellulose take place at its surface (Heiner, A.P. & Teleman, O., 1998). The general understanding of cellulose degradation at present is that for efficient breakdown of crystalline cellulose, three types of enzymes are needed (Figure 2). They are EGs (which cut cellulose chains randomly), cellobiohydrolases (CBHs, which cleave cellobiose from the cellulose chain ends) and β-glucosidases (which hydrolyze cellobiose and cellodextrins into glucose) (Boisset, C. *et al.*, 2000). Efficient cellulose degradation by microbes is achieved in two major ways. Production of a 'soup' of different enzymes that can work synergistically is one method. These different enzymes could have differential expression depending on the substrate, time and other stimuli. The second method is cellulosome activity. In this approach, different enzymes are found associated as one complex (the cellulosome) for fast breakdown of cellulose.

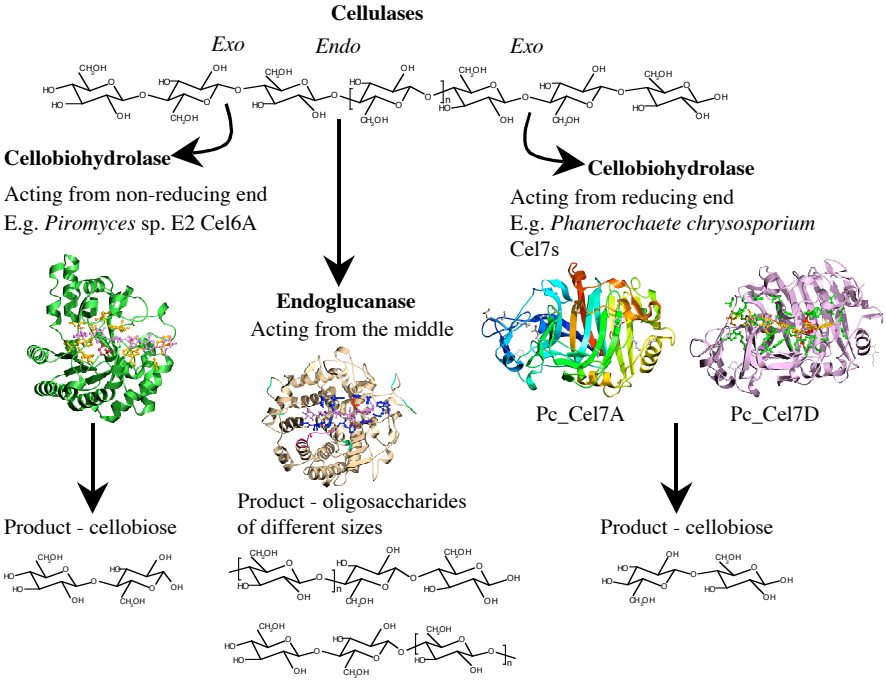

Figure 2. Cellulose active enzymes

The two major fungal cellulase types are EG (1,4-β-D-glucan glucanohydrolase) and CBH (1,4-β-D-glucan cellobiohydrolase). EGs are thought to cleave cellulose chains at random positions, so exposing new chain ends and creating starting points that give CBHs the opportunity to act. CBHs are categorized as exocellulases but recent knowledge has suggested that these enzymes could be able to perform an initial endo attack to generate ends from which they proceed (Muñoz, I.G. *et al.*, 2001). Once the substrate is bound, the enzyme continues the action on the same cellulose chain. This continuity of the action is called 'processivity'. The main features of the enzymes having this sort of behaviors are an enclosed active site cleft by long loops and the presence of multiple binding sites for the substrate along the tunnel starting from the substrate entry side. These features require the enzyme to remain bound to the cellulose chain after the initial binding event. The enzyme cleavage is possible for only every second bond. In the case of an EG, shorter loops that convert the binding site into a more open cleft, combined with fewer glycosyl-binding sub-sites, give the opportunity for this type of enzyme to attack the middle of the cellulose chain frequently (Harhangi, H.R. *et al.*, 2003). CBHs release cellobiose from the non-reducing or reducing end of cellulose due to their exo activity. A cellulose chain is thus threaded through CBH until it is cut at the active site, so producing cellobiose. The shape of this tunnel suggests the threading of the substrate into the active site for catalysis, restricting the reorientation of the cellulose chain due to the limited space available. CBHs display a broad specificity in the hydrolysis of crystalline and amorphous cellulose. Most of the fungal cellulases as well as some of the bacterial cellulases are glycoproteins. This glycosylation possibly helps to provide protection from proteolytic attack and/or thermostability as well as having a possible involvement in the adsorption of cellulases to insoluble cellulose (Clarke, A.J., 1997; Zou, J. *et al.*, 1999).

2.2.2.1 *Phanerochaete chrysosporium* Cel7s

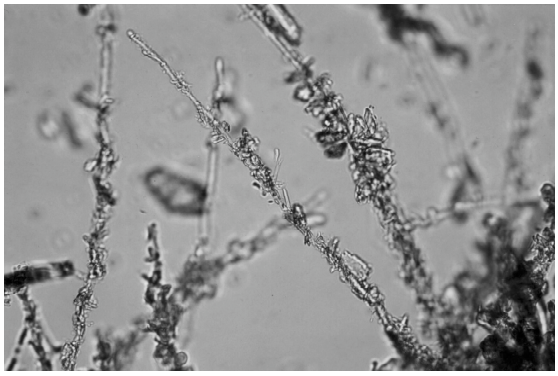

Figure 3. Light microscopic picture of *Phanerochaete chrysosporium* hyphae (courtesy of Dr. Kiyohiko Igarashi, University of Tokyo, Tokyo, Japan)

The basidiomycete fungus *P chrysosporium* (Figure 3) is a white rot fungus, which has the capacity of producing multiple isozyme forms of EG, CBH and β-glucosidase. The filamentous soft rot fungus *Trichoderma reesei* possesses the best-studied cellulase system, with at least two cellobiohydrolases and five EGs. These two cellobiohydrolases, Tr_Cel7A (CBHI) and Tr_Cel6A (CBHII), belong to GH family 7 and 6 respectively (Becker, D. *et al.*, 2001). Similar to these well-characterized *T. reesei* cellobiohydrolases, there are six CBHI-like and single CBHII-like genes in *P. chrysosporium*. Except for one (*cbh1-1*) all *cbh1* genes show a tripartite architecture (coding for catalytic and cellulose-binding modules connected by a linker) (Vallim, M.A. *et al.*, 1998). *P. chrysosporium* CBHI-like enzymes are classified into GH family 7. The naming convention for these proteins is summarized in the Table 1 (Muñoz, I.G. *et al.*, 2001).

Gene	Proposed enzyme name	Previous enzyme name	Identity to Pc_Cel7D %
cbh1-1	Cel7A	No report	65
cbh1-2	Cel7B	No report	65
cbh1-3	Cel7C	CBH1.1/CBH62	81
cbh1-4	Cel7D	CBH1.2/CBH58	-
cbh1-5	Cel7E	No report	84
cbh1-6	Cel7F	No report	85

Table 1. *P. chrysosporium* Cel7 genes and enzyme classification (Muñoz, I.G. *et al.*, 2001)

Both transcription analysis and protein purification studies have shown that cellobiohydrolase 58 (CBH58) or according to the above naming convention, Pc_Cel7D (the product from *cbh1-4* gene) is the major enzyme among the six Cel7 isozymes under many conditions (Uzcategui, E. *et al.*, 1991; Vallim, M.A. *et al.*, 1998). The exact role and interaction of the individual genes/enzymes in cellulose breakdown are unclear. However, there are reports suggesting that the combined activity of a soup of enzymes degrades cellulose. The involvement of extra-cellular oxidative enzymes such as cellobiose dehydrogenase in cellulose degradation in cooperation with cellulases has also been investigated (Igarashi, K. *et al.*, 2002).

2.2.2.2 *Piromyces* cellulosome

Piromyces sp. strain E2 (Figure 4), originally isolated from Indian elephant (*Elephas maximus*) faeces, is an obligatory anaerobic fungus, with an effective cellulolytic system. This fungus plays a significant role in the process of plant cell-wall material degradation and fermentation. In the case of crystalline cellulose, the degradation machinery is based on the high-molecular-mass (hemi)cellulolytic complex that is similar to the bacterial cellulosome. This fungal cellulosome consists of catalytic components connected to fungal dockerins. At present, the identified cellulase genes of anaerobic fungi belong to GH families 3, 5, 6, 9,45 and 48 (Harhangi, H.R. *et al.*, 2003; Steenbakkers, P.J. *et al.*, 2002).

Figure 4. Developing thallus of *Piromyces* sp. E2 with a bulging sporangium at the top; underneath is a highly branched rhizoid (courtesy of Dr. Huub Op den Camp, Radboud University, Nijmegen, The Netherlands (Copyright Claudius K. Stumm))

2.3 Chitin and chitin active enzymes

2.3.1 Chitin

Chitin (Figure 5) consists of 1,4 glycosidic bonded units of 2-acetamido-2-deoxy-β-D-glucan (N-acetyl glucosamine, GlcNAc). Discovered by Henri Braconnot in 1811,this polymer was named "chitine" by A. Odier in 1823 (from the Greek "χιτων", *i.e.* "chiton", which means tunic or coverage). C. Rouget discovered chitosan, the deacetylated form of chitin, in 1859. Chitin and chitosan polysaccharide structures have interesting physical and physicochemical properties. Due to the peculiar glycosidic linkage, this polymer has a relatively rigid structure. As there are in fact no fully acetylated or deacetylated polymers in nature, the terms chitin and chitosan are largely suggestive of the respective acetyl contents (Domard, A., 1996).

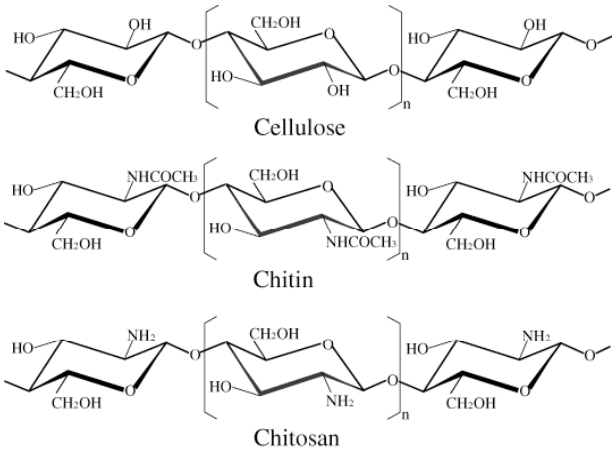

Figure 5. Chemical structures of cellulose, chitin and chitosan

Chitosan (Figure 5) has been proven to be a wonderful material. The versatility of this polymer is reflected by the vast range of applications in many areas of science, such as medicine, phytoprotection, cosmetics, food processing and packaging and technical usages (Stevens, W.F., 1996). As in all cases, nature has created a system to break down its own creations, which applies to chitin/chitosan by the chitin active enzymes produced by chitinolytic organisms.

2.3.2 Chitin degradation

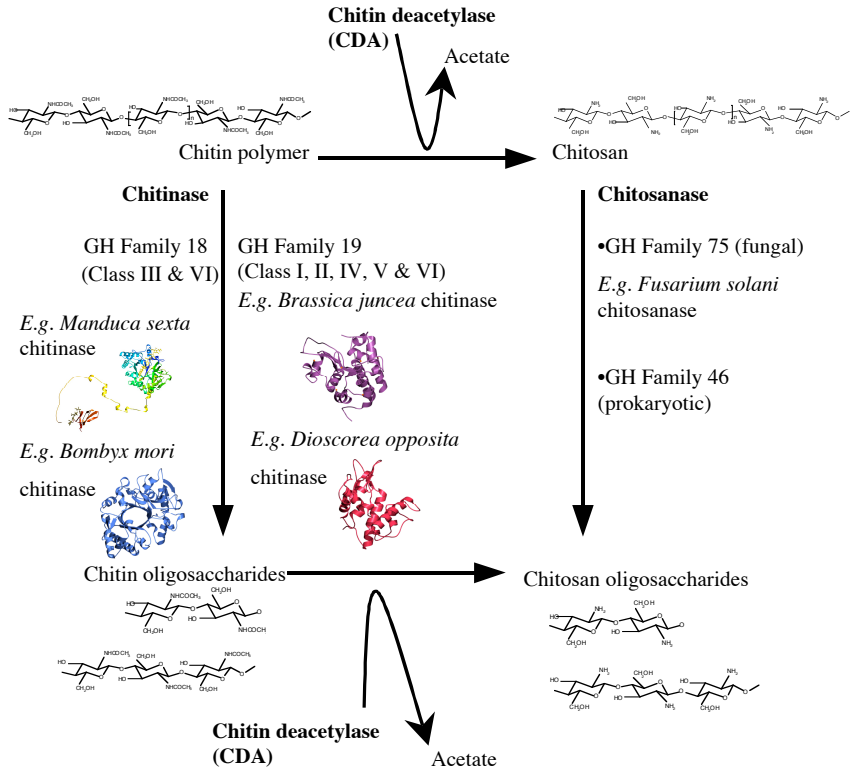

Figure 6. Chitin active enzymes

A variety of organisms possess a chitinolytic capacity. These include plants, microorganisms such as bacteria and fungi, arthropods and some of the other higher animals such as amphibians, fish and mammals.

Chitinase (Figure 6) is one of the major weapons for the defense against pathogens not only in higher plants and seaweeds but also in fish and mammals. Microorganisms digest the chitinous substrates for nutrients or hydrolyze their own chitinous cell walls for cell proliferation, while arthropods such as insects and crustaceans produce chitinase in order to degrade the exo-skeleton during eckdysis (Koga, D. *et al.*, 1999). Chitinases are classified into GH families 18 and 19, and therefore possess exo acting retaining mechanism and endo acting inverting

mechanism, respectively (Henrissat, B., 1999). These mechanisms are further discussed in section 2.5.

Pathogenesis-related proteins (PR) are defined as plant proteins that are induced in pathological or related situations. This group of proteins is expressed in plants as a response to a pathogenic attack, various stress factors (*e.g.* drought, salinity, wounding, heavy metals in their environment, endogenous and exogenous elicitor treatment and plant growth regulators) (Kasprzewska, A., 2003). These pathogenesis related proteins are divided into 11 classes based on biological properties, enzyme activity and sequence similarities (van Loon, L.C. *et al.*, 1994). Plant chitinase is one of the well-documented PR proteins.

Plant chitinase up-regulation is stimulated by a variety of stress conditions (both biotic and abiotic), phytohormones (*e.g.* ethylene, jasmonic acid and salicylic acid). The involvement of number of physiological events has also been reported including development and growth processes (Kasprzewska, A., 2003). It is clear that responses expressing chitinases are defensive. These chitinases act on chitin in pathogen cell walls, bacterial peptidoglycan and lipochito-oligosaccharides (Nod factors) produced by nitrogen fixing bacteria. Although there are no reports on endogenous substrates for chitinase, supporting evidence for the hydrolysis of arabinogalactan proteins and N-acetylglucosamine containing glycoproteins in cell walls has been reported (Kasprzewska, A., 2003).

2.3.3 Plant chitinase genes and classes

Plant endochitinases, a diverse group of enzymes, with differences in primary structure, isoelectric point and cellular localization, have been classified into 6 different classes, as summarized in Table 2.

GH family	PR proteins	Class	Gene name
19	PR-3	I	*Chia1*
		II	*Chia2*
		IV	*Chia4*
		V	*Chia5*
		VI	*Chia6*
18	PR-8	III	*Chib1*
18	PR-11	VI	*Chic1*
-	PR-4	I	*Chid1*
		II	*Chid2*

Table 2. Nomenclature of plant chitinase genes (Neuhaus, J.M. *et al.*, 1996)

The plant chitinases have been classified into classes based on amino acid sequence similarities, and the presence and absence of an N-terminal cysteine rich domain as follows.

Class I: >50% sequence identity to tobacco class I chitinase, conserved N-terminal domain
Class II: >50% sequence identity to tobacco class II and class I chitinases, lacks N-terminal domain
Class III: >30% sequence identity to tobacco class III chitinase/lysozyme, no sequence similarity to tobacco class I and class II enzymes
Class IV: >50% sequence identity to *Phaseolus vulgaris* PR 4 chitinase; there is a deletion in the N-terminal domain, several deletions in the CM (~22 amino acids deleted) and a truncated C-terminal end
Class V: >50% sequence identity to stinging nettle (*Urtica dioica*) lectin precursor and a duplicated N-terminal lectin domain
Class VI: >50% sequence identity to tobacco endochitinases and significant sequence similarity to bacterial exochitinases from *Bacillus circulans*, *Serratia marcescens* and *Streptomyces plicatus*, but no sequence similarity to the class I – V proteins (Meins, F. *et al.*, 1994).

Three different chitinases have been proposed in various papers to represent class VII chitinases. To be able to use the 'classes' to describe all plant chitinases, the following three suggestions must also be taken into account. As the rice chitinase sequence was submitted to the gene bank prior to the Indian mustard chitinase and cotton chitinase sequences, these sequences could be classified into class VII, VIII and IX respectively.

Class VII: *Oryza sativa* chitinase (gi|2055262|dbj|BAA19793.1), lacks N-terminal domain and C-terminal end, deletions in the CM compared to class I and class II tobacco chitinase (Li, W.L. *et al.*, 2001; Nakazaki, T. *et al.*, 1997; Truong, N.H. *et al.*, 2003).

Class VII: *Brassica juncea* chitinase (gi|6048743|gb|AAF02299.1) shows 62% identity to tobacco class I chitinase and has two N-terminal chitin binding domains (Zhao, K.J. & Chye, M.L., 1999).

Class VII: *Gossypium hirsutum* (gi|32401255|gb|AAP80801.1) chitinase shows less than 30% identity to tobacco class I chitinase, no N-terminal chitin binding domain, lacks short C-terminal extension (Li, J. & Liu, J.-Y., 2003).

Structures have been determined for GH family 18 chitinases of different origins such as bacteria, fungi, human and higher plants. The chitinase domain of this family consists of an eight-stranded α/β barrel (TIM-barrel) whereas eight strands of parallel β sheet are laid down with a helix as the "return stroke". The eight strands of the sheet bend into a barrel structure with the helices forming a ring around the outside (Robertus, J.D. & Monzingo, A.F., 1999).

There are only a few structures available for GH family 19 enzymes, and no ligand bound structures have been reported. This is partly due to the unavailability of inhibitors for this family. However, mercury has been reported to inhibit bacterial exochitinases (Kim, K.J., Yang, Y.J. & Kim, J.G., 2003; Wen, C.M. *et al.*, 2002). These enzymes belong to either bacteria or plants but are distantly

related to family 18 enzymes. Family 19 chitinases have a bilobal structure with a high α-helical content. Conserved hydrophobic residues form the core of the protein where the polar conserved residues line up along the large cleft suggesting their importance in substrate binding and catalysis.

2.4 Auxiliary modules

Multi modular glycoside hydrolases are a frequent observation in most genomes. These modules generally have functions such as substrate binding, vacuolar (or any other) targeting or catalytic action, and are connected by linker/hinge regions; some other modules have unknown function. One enzyme could possess one or several substrate binding modules (SBMs). Stability of the SBMs is mainly governed by conserved disulfide bridges, especially in extracellular proteins. Therefore, these modules are usually cysteine rich (Beintema, J.J., 1994).

2.4.1 Cellulose binding

All cellulolytic bacteria and fungi produce an array of cellulases that act on crystalline cellulose. These cellulases can be associated into multi-enzymatic complexes (known as a cellulosome) or act as individual enzymes. A modular structure is usually seen in both cases. The individual enzymes having multi-modular structures possess a CM and cellulose-binding module (CBM). The CBM binds to cellulose while the CM hydrolyzes the substrate. These CBMs may be linked at either N-terminus or C-terminus of the CM by a linker, which usually rich in Ser/Thr/Pro. Fungal CBMs are much smaller (32-36 residues) than the CBMs found in bacterial cellulases (90-100 or 130-172 residues). CBMs display varying properties giving a vast diversity in substrate binding affinities and specificities. Some of these binding modules bind to crystalline cellulose whereas others are limited to amorphous substrate. This binding may be reversible or irreversible (Carrard, G. *et al.*, 2000). CBMs consist of a β-sheet core, with aromatic residues (two or more, usually three) across one face making a platform for face-to-face stacking with glucose residues in crystalline cellulose. These aromatic residues are believed to interact with every second glucose residue on a single cellulose strand (*i.e.* on one side of each cellobiose unit) (Raghothama, S. *et al.*, 2000).

2.4.2 Cellulosome

The cellulosome, a multi-enzyme complex, plays a distinct role in degrading insoluble forms of cellulose. This complex is composed of a battery of subunits with interacting functional domains, both enzymic and non-enzymic, non-covalently linked. Scaffoldin subunits of cellulosome play a central role in organizing cellulolytic subunits into a multi-enzyme complex. This is achieved by the interaction between cohesin domains on scaffoldin and the dockerin domain on each enzymatic subunit; thus this interaction defines the cellulosomal structure. Other functions of scaffoldin may include cellulose binding and anchoring of the complex on to the cell surface. The anchoring of the cellulosome to cellulose is most often achieved through a CBM. Multiple copies of cohesin domains

integrate catalytic subunits into the complex via their dockerin domains (Bayer, E.A. *et al.*, 1998).

2.4.3 Chitin binding

Representative plant chitin binding modules (CtBMs), known as hevein, are composed of 30-43 amino acid residues including 8 cysteines, three aromatic residues and glycines (Beintema, J.J., 1994). Invertebrate CtBM (from Japanese horseshoe crab, *Tachypleus tridentatus*), tachycitin, has 73 residues, 5 disulfide bridges and aromatic residues in similar manner to plant CtBM (Suetake, T. *et al.*, 2000). These aromatic residues are involved in the platform that binds chitin to the binding module. There are also reports on a different CtBM from coconut rhinoceros beetle, *Oryctes rhinoceros*, which is known as scarabaecin. Although scarabaecin shows no similarity to other CtBMs, it has shown strong antifungal activity against fungal phytopathogens (Hemmi, H. *et al.*, 2003).

Early reports (Iseli, B., Boller, T. & Neuhaus, J.M., 1993), showed that the chitin binding module possesses chitin binding properties that are helpful in increasing the activity of the enzyme. This study showed further that the same enzyme without a binding domain had low affinity for its substrate, and that the effectiveness is three times higher in the enzyme with binding module. These observations suggest that binding and CMs act independently, but that the binding modules make the enzyme more effective by keeping it closer to its substrate. In another study (Arakane, Y. *et al.*, 2003), it has been reported for *Manduca sexta* chitinase that CM and CtBM can function separately. However, it was obvious that when the two modules were linked together, activity toward the insoluble chitin polymer was enhanced. Further, the CtBM plays a vital role for the effectiveness of antifungal activity but it is not necessary for the antifungal or catalytic activity (Iseli, B. *et al.*, 1993). However, (Broekaert, W.F. *et al.*, 1989) have reported that the chitinase-free chitin binding lectin from stinging nettle rhizome (*Urtica dioica* L.) inhibits fungal growth.

2.4.4 Linker

Most of these GHs have a SBM connected to the CM by a linker. There are no structures of the linker alone or together with binding and CMs known for chitinases or cellulases. It has been speculated for *Manduca sexta* (tobacco hornworm) chitinase (Figure 7) that binding modules and CM can act independently and are functional separately, although the linking of the two modules clearly allows the CM to gain better access to substrate. Binding modules bind to the solid substrate giving the possibility for the CM to act nearby. The module-connecting linker is protected from proteolytic cleavage possibly by its heavy glycosylation, which may also act in facilitating protein secretion and stabilizing the enzyme (Arakane, Y. *et al.*, 2003).

These linkers are Ser/Thr/Pro rich. By analogy to the linkers known for cellulases, chitinase linkers are long and flexible enough for allowing the modules to orientate in a way needed to gain the maximum efficiency (Carrard, G. *et al.*, 2000).

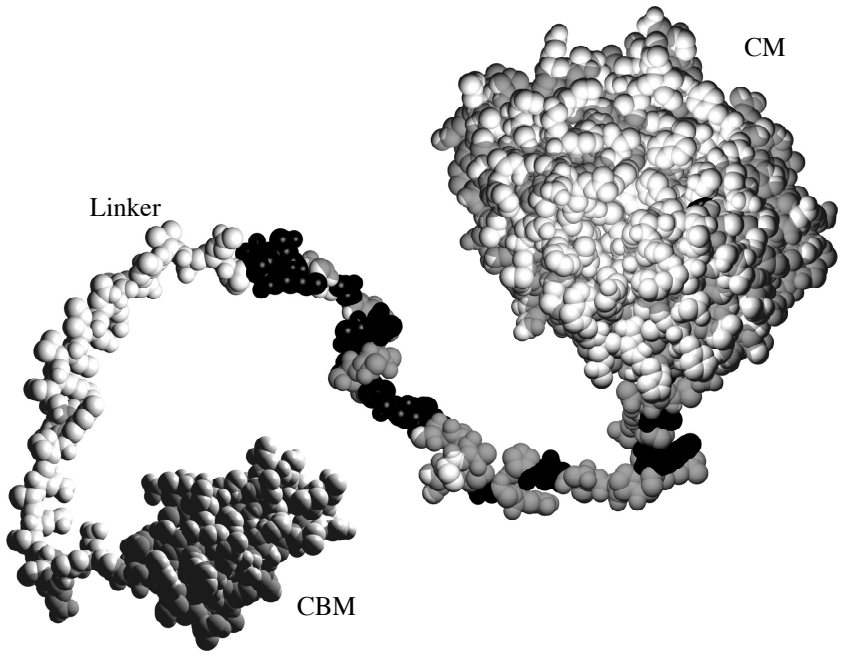

Figure 7. Homology models of *Manduca sexta* chitinase catalytic module and chitin binding module with a fantasy model of the linker showing the possible glycosylation sites in black

2.4.5 Vacuolar targeting

There are reports indicating that at the C-terminus of several family 19 chitinases, an extension of amino acid sequence represents a vacuolar targeting signal; this signal could be cleaved off post-translationally (Beintema, J.J., 1994). The absence of such a C-terminal extension could be correlated with extracellular targeting; *i.e.* class II and class IV chitinases. Therefore, these extracellular (unlike intravacuolar) chitinases could be a defensive tactic against fungal pathogens (Hamel, F. *et al.*, 1997). These chitinases also have a signaling function for releasing elicitors from either invading fungal hyphae or glycolipids present in the cell walls during development.

2.5 Glycoside hydrolase catalytic mechanisms

There are two mechanistic classes found among the glycoside hydrolases, namely inverting and retaining (Figure 8 shows the schemes of the two mechanisms.).

The inverting enzymes hydrolyze the glycosidic bond with net inversion of the anomeric configuration whereas the others act with net retention. Inverting glycosidases act via a single/direct displacement step by replacing the leaving group with water. Retaining glycosidases utilize a double displacement generating a covalent intermediate, *i.e.* a glycosyl-enzyme complex. In both these types of reaction, the catalytic center comprises a pair of carboxylic acids.

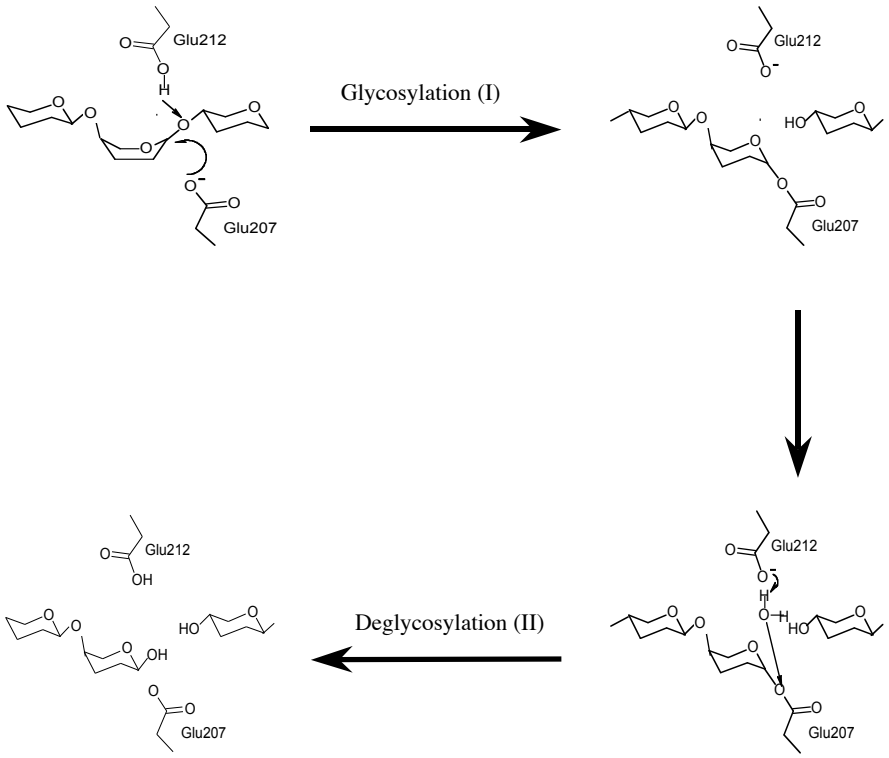

Figure 8A. Proposed retaining mechanism for Pc_Cel7D

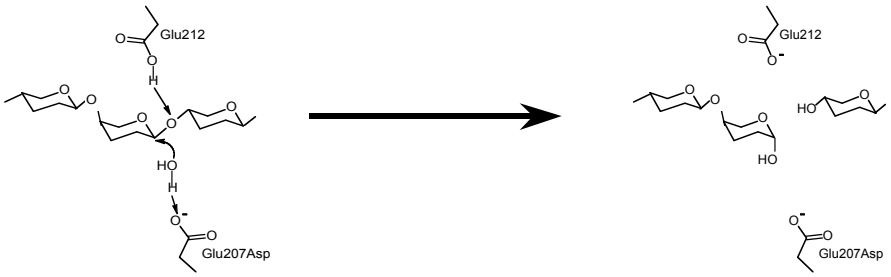

Figure 8B. Inverting catalytic mechanism using Pc_Cel7B as an example

The carboxylic acid residues in the active center of inverting enzymes act as a general acid and a general base, whereas in the case of retaining enzymes, one functions as the general acid/base while the other acts as a nucleophile/leaving group. Both these enzyme types function via transition states with oxocarbenium ion-like character. The active sites of these enzymes are located in clefts or tunnels. A glutamate or an aspartate appears to function as the catalytic nucleophile in all cases and in most cases such acidic residues also function as a general acid/base. There are number of other polar residues hydrogen bonding to catalytic residues,

probably modulating their charge states. In the retaining enzymes, the general acid/base and the nucleophile are located on opposite sides of the active site cleft. One of the two residues involved in acid catalysis is believed to support by other anionic residue by providing further electrostatic stabilization during the transition state. The architecture of the catalytic center of the inverting enzymes is similar to that of the retaining enzymes with two carboxylic acids on opposite sides of the substrate-binding cleft. Closer observation reveals that the distance between catalytic residues of inverting enzymes (average 9.0 – 9.5Å) are significantly larger than that of retaining enzymes (average 4.8 – 5.3Å). The greater distance in inverting enzymes is possibly due to the need for positioning a water molecule below the substrate in-between the catalytic residues (McCarter, J.D. & Withers, S.G., 1994). The two mechanisms, inverting and retaining, are quite distinct while having significant similarities as described above. Interestingly, the similar transition states have allowed the possibility of converting a retaining enzyme (*e.g.* a retaining β-glucosidase from *Agrobacterium faecalis*) to an inverting enzyme by mutating the nucleophile, Glu358Ala (Wang, Q.P. *et al.*, 1994).

The natural substrates of many glycoside hydrolases are polymers. Therefore, in principle, it is possible to distinguish the enzymes according to the preference of the substrate cleavage. Some enzymes cleave somewhere within the polymeric chain and others at the ends. The former are named "*endo*" whereas the latter is "*exo*". As the ends of the polysaccharide chains are nonequivalent, there are two types of "*exo*" enzymes, that is, ones that act beginning from the reducing or non-reducing ends of the sugar chain. A very well known property of such enzymes is that retaining enzymes are *exo* and inverting enzymes are *endo* (Davies, G., Sinnott, M.L. & Withers, S.G., 1998). The sugar binding subsites in glycoside hydrolases are numbered from the cleavage site labeling from $-n$ to $+n$ (where n is an integer). Negative numbers ($-n$) represents the non-reducing end of the sugar chain and positive numbers ($+n$) show displacement towards the reducing end, with the cleavage site located between the -1 and $+1$ subsites (Davies, G.J., Wilson, K.S. & Henrissat, B., 1997).

3 Results and discussion

3.1 *Phanerochaete chrysosporium* cellobiohydrolases

3.1.1 Pc_Cel7D structure and function (Paper I)

3.1.1.1 Pc_Cel7D introduction

The major cellulase produced by the white-rot fungus *P. chrysosporium* in a minimal medium is Pc_Cel7D. This enzyme is classified into glycoside hydrolase family 7 and acts with net retention of the anomeric carbon configuration. Pc_Cel7D was prepared for crystallization after papain cleavage of the intact protein and deglycosylation (Muñoz, I.G. *et al.*, 2001). Then, the CM was crystallized (Figure 9) (Muñoz, I.G. *et al.*, 2001) and the apo enzyme structure solved by molecular replacement using *T. reesei* cellobiohydrolase (Tr_Cel7A) as the search model (Muñoz, I.G. *et al.*, 2001).

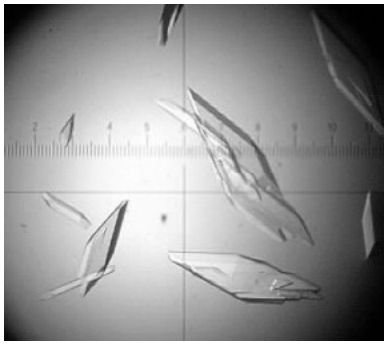

Figure 9. Pc_Cel7D crystals

Pc_Cel7D CM is a globular, elongated structure with approximate dimensions of 62Å x 41Å x 48Å. It consists of a characteristic glycoside hydrolase family 7 cellulase fold, *i.e.* a β-sandwich composed of two large, mainly antiparallel, β-sheets packed onto each other. On one face of the sandwich, a long cellulose-binding site is defined by loops. The catalytic residues are Glu207, Asp209 and Glu212. Substrate binding is expected to be similar to Tr_Cel7A, although the tunnel/cleft opening is slightly wider due to some deletions in the tunnel forming loops. Superimposition of Tr_Cel7A structures allowed docking of modeled oligosaccharide chain into Pc_Cel7D CM to map the cellulose binding along the 50Å long active site tunnel (Muñoz, I.G. *et al.*, 2001), which was now supported by ligand-bound structures. There are subsites to accommodate nine glycosyl units in the tunnel, which are numbered –7 to +2 (numbering starts from the point of glycosidic bond cleavage, between -1 and +1; negative numbers indicate the non-reducing end of the cellulose chain, and positive numbers, the reducing end (Davies, G.J. *et al.*, 1997). Subsites +1 and +2 are the product sites, which bind the product, cellobiose, after cleavage at the reducing end of the chain. These sites are placed at the end or exit of the binding site cleft/tunnel where the disaccharide is released. Product inhibition is a common observation for cellobiohydrolases.

This study covers the structures of Pc_Cel7D in complex with disaccharides, the product (cellobiose) and two inhibitors (lactose and cellobioimidazole), as well as a structure obtained in the presence of cellobiose, GG-S-GG, TRIS-HCl and calcium.

3.1.1.2 Overall structures

Deglycosylated Pc_Cel7D CM was crystallized and then single soaks with cellobiose, lactose and cellobioimidazole were performed. A double soak experiment was carried out with cellobiose followed by thio-linked sugar GG-S-GG. Complete data sets to 1.7Å or better were collected using synchrotron radiation. Clear electron density for all ligands, except GG-S-GG, was found prior to theirs inclusion in the models. All these structures are similar to the CM structure (PDB entry 1GPI), with an N-acetylglucosamine residue bound to Asn286 that remained after deglycosylation. One molecule of each bound ligand and a number of bound waters also could be found in the structures.

3.1.1.3 Binding of ligands to product site of Pc_Cel7D

Product inhibition is one of the characteristic features of Pc_Cel7D, which is supported by the observation of cellobiose bound to the +1/+2 (product) sites in the structure (Figure 10A). The non-reducing end of the disaccharide is in the +1 site with stacking interactions to Trp373. The hydrophobic β-face of the sugar makes non-polar contacts with Trp373. The interactions on the α-face are polar. The other residues make bonds to cellobiose in the product site are Asp248, Arg240, Glu212, Arg391 and Asp336. Both glycosyl rings show a regular $^4C_1$ chair that is the most favorable conformation in solution. The above-mentioned interactions lead to tight binding of disaccharide between the walls of the product site (in a sandwich type binding). Hydroxyls on one edge of the disaccharide point into the binding cleft making interactions with water while the hydroxyls in the other edge point towards the bulk solvent.

Lactose binding to Pc_Cel7D is similar to that of cellobiose with a few differences. This disaccharide is also an effective competitive inhibitor of Pc_Cel7D although it has a different configuration at C4 in the galactosyl unit in the +1 site compared to the corresponding glucosyl unit of cellobiose. The orientation of the hydroxyl group at C4 is equatorial in cellobiose whereas it is axial in lactose, and in the latter case it is making a direct hydrogen bond to Arg240. This allows Arg240 to make favorable interactions with Gln172 and Asp248, as well. The glucosyl unit in the +2 site has the same pattern of interactions as described for cellobiose at this position. However, the electron density of lactose and the immediate area in the protein is significantly better, with lower temperature factors than observed for cellobiose. These observations provide a structural basis for the fact that lactose is a stronger inhibitor of Pc_Cel7D than cellobiose.

Cellobioimidazole was used as an inhibitor due to the presence of a charge distribution mimicking the transition state of some exoglycosidase reactions and an sp^2-hybridized anomeric centre. This disaccharide was bound to the +1/+2 sites

of Pc_Cel7D in a similar manner to both cellobiose and lactose. The major difference is that cellobiose is shifted more than 2Å along the cleft towards the catalytic centre. The glucosyl unit in the +1 site makes stacking interactions with Trp373. It maintains hydrogen bonding to Asp248, His223, Asp209 and Glu212. At the +1 site, the glucosyl unit of cellobioimidazole adopts a regular $^4C_1$ chair conformation similar to cellobiose and lactose although some atoms of cellobioimidazole have been displaced. At the +2 site, the glucoimidazole ring cannot adopt a $^4C_1$ chair due to its C1-N5 double bond; the conformation is closest to an envelope form, with C3 out of plane. The glucoimidazole ring makes hydrogen bonds to Arg240, Arg391 and to several solvent molecules. Structural data such as electron density, temperature factors and good interactions with protein suggest that cellobioimidazole is an effective cellobiohydrolase inhibitor.

3.1.1.4 Binding of thio-linked substrate analogue GG-S-GG

Based on the premise that product-inhibited enzyme would have an empty substrate binding site with sufficient affinity for the substrate to bind, Pc_Cel7D crystals were soaked first with cellobiose then with the thio-linked sugar, methyl 4-S-β-cellobiosyl-4-thio-β-cellobioside $(Glc)_2$-S-$(Glc)_2$, in hopes of obtaining a complex that included sugars bound in both ends of the active-site cleft. The electron density in the product site clearly shows the bound disaccharide whereas the substrate sites are only partially occupied with the longer sugar. Electron density compatible with only TRIS, among the known crystallization reagents, was also found in the catalytic site, (Figure 10A).

3.1.1.5 Inhibition experiments

The discovery of TRIS in the active site leads us to carry out a systematic study of the components in the crystallization solution. Inhibition experiments showed that both 10 mM TRIS-HCl and 5 mM $CaCl_2$ individually and in combination inhibit Pc_Cel7D. No inhibition effect was observed with 10 mM NaCl, suggesting that TRIS and calcium ions are indeed the inhibiting species. TRIS or TRIS with calcium used in combination with 0.1 mM cellobiose gave a stronger inhibition, indicating cooperative binding.

3.1.1.6 Discussion of ligand binding

The Pc_Cel7D structural data have provided the background knowledge needed to understand why the binding of lactose is tighter than that of cellobiose, and why the opposite is the case in Tr_Cel7A. The complex with cellobioimidazole exhibited a distinct binding mode, leading us to look closely at binding data in the +1 and +2 sites in cellobiohydrolases generally. The ligands that bind to the product site place the non-reducing end of the disaccharide in the +1 site, and the reducing end in the +2 site. There were two clear binding modes seen when all of the data are compared. In the first mode the hexose in site +1 is closer to the active site centre, with a direct hydrogen bond to Glu212 (equivalent to 217 in Tr_Cel7A). The examples of the structures exhibiting this binding mode are the Pc_Cel7D/imidazole complex, and the complexes of Tr_Cel7A with IBTG, cellopentaose, and cellobiose + cellohexaose (1CEL, 6CEL, 7CEL). In the second mode, the sugar is shifted ~2 Å away from the catalytic centre leaving sufficient

room for a water molecule between sugar and the catalytic acid group. Examples for this type of binding are the Pc_Cel7D complexes with cellobiose and lactose, as well as complexes of cellobiose or cellotetraose with Tr_Cel7A (3CEL and 5CEL, respectively). The location of the 2-hydroxyl in the +1 site and the 6-hydroxyl in the +2 site is very similar in all structures. Therefore, the primary binding could be described as taking these points as key positions in release of the product. When the disaccharide moves away from the catalytic residues, the sugar in the +1 site moves out towards the bulk solvent and the sugar in the +2 site moves deeper into the cleft/tunnel of the enzyme.

To begin the catalytic action, first a cellulose chain threads into the enzyme. In the case of family 7 enzymes, the reducing end of the cellulose chain has to enter via the –7 or –8 subsite. To reach the catalytic site, the glycosyl units must pass through a number of subsites where each sugar unit binds in a particular position/ orientation. These bindings have to be weak so that the sugar unit will not sit in one subsite forever and instead can "slide" on to the next. For the cleavage to start, the sugar chain must be threaded all the way through the substrate-binding tunnel of the enzyme. During catalysis re-adjustment in glycosyl units binding might be needed.

The size of the tunnels of Pc_Cel7D and Tr_Cel7A are sufficient for a cellulose chain to thread through without any major changes, as there are no conformational changes observed between the apo enzyme and complexes. According to Divne *et al*. (1998), sugar-binding elements in the tunnel display the following features that support sliding of the substrate.

i.) Distribution of tryptophan-indole rings along the tunnel making large hydrophobic platforms
ii.) The hydrogen bonding network provided by the water molecules lining the tunnel
iii.) A smoothing of the total energy profile (energy profiles for the aromatic and hydrogen-bonding interactions)
iv.) A modeled "up-side-down" cellulose chain could give the same sort of interaction pattern.

Sliding is more critical in the beginning of the tunnel while the interactions in the subsites near the catalytic sites must be specific in order to assist in the catalysis. The product site clearly shows alternative binding modes, as illustrated by this work.

Family 7 enzymes utilize a double displacement mechanism involving a glycosyl-enzyme intermediate and net retention of the anomeric carbon configuration (McCarter, J.D. *et al.*, 1994). Based on the results with Pc_Cel7D, we propose that during substrate binding and catalysis there are two docking modes. One docking mode is the "cut" mode where the disaccharide unit is placed immediately at the active site, as observed for Pc-Cel7D complex with cellobioimidazole. The other will be referred as "slide" mode, where the

disaccharide is placed slightly further away, as seen for the cellobiose and lactose complexes.

In the process of threading, the cellulose chain has to be docked in the "cut" mode in order to be cleaved. This allows the O3 hydroxyl to point downwards to the deepest part of the cleft, making a direct interaction with Glu212-OE1. In this position, the hydroxyl lies within hydrogen-bonding range of the catalytically important Asp209 and His223. The other carboxylate oxygen of Glu212, OE2, makes a hydrogen bond to O4 of the glycosyl unit. In the real enzyme-substrate complex, this O4 hydroxyl represents the glycosidic oxygen that links the sugar in site +1 to that in site –1; hence this hydrogen bond between Glu212 and O4 indicates the protonation of the glycosidic oxygen in the transition state. The reducing end of the cellulose substrate (represented by the cellobiosyl unit in sites +1/+2) remains in the "cut" mode in the transition state where the glycosidic oxygen is bound to Glu212. As cellobiose clearly makes more favorable interactions in the "slide" mode, it is obvious that maximizing the transition-state stabilization is not a major consideration. As soon as the cellulose chain is cleaved, then the product takes on the "slide" mode of binding. The former positions of O3 and O4 hydroxyls are occupied by water molecules that make hydrogen bonds to O3 and O4. Therefore, the "slide" mode of docking favors the binding of product even after the cleavage. The water molecule that lies between O4 of the hexose in site +1 and the nucleophile Glu212 represents conditions in the intermediate, where the possibility to perform a nucleophilic attack on its anomeric carbon exists. Therefore, the structures suggest that the cleavage of the intermediate is followed by release of the product.

Classical understanding is that the difference between an EG and a cellobiohydrolase lies solely in the length of the loops that cover the substrate-binding groove. This difference could be visualized by superimposing Tr_Cel7A (PDB entry 8CEL), Tr_Cel7B (an EG) and Pc_Cel7D. In cellobiohydrolases, it is very clear that sugar-protein interactions (platforms for sugar binding) are highly preserved.

In the case of an EG, the carbohydrate interacting platforms are conserved near the catalytic site, *i.e.* in sites –2 to +1, but not towards the end of the tunnel. In the case of substrate binding sites –7 to –3, a few interacting residues are conserved, but most of the binding platforms are not preserved. A possible interpretation to this is that the cellulose binding is quite different in the two types of enzymes. During the substrate threading along the tunnel of cellobiohydrolases there is a gradual twisting of the cellulose chain until it is effectively turned upside-down at the exit of the tunnel compared to the entrance. However, an EG cleaves within the rather stiff cellulose chains, which are possibly fixed at both ends by the interactions of other cellulose molecules. Therefore, for a similar twisting within an EG's binding cleft, a second twist will also be needed to return the cellulose to the original orientation. This implies that more structural information is needed for a detailed understanding of the cellulose binding in family 7 EGs.

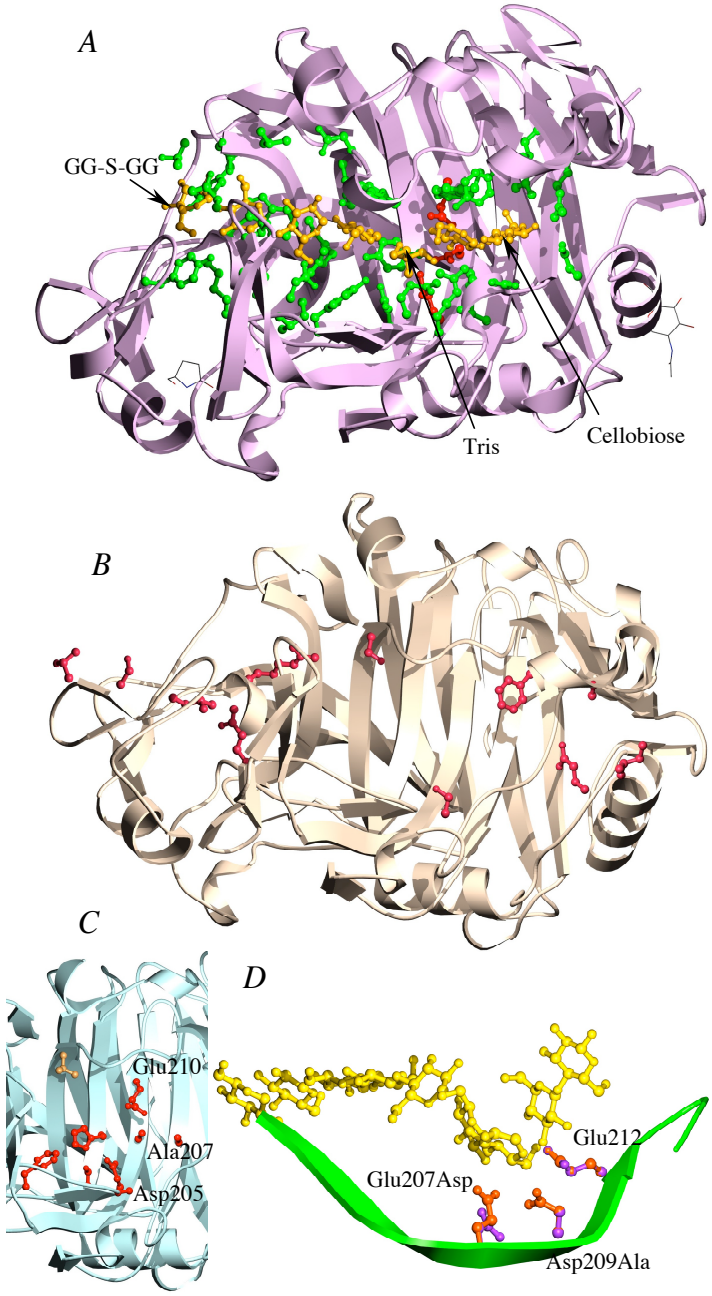

Figure 10(*A*). Pc_Cel7D complex with cellobiose and GG-S-GG; residues involved in catalysis and substrate binding are marked red and green, respectively. **(*B*)** Pc_Cel7A shows changes in substrate binding residues suggesting a difference in enzyme dynamics **(*C*)** & **(*D*)** Mutations in Pc_Cel7B catalytic residues and in the vicinity show a possible change in the catalytic mechanism. Panel ***D*** shows the mutations Glu207Asp and Asp209Ala providing sufficient space for a water molecule under the cellulose chain

Inhibition of Pc_Cel7D could be obtained by TRIS and calcium alone and in combination. Further, the inhibition data suggested that when the disaccharide cellobiose is included in the mixture, it binds in a synergistic fashion. As these reagents are included in the crystallization solutions, it is possible to observe TRIS in the active site of the complex with GG-S-GG. The catalytically stable tetrasaccharide, however, is observed at only partial occupancy (Figure 10*A*). TRIS is positioned near the nucleophile Glu207, Asp209 and catalytic acid/base Glu212 providing a clear explanation for its inhibition. Re-inspection of previous structures confirmed that TRIS is not present in the apo enzyme or the disaccharide structures. This observation suggests that the existence of some degree of synergy in binding of TRIS with the thio-linked sugar as well. As the catalytic site of Tr_Cel7A is similar to that of Pc_Cel7D, it is suggestive that TRIS could also bind to Tr_Cel7A and inhibit it. There are no earlier reports on TRIS inhibition of family 7 glycoside hydrolases, although unrelated proteins with similar catalytic sites are known to be inhibited by TRIS; *e.g.* the family 13 amylase (Brzozowski, A.M. *et al.*, 2000).

3.1.2 Other cellobiohydrolases of P. chrysosporium *(Paper II)*

3.1.2.1 Homology modeling

P. chrysosporium has six Cel7 enzymes expressed at different times at varying levels. The biggest puzzle to unfold is the reason for this differential expression (Vallim, M.A. *et al.*, 1998). In this work, the knowledge about the Pc_Cel7D structure gave us to gain a better understanding of the sequence data of the other isozymes via homology modeling. The folds of the enzymes are expected to be very similar due to the high sequence identity (66-83%), thus justifying the modeling based on Pc_Cel7D coordinates. The expected r.m.s. difference between the backbone atoms in the core regions of the real protein structures is <0.8 Å (Chothia, C. *et al.*, 1986). Many side-chain conformations could also be predicted with confidence, due to the similar side chains of the model sequences compared to those of the PcCel7D structure; others could be decided by considering the packing in the protein core. To understand the substrate binding in the various proteins, an oligosaccharide that has been modeled using a number of Tr_Cel7A structures was used (Divne, C. *et al.*, 1998). These docking studies provided the possibility to evaluate the effect of sequence similarities and differences on functional properties. The following discussion is based on the residue numbering of Pc_Cel7D.

The sequences and structures of Pc_Cel7C, Pc_Cel7E and Pc_Cel7F are closely related to Pc_Cel7D due to the high sequence similarity and the limited number of changes in the amino acids in the core regions. There are no insertions and deletions. The conserved catalytic residues suggest that the enzyme mechanism will be the same. However, there are some differences in the binding tunnel. These changes could lead to production of enzymes that slightly differ from Pc_Cel7D in dynamic properties.

There are relatively larger changes in Pc_Cel7A and Pc_Cel7B compared to each other and to Pc_Cel7D. Homology models of Pc_Cel7A and Pc_Cel7B are shown in Figure 10*B*, *C* and *D*, respectively.

Pc_Cel7A is different from the other isozymes in that the C-terminal part with the linker and CBM are missing. The expected result of this single deletion is a lower activity on solid substrates. The conserved catalytic residues suggest the same catalytic mechanism as Pc_Cel7D. Some changes in the residues that are important in substrate binding can have a considerable impact on substrate specificity or dynamics of this enzyme. The mutations Trp38Arg, Arg240Ser and Trp364Ser, are most interesting due to their possible interactions with substrate in Pc_Cel7D. Trp38 makes stacking interactions on glycosyl unit at the –4 site, but has been replaced by Tyr, Ser and Ile in EGs of two *Trichoderma* species, EGs from *Humicola insolens* and *Fusarium oxysporum*, respectively. Trp364 is conserved in all known Cel7 sequences, except Pc_Cel7A and Pc_Cel7B. The contribution of the stacking interactions of this tryptophan in the –2 site is expected to play a significant role in the substrate distortion in the –1 site during the catalysis (Divne, C. *et al.*, 1998). Therefore, this replacement must cause prominent changes in both binding and catalysis. Arg240 is a highly conserved residue among cellobiohydrolases and makes interactions in the +1 and +2 sites, although it is frequently missing in EGs (as well as in Pc_Cel7B). Mutations at Tyr47, Gly193, Tyr378 and Lys382 can be seen in cellobiohydrolases but replacements at Asn49, Tyr51, Lys178, Asn195 and His367 are mostly observed in EGs. Most of these changes can alter the cellulose binding and others may change the loops resulting in a relatively larger tunnel opening. In summary, Pc_Cel7A has some EG character. The changes also are suggestive that the substrate of this enzyme may not be cellulose but some soluble form of linear polysaccharide.

The homology model of Pc_Cel7B shows that three deletions in the sequence of Pc_Cel7B lie near each other in the folded structure, so creating a more open substrate-binding tunnel. This tunnel with a wider opening will be expected to give greater exposure in subsites between –2 and –5, suggesting the possibility of increased endo activity. The dynamics of the protein and its interactions with the substrate could also be affected by these deletions. The fourth deletion closer to the C-terminus (residues 429-430 of Pc_Cel7D) may not have any significant functional importance. The most interesting change in this enzyme is that there are mutations in catalytic residues, Glu207Asp and Asp209Ala (Figure 10*C* & 10*D*). These changes in highly conserved catalytic residues and others in the immediate neighborhood are certain to affect the properties of the enzyme dramatically. The distance between the putative catalytic residues of Pc_Cel7B is predicted to be ~7 Å, which is an intermediate value between those observed for an inverting enzyme (9-9.5 Å) and for a retaining enzyme (~5.5 Å) (McCarter, J.D. *et al.*, 1994). The shorter side chain of the aspartate that replaces the nucleophile (Glu207Asp) creates sufficient space for a water molecule under the anomeric carbon of the substrate in our model, suggesting a possible switching of the reaction to an inverting mechanism. However, a change in the catalytic mechanism within a glycoside hydrolase family is unprecedented, it is clear the experimental data will be needed

to resolve this issue. In another case, change of the nucleophile (of the β-glucosidase of *Agrobacterium*) from glutamate to aspartate reduced the catalytic rate by 2500-fold (Withers, S.G. *et al.*, 1992). Changes in the residues in the –1 site (at Tyr142, Asp170 and Trp364), which are important in binding and stabilizing the transition state, could also have considerable effects on catalytic properties. More space is created by the changes to smaller residues at Try168 and Asp176. Changes in the highly conserved (Tyr142, Tyr168, Asp170 and Trp364) and other residues near the active site can also have a major impact on the catalytic properties. The differences in the substrate-binding tunnel could, for example, be associated with a different specificity.

3.2 *Piromyces* sp. strain E2 Cel9A (Paper III)

Piromyces sp. strain E2 is an obligatory anaerobic fungus, which has been isolated from the feces of an Indian elephant. This type of fungus has a rich cellulolytic system and lives in the digestive tract of herbivores where it plays a significant role in the degradation and fermentation of plant cell wall materials. Crystalline cellulose degradation by this fungus occurs in a high molecular weight (hemi) cellulolytic complex, the cellulosome. This fungal cellulosome includes a number of catalytic components each linked to at least one copy of dockerin. The aim of this structural study was to support the process of identification of the major components of the *Piromyces* sp. strain E2 cellulosome (Steenbakkers, P.J. *et al.*, 2002).

3.2.1 Homology modeling

The CM of the *Piromyces* sp. strain E2 Cel9A (Figure 11*A* & *B*) enzyme was modeled using the appropriate portion of the *Thermobifida fusca* E4 endo/exo-1,4-glucanase (PDB entry 1TF4) as the template in the program SOD (Kleywegt, G.J. *et al.*, 2001). The model was modified in the graphics program O (Jones, T.A. *et al.*, 1991) using rotamers to improve the packing of the interior of the protein and to model insertions and deletions in loop regions. While models of such loops are very likely to contain errors, it is useful to have a visual reminder that the changes exist, color-coded to ensure appropriate caution is used in their interpretation. Likewise, homology models of the three dockerins were generated using the structure of the cellulosome dockerin module from *Piromyces equi* (PDB entry 1E8P).

3.2.2 Results and discussion

The mature protein of Cel9A was predicted to have a modular organization based on BLAST searches, with a CM belonging to glycoside hydrolase family 9. Therefore, to get a deeper understanding of the structure and function of the *Piromyces* sp. strain E2 CM, homology modeling was used. Cel9A showed highest sequence homology to the E4 enzyme from *T. fusca* but not to the family 9 enzymes from clostridial cellulosomes. In comparison with the *T. fusca* enzyme, residues 1-445 of the mature sequence were identified as belonging to the CM with high sequence similarity (45% identity).

The core regions of the structure and model are thus expected to be very similar, with backbone atoms matching with a root mean square difference of ~1.0 Å (Chothia, C. *et al.*, 1986). The structure of GH family 9 CM is an $(\alpha/\alpha)_6$ (Figure 11*B*) barrel having a shallow binding cleft running along one face. All catalytic residues are conserved. The *Piromyces* sp. strain E2 enzyme, similar to its counterparts in family 9, is expected to have inverting mechanism converting the configuration at the glycosidic bond of the substrate to its enantiomer in the product. Catalytic residues are Glu425 (acid for catalysis), Asp54 and/or Asp57 (base activating the catalytic water). It has a conserved calcium-binding site stabilizing two loops near the active site.

The comparison of the previously described six sub-sites in the *T. fusca* structure (Sakon, J. *et al.*, 1997) with those of the model provided additional information. Most residues that are important in substrate binding are conserved. . There can be some impact on the +2 and potential +3 sites due to a change near residue 390. Replacement of Trp256 by Gly257 removes an aromatic stacking interaction with the substrate, which is the basis for the –4 site of the *T. fusca* enzyme. A deletion (3 residues near 295) and an insertion (near 255, one residue) can also change the shape of the substrate-binding cleft near the –4 site. The *T. fusca* enzyme has rather open substrate-binding cleft, a general feature of EGs, although it shows a considerable amount of exo-activity. The probable reasons for this unusual property are the blocking of the cleft at the –4 end, the interaction with Trp256 and function of the cellulose binding modules (CBMs) that binding and orient substrate within the enzyme (Irwin, D. *et al.*, 1998). The *Piromyces* sp. strain E2 enzyme lacks the CBMs, the blockage and the tryptophan, and has no compensating loop extensions that could close the active site to give a tunnel effect. The other only available structure of a GH family 9 is *Clostridium thermocellum* EG, CelD (Juy, M. *et al.*, 1992) which is distantly related to the *T. fusca* enzyme (identity ~ 23%). This enzyme has an open active-site cleft and lacks the aromatic residue in the –4 site as found for the *Piromyces* sp. strain E2 Cel9A; it is presumably an EG. These observations lead us to suggest that Cel9A from *Piromyces* sp. strain E2 is also an EG.

3.2.3 Auxiliary modules in Cel9A

The Cel9A CM is followed by a domain with unknown function, then by a serine/threonine-rich linker; three fungal dockerin domains are found at the C-terminal end of the mature protein. The domain with unknown function has no detectable sequence identity to structure known. There is significant similarity (57%) to one GeneBank entry, that of a partial sequence of *Orpinomyces* sp. PC-2 designated as CelJ (Steenbakkers, P.J. *et al.*, 2001). Secondary structure prediction suggested that this region is likely to have some form of beta-sheet structure.

The three C-terminal modules placed after the linker are designated as dockerin I, II and III. Fungal dockerins are classified into three sub families based on the number, location and sequence context of cysteine residues (Steenbakkers, P.J. *et al.*, 2001).

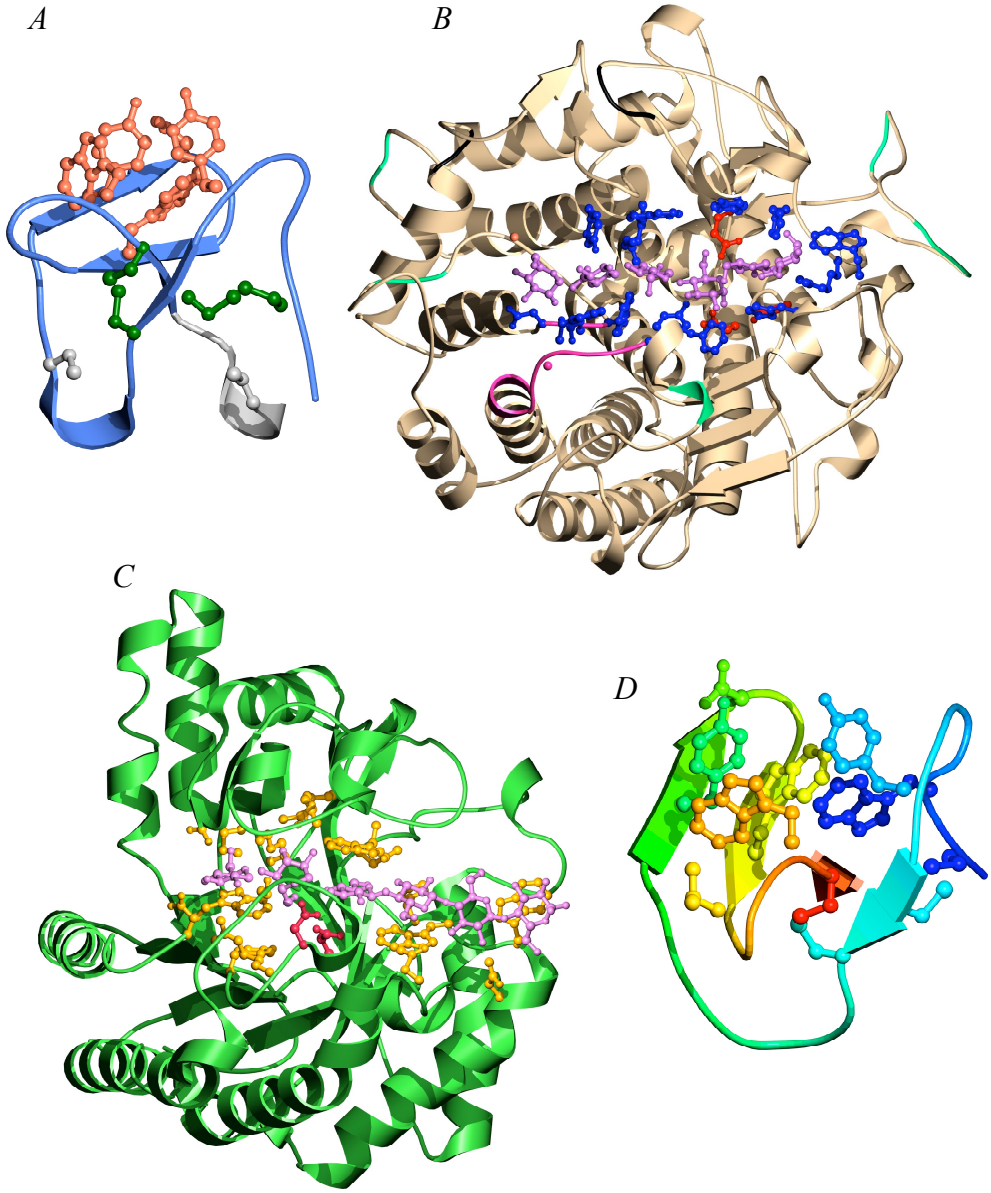

Figure 11. Homology models of *Piromyces* sp. E2 Cel9A and Cel6A and *Piromyces equi* Cel6A cellulosome components **(*A*)** Dockerin I of Cel9A shows disulfide forming residues in green and conserved scaffoldin binding side chains in tomato color **(*B*)** Cel9A catalytic module; modeled cellulose chain in medium orchid, catalytic residues in red, cellulose binding residues in blue, calcium binding site in deep pink, insertions in spring green and deletions in black **(*C*)** Catalytic module of *Piromyces* sp. E2 Cel6A; conserved amino acid side chains are shown in light coral, catalytic residues in crimson and the modeled cellulose chain in orchid **(*D*)** Dockerin of *Piromyces equi* Cel6A; conserved scaffoldin binding residues modeled as ball-and-stick side chains. Rainbow coloring blue to red is from the N-terminus to C-terminus

Accordingly, dockerin I classifies as type 2 (with 4 cysteine residues) and dockerins II and III as type 3 dockerins (with 6 cysteine residues). The *Piromyces equi* fungal dockerin type 2 structure (Raghothama, S. *et al.*, 2001) showed the disulfide bridges formed by these four cysteines, which play a major role in binding the fungal scaffoldin protein. This *Piromyces equi* fungal dockerin structure allows us to model the Cel9A dockerins (identities to 1E8P of the various dockerins are I-52%, II-35%, III-47%); a representative of these models is shown in Figure 11*A*. As dockerins II and III are type 3, they are slightly different than the template. The residues important in scaffoldin binding are placed on the surface and are totally conserved. The deletions in the loops of the dockerin are not expected to affect the binding surface. Some other residues are conserved due to structural reasons such as disulfide bonds *etc.* The number of disulfide bonds in Cel9A dockerins follows the dockerin classification criteria, and homology models help to extract this information. Two exposed cysteine residues in dockerin II and III seem likely to form a pair of inter-domain disulfide bonds (*e.g.* between dockerins II and III, or between two molecules of Cel9A). However, some anaerobic fungal enzymes have only a single copy of this dockerin type, which suggests that the inter-domain disulfides are not an essential property for the function. These observations confirm the significant contribution of cysteines in dockerins.

Thus the homology modeling of *Piromyces* sp. strain E2 enzyme shows that it includes a GH family 9 CM at the N-terminus that probably functions as an EG, and three fungal dockerin domains at the C-terminus, a domain in-between that has unknown function.

3.3 *Piromyces* sp. strain E2 Cel6A (Paper IV)

Cel6A enzymes of *Piromyces* sp. E2 (Cel6A(e2)) (Figure 11) and *Piromyces equi* (Cel6A(pe)) also have modular structures, and form part of the cellulosome. Thus, in addition to the CMs, both enzymes contain dockerin domains.

3.3.1 Homology modeling

Homology models of the CMs of *Piromyces* sp. E2 and *Piromyces equi* Cel6A enzymes were generated using two representative structures of the CM of *T. reesei* Cel6A (Zou, J. *et al.*, 1999) (PDB entry 1QJW and 1CB2, respectively) as templates in the programs SOD (Kleywegt, G.J. *et al.*, 2001) and O (Jones, T.A. *et al.*, 1991). Similarly, homology models of the non catalytic dockerin domains (dockerins) were generated using the sequence and structure of a Cel45A-linked dockerin module from *Piromyces equi* (PDB entry 1E8Q (Raghothama, S. *et al.*, 2001)).

3.3.2 Cel6A Results and discussion

Both enzymes, Cel6A(e2) (Figure 11*D*) and Cel6A(pe), include two dockerins at the N-terminus followed by the CMs. A short linker connects the dockerins. These dockerin sequences are highly similar to each other, Cel6A(e2) dockerins having 67% identity to each other whereas there is 84% identity between the two of

Cel6A(pe). These dockerins can be classified into type 2 (Steenbakkers, P.J. *et al.*, 2001). Viewing sequence information in the context of the available structural data was done using the homology models of dockerins generated based on the NMR structure of the dockerin lying at the N-terminal end of the Cel45A enzyme from *Piromyces equi* (Raghothama, S. *et al.*, 2001). This structure was a good template as the sequence identity is 50-60%. The conservation of disulfide bridges, hydrophobic residues in the core of the molecules and surface residues thought to interact with scaffoldin are very high. All insertions and deletions occurred in one loop. It is thus expected that there is no significant difference between the bindings of the various dockerins to the cellulosome. The accumulated effects of binding of closely connected dockerins in tandem to the multi-enzyme complex could be stronger.

There is a 82% sequence identity between the CMs of Cel6A(e2) (Figure 11*C*) and Cel6A(pe) with no insertions or deletions. Therefore, the following discussion is only based on Cel6A(e2), but could be applied to Cel6A(pe) as well.

These C-terminal units are similar to the CMs of other GH family 6 cellobiohydrolases. Therefore, homology models were generated using the *T. reesei* Cel6A structures that were most similar according to hidden Markov model-based searches (sequence identity 42-44%). The model obtained is expected to be similar to the true structure with an approximate error of only ~1 Å in core regions (Chothia, C. *et al.*, 1986).

The CM of *Piromyces* Cel6A (Figure 11*C*) is a single domain with a distorted α/β barrel topology; a long cleft containing the active site is located at the C-terminal end of the parallel β-strands that make up the barrel. The connecting loops enclose the tunnel as is needed for the processivity of the enzyme action.

T. reesei Cel6A is a cellobiohydrolase, that similar to Pc_Cel7D, it produces cellobiose as the main product of its reaction. The *Piromyces* Cel6As are much closer to cellobiohydrolases (amino-acid identity ~43%) than to the EGs (~25% amino-acid sequence identity). The small changes due to the insertions/deletions in the *Piromyces* sequences may not have much effect on the substrate binding. There is no difference in the four substrate-binding sub-sites at the entry side of the tunnel and two product binding sub-sites (for cellobiose). All essential residues are conserved along with the catalytic residues, Asp243 (proton donor) and Asp427 (catalytic base). The overall observations strongly suggest that the *Piromyces* enzymes will also be processive cellobiohydrolases. These enzymes are expected to show classical single-displacement (inverting) mechanism (Sinnot, M.L., 1990).

3.4 *Brassica juncea* chitinase

Brassica juncea is a representative of the genus of most common leafy vegetables consumed in East Asia. Chitinase investigations have been prompted by the susceptibility of *Brassica* species to fungal pathogenic attacks. Isolation of the chitinase cDNA revealed that this enzyme is structurally distinct from earlier

known examples, due to the presence of two chitin-binding modules. The expression of this chitinase is induced by wounding, methyl jasmonate treatment, *Aspergillus niger* infection and *Pieris rapae* feeding, showing their role in defense. This unusual chitinase was expressed in *Pichia* recombinantly, with both binding modules (BjCHI1), with only one binding module (BjCHI2) and without any binding module (BjCHI3) (Fung, K.L. *et al.*, 2002; Tang, C.M. *et al.*, 2004; Zhao, K.J. *et al.*, 1999). Knowledge on the action of the enzymes in this whole family (glycoside hydrolase family 19) is very poor due to the unavailability of an effective inhibitor, and the absence of any structures with bound carbohydrates.

3.4.1 Model of the catalytic module (Paper V)

3.4.1.1 Homology modeling

Brassica juncea chitinase CM has 60% identity to the barley seed chitinase sequence. A homology model of BjCHI3 was thus built based on the barley chitinase structure (PDB entry 2BAA) using SOD (Kleywegt, G.J. *et al.*, 2001) and O (Jones, T.A. *et al.*, 1991).

3.4.1.2 Model and mutant designing

The homology model of BjCHI3 assisted in the design of mutations that test the role of various residues in its function. All of the amino acid residues chosen for site-directed mutagenesis lie within the cleft where the active site is located (Figure 12). Two conserved glutamate residues (equivalent to Glu212 and Glu234 of BjCHI3) are postulated to play the main role in the reaction (Hart, P.J. *et al.*, 1995; Holm, L. & Sander, C., 1994). Single amino acid substitutions were performed for residues His211, Glu212, Glu234, Tyr269, Glu349, Asp360 and Arg361 (numbered according to the intact pre-protein of BjCHI1). These mutations were designed to investigate the importance of every well-conserved charged residue in the entire active-site cleft and in the immediate vicinity. Among them, His211, Glu212, Glu349 and Arg361 are completely conserved in classes I, II and IV of family 19 enzymes whereas Glu234 is ~90% conserved. There are no earlier reports of mutation studies of residues His211, Glu349 and Arg361 in any family 19 enzyme.

Chitinase assay results showed that the Glu212Ala chitinase mutant has the largest reduction in the activity. No activity was observed for this mutant within the detection limits of the assay. The mutant of His211Asn showed 9% activity, while the Arg361Ala substitution had similar effects, with 7% activity remaining. Glu234Ala was a less detrimental mutation with 36% of wild-type activity. Mutation of Tyr269Asp gave an enzyme with 50% activity. Compared to the other mutants, Glu349Ala and Asp360Ala showed the highest remnants of wild-type activity, 73% and 68%, respectively, denoting the lesser importance of these residues.

In family 19 structures, the distance between the relevant atoms equivalent to Glu212 and Glu234 of BjCHI3 is ~9 Å as is generally observed for inverting enzymes. The distance between the catalytic residues observed for retaining enzymes is ~5.5 Å (Henrissat, B. *et al.*, 1995).

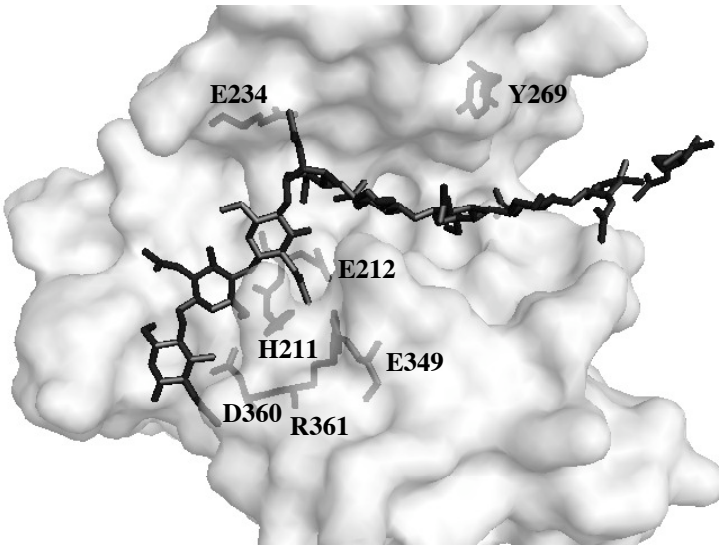

Figure 12. Surface diagram of the catalytic cleft of BjCHI3 homology model is shown with the indicated amino acid side chains for mutation studies in sticks. Modeled chitin chain is docked according to the suggestions of earlier authors (Sasaki, C. *et al.*, 2003).

Therefore, Glu212 is thought to act as the general acid in the reaction while Glu234 functions as the general base that promotes the nucleophilic attack by water on the anomeric carbon of the sugar. Residue Glu234 is also thought to stabilize the charged intermediate. The mutant results support the proposed role of Glu212, but do not indicate a strong role for Glu234. The large reduction in the activity of mutants His211 and Arg361 is due to the adjacent positioning to the putative general acid (Glu212). Tyr269 is thought to be playing a role in substrate binding rather than catalysis. Glu349 and Asp360 seem to be less directly involved in substrate binding. Thus, this study has broadened the knowledge of the residues involved in the activity of the proteins in this family. The mutants generated will also be useful for further investigations of family 19 chitinase function.

3.4.2 Brassica juncea *crystal structures (Paper VI)*

3.4.2.1 Crystallization data collection and structure determination

Purified, deglycosylated *Brassica juncea* proteins were received from Dr. Mee-Len Chye's laboratory, The University of Hong Kong, as ammonium sulfate precipitates. To prepare for crystallization, these chitinases were dialyzed against 10 mM HEPES buffer, pH 7.0, at 6°C and concentrated.

Crystallization screening trials of *Brassica juncea* chitinases gave crystals in many different conditions. *B. juncea* chitinase CM (BjCHI3) crystallized (Figure 13) in the presence of 10% mono methyl PEG 5000, un-buffered 0.2M sodium acetate,

0.1M sodium acetate at pH 4.2. A 63% complete x-ray data set was collected without a cryo-protectant.

Figure 13. BjCHI3 crystals

The CM connected to a single chitin binding module (BjCHI2) gave crystals in three different conditions. A native x-ray data set of BjCHI2 was collected using the crystals grown in 10% mono methyl PEG 5000, un-buffered 0.2M sodium acetate and 0.1M sodium acetate at pH 4.2. Single- and double-soak experiments were performed with $HgCl_2$ alone, and then $HgCl_2$ followed by glycol chitin, respectively. Crystals used for single-soak experiments were grown in 50% mono methyl PEG 5000 and 10M LiCl at pH 6.6 in cacodylate buffer, and double-soaked crystals were grown in 50% mm PEG 5000 and 10M LiCl at pH 7.6 in HEPES. $HgCl_2$ and glycol chitin were dissolved in cryo-protectant solution that contained 30% PEG 5000 and 4M LiCl in 0.1M cacodylate buffer at pH 6.6. Mutant Glu234Ala of the CM BjCHI3 gave crystals in 20% PEG 3350 and 0.2M ammonium formate at pH 6.6. To collect the native data set, Paratone N was used as the cryo-protectant whereas in the case of the triple-soak of $HgCl_2$, cellobiose and glycol chitin the cryo-protectant solution mentioned above was used. In all above-mentioned Pc_Cel7D crystals, 2 μl of the respective ligand solution were added to the protein-mother liquor drop with crystals and after 10 minutes these crystals were picked up directly from the drop using ready-made cryo-loops (Hampton Research). These were then flash frozen by immersion in liquid nitrogen. In the case of chitinase soaks, crystals were transferred to the ion/ligands dissolved in cryo-protectant solution. Crystals were picked and frozen as mentioned above. X-ray data sets for all the structures were collected at 100 K at the different beam lines of the European Synchrotron Radiation Facility (ESRF), in Grenoble, France. These images were processed by MOSFLM (Leslie, A.G.W., 1992), then the reflections were scaled and merged with the programs of the CCP4 package (CCP4, 1994).

The structure was determined by molecular replacement (MR) using the program AMoRe (Navaza, J., 1994) for BjCHI3 native CM, using the *Brassica juncea* CM homology model described above in section 3.4.1 (Paper V) as the search model.

Then rigid-body and restrained refinements were carried out with REFMAC5 (Murshudov, G.N., Vagin, A.A. & Dodson, E.J., 1997). For the other Bj chitinase structures, MOLREP (Vagin, A. & Teplyakov, A., 1997) and PHASER (Read, R.J., 2001) were used as the programs for MR and the first BjCHI3 structure was used as a search model. The graphics program O was used for model building (Jones, T.A. *et al.*, 1991).

3.4.2.2 Overview of structures

The Bj CM is a bilobed structure with high α-helical content (~46% of the 244 amino acids are helical). There are three small strands arranged into a β-sheet. BjCHI3 structures are 60%, 59% and 56% identical at the amino-acid sequence level to previously known family 19 chitinase structures of barley (PDB entries 2BAA and 1CNS) and jack bean (1DXJ). Approximate dimensions of the BjCHI3 are 40Å x 50Å x 50Å (Figure 14). There are three disulfide bridges (Cys242 & Cys251, Cys350 & Cys382, Cys230 & Cys168) and three free cysteines. These disulfide bridges are located in loops, stabilizing them by providing rigidity to the structure. Statistics for the Bj structures are listed in appendix A. The BjCHI3 structure was solved in space group P1 with four molecules in the asymmetric unit.

Flexibility of loops in the various molecules in the BjCHI3 structure could be seen, probably due to differences in the crystal contacts, although it may have some functional significance (see below). Other structures were obtained from BjCHI2 and E234A mutant proteins. Structures of BjCHI2 were supposed to contain the CM plus the CtBM. SDS PAGE prior to the crystallization trials provided evidence for the correct sized band on the gel. However, in the structure, CtBM is missing, suggesting either that the CtBM had been cleaved off during the crystallization, or that it is disordered in the crystal. Therefore, all structures from BjCHI2, except one, show only the CM. The exception was a structure solved in space group $P2_12_12_1$ having one molecule in the asymmetric unit. This structure includes the CM as well as extra six residues at the N-terminus, which belong to the linker between the CtBM and the CM. In addition to this, there are six chloride ions bound to this structure. Five of them make interactions that stabilize the loops of the structures and water. Such interactions could be biologically relevant. The other BjCHI2 and BjCHI3 E234A mutant structures were solved in space group P21, having two molecules in each asymmetric unit in each case. Therefore, all the structures may be considered as representing BjCHI3, which includes only the CM.

The two molecules in the asymmetric unit of the BjCHI CM structure soaked with $HgCl_2$ followed by glycol chitin shows slightly differing conformational changes in several loops. The same sort of change is visible also for the non-soaked apo structures. A comparison of these conformational changes suggests a possible open-close mechanism during catalysis. The movements of loops in non-soaked structures are rather small compared to those in the soaked structures, and the largest movements could be observed in the double soaked structures with $HgCl_2$ and glycol chitin.

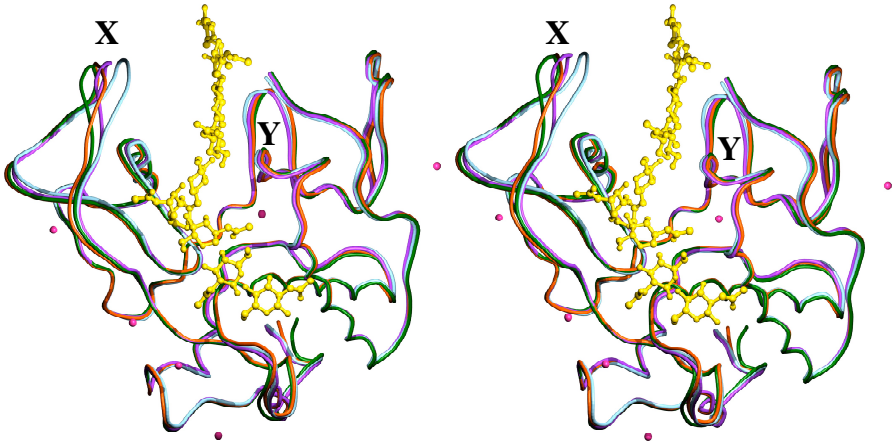

Figure 14. *Brassica juncea* chitinase crystal structures (orange-red: BjCHI3 apo structure A molecule, darkgreen: BjCHI3 apo structure D molecule; purple: BjCHI3 soaked with $HgCl_2$, sky blue: BjCHI3 E234A mutant soaked with $HgCl_2$, cellobiose and glycol chitin) showing loop movements upon soaking suggesting an open-close mechanism during the catalysis. Modeled chitin in gold is docked according to (Sasaki, C. *et al.*, 2003). The six chloride ions bound are shown as deep pink spheres

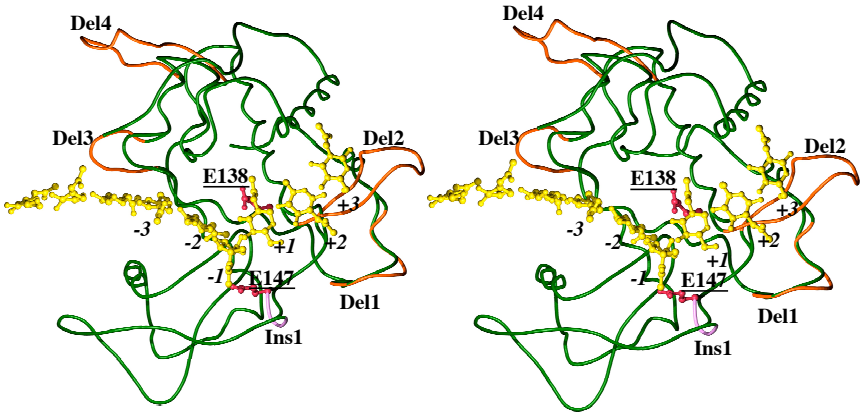

Figure 15. A stereo diagram of the homology model of *Dioscorea opposita* (yam) chitinase E. Catalytic residues Glu138 and Glu147 are shown in ball-and-stick in crimson. Predicted insertions (Ins1) and deletions (Del1-4) compared to barley chitinase structure are marked in violet and orange-red respectively. Modeled chitin chain is docked according to the suggestions by (Sasaki, C. *et al.*, 2003)

The flexibility of loops marked X and Y in the Figure 14 are clear, with movements towards the cleft, whereas in the apo structures loops maintain a more open position. This loop movement is largest in the BjCHI3 mutant E234A structure (Figure 14) that was soaked with cellobiose, $HgCl_2$ and glycol chitin. These inward movements, although small, suggest a possible function in the catalytic mechanism of the endochitinase. Usually, the cleft of a GH family 19 endochitinase is wide open, compared to the more restricted active sites of the family 18 enzymes. There are only three family 19 structures in the PDB, two for barley chitinase and the other for jack bean chitinase. All are in the apo form; no inhibitors or ligand bound structures are known. Further, no good inhibitors have yet been identified. One possible reason for these observations could be related to the large cleft. The interactions with the substrate inside the cleft could be weak in the open form. So, there is no strong grip on bound substrate, which must be an essential feature for endo acting enzymes. However, to facilitate the catalysis, substrate must be docked in a stable manner. This stability could be achieved through small conformational changes of the loops located either side of the cleft that would allow the chitinase to close down on the substrate once it is bound. The observed differences in the structures reported here could thus reflect features important in the function of family 19 chitinases.

3.5 Yam chitinase (Paper VII)

The amino-acid sequence of yam (*Dioscorea opposita* Thunb) chitinase was obtained after cloning and sequencing of the genomic DNA from leaves (Mitsunaga, T. *et al.*, 2004). The deduced amino acid sequence shows an identity of 50 to 59% to other class IV plant chitinases.

At the C-terminus of the sequence, an extra extension comprised of 8 amino acids could be seen after the cysteine that was reported to be the last amino acid for other class IV chitinases. This additional extension is thought to be important in subcellular localization.

3.5.1 Homology modeling

Yam chitinase is 50% identical to the structure of class II barley seed endochitinase. Therefore, using the barley seed chitinase structure (PDB entry 2BAA) as the template, yam chitinase was modeled using SOD (Kleywegt, G.J. *et al.*, 2001) and O (Jones, T.A. *et al.*, 1991).

3.5.2 Overview of the model

The yam chitinase model (Figure 15) offered new insights into the differences between the class I/II and class IV plant chitinases. The usefulness of this model is increased by the fact that there are no known structures of glycoside hydrolase family 19 class IV enzymes. The sequence identity observed suggests the error of the backbone atom positions in the core region of the model will be ~1 Å (Chothia, C. *et al.*, 1986); the structures of loop regions, especially those with insertions and deletions, have larger uncertainty. Placement of chitin was done according to the suggestions taken from the docking studies of class I chitinase

published elsewhere (Sasaki, C. *et al.*, 2003). Comparable analysis of chitin binding domain model was impossible due to the unavailability of similar structures. Chitinase classes II and IV have an inverting catalytic mechanism. All the residues lining the active site cleft are conserved, implying the substrate (chitin) binding pattern in class I/II and class IV to be the same. In comparison to class I/II structural data, the deletions in class IV chitinases clustered at the end of the substrate-binding cleft resulting a shorter cleft by approximately one glycosyl unit at each end. According to (Sasaki, C. *et al.*, 2003) class I/II chitinases have -2, -1, +1and +2 core subsites whereas binding of –3 and +3 subsites is weaker. Consequently, little loss in binding affinity with a shorter active site cleft may be expected.

Like other enzymes in this family, yam chitinase possesses endo-activity, *i.e.* it cleaves in the middle of the chitin chain (Koga, D., Yoshioka, K. & Arakane, Y., 1998). A shorter cleft of yam chitinase could be useful in recognizing shorter segments of the chitin chain, which is probably an advantage in its attack on a solid substrate such as the fungal cell wall. In other words, class IV chitinase can grasp a smaller section of exposed chitin chain on a fungal hyphal wall and begin its attack. Therefore, this can be advantageous for the plant's defense against pathogens and possibly be correlated with the biological observation of more effective lytic activity towards plant pathogen cell walls (Karasuda, S. *et al.*, 2003). There is at present no clear evidence for the effect of the N-terminal chitin binding module on its activity in the native setting. The C-terminal extension of the sequence cannot be modeled with any accuracy. However, its suggested role as a secretion signal might not require a special structural feature.

4 Concluding remarks

Comparative studies of the properties of Pc_Cel7D, Pc_Cel7C and Tr_Cel7A will help guide the industrial applications of *Phanerochaete* cellobiohydrolases. In future work, it will be interesting to obtain a structure with bound substrate having high occupancy in the substrate-binding sites. This is a difficult task at the moment due to the unavailability of mutants/ recombinant protein as it is produced in the natural fungal system. Recombinant Pc_Cel7D will, for example, allow the possibility of designing more stable enzymes. A comparative molecular biological and biochemical study of the Pc_Cel7 isozymes can lead to more utility industrially. Apart from the industrial applications such as saccharification, de-inking and pulp production, understanding of the enzymology will add to the current knowledge on carbon recycling and handling of the complex problems involved in degrading complex substrates by *Phanerochaete*.

In the case of *Piromyces* sp. E2 and *Piromyces equi* cellulases, they also will need to be expressed recombinantly to allow the necessary experiments to gain a clear understanding about the biochemistry and structural features of the proteins. It is important that the structural information such as boundaries of the modules should be utilized in the process of primer designing for cloning. The ability to make mutants would also open the opportunity for further structural and biochemical analysis to understand the catalytic processes.

The *Brassica juncea* CM homology model broadened out knowledge on the catalysis and was useful for mutant designing. The CM structures of apo enzyme and chloride ions bound complex together with the mutant allowed us to investigate the conformational changes during the catalysis, and to propose an open-close mechanism. In the future, the project will be focused on obtaining ligand bound structures, as well as structures of BjCHI1 and BjCHI2.

The yam (*Dioscorea opposita*) chitinase homology model suggested the functional significance of the large deletions in class IV chitinases that would make the substrate-binding cleft shorter while retaining good binding affinity. This needs to be tested experimentally, together with our speculation about the enzyme being able to grasp a smaller chitin chain. For a wider understanding of class IV chitinases, the crystal structure of this type of protein has to be determined. As the first step crystallization trials have to be performed with recombinant or isolated natural enzyme with and without the chitin-binding module.

Any of the above mentioned chitinase structures would extend the horizons of our knowledge on GH family 19 chitinases and increase our capacity to use these chitinases industrially (*e.g.* as a fungicide) and in the pathogen resistant plant production.

5 Reference

Arakane, Y., Q. Zhu, M. Matsumiya, S. Muthukrishnan & K.J. Kramer (2003). Properties of Catalytic, Linker and Chitin-Binding Domains of Insect Chitinase. *Insect Biochem Mol Biol* 33(6): 631-48.

Bayer, E.A., E. Morag, R. Lamed, S. Yaron & Y. Shoham (1998). Cellulosome Structure: Four-Pronged Attack Using Biochemistry, Molecular Biology, Crystallography and Bioinformatics. *Carbohydrates from Trichoderma Reesei and Other Microorganisms Structure, Biochemistry, Genetics and Applications*. M. Claeyssens, W. Nerinckx & K. Piens. The Royal Society of Chemistry. 39-65.

Becker, D., C. Braet, H. Brumer, 3rd, M. Claeyssens, C. Divne, B.R. Fagerstrom, M. Harris, T.A. Jones, G.J. Kleywegt, A. Koivula, S. Mahdi, K. Piens, M.L. Sinnott, J. Stahlberg, T.T. Teeri, M. Underwood & G. Wohlfahrt (2001). Engineering of a Glycosidase Family 7 Cellobiohydrolase to More Alkaline Ph Optimum: The Ph Behaviour of Trichoderma Reesei Cel7a and Its E223s/ A224h/L225v/T226a/D262g Mutant. *Biochem J* 356(Pt 1): 19-30.

Beintema, J.J. (1994). Structural Features of Plant Chitinases and Chitin-Binding Proteins. *FEBS Lett* 350(2-3): 159-63.

Boisset, C., C. Fraschini, M. Schulein, B. Henrissat & H. Chanzy (2000). Imaging the Enzymatic Digestion of Bacterial Cellulose Ribbons Reveals the Endo Character of the Cellobiohydrolase Cel6a from Humicola Insolens and Its Mode of Synergy with Cellobiohydrolase Cel7a. *Appl Environ Microbiol* 66(4): 1444-52.

Broekaert, W.F., J. Vanparijs, F. Leyns, H. Joos & W.J. Peumans (1989). A Chitin-Binding Lectin from Stinging Nettle Rhizomes with Antifungal Properties. *Science* 245(4922): 1100-1102.

Brzozowski, A.M., D.M. Lawson, J.P. Turkenburg, H. Bisgaard-Frantzen, A. Svendsen, T.V. Borchert, Z. Dauter, K.S. Wilson & G.J. Davies (2000). Structural Analysis of a Chimeric Bacterial Alpha-Amylase. High-Resolution Analysis of Native and Ligand Complexes. *Biochemistry* 39(31): 9099-107.

Carrard, G., A. Koivula, H. Soderlund & P. Beguin (2000). Cellulose-Binding Domains Promote Hydrolysis of Different Sites on Crystalline Cellulose. *Proc Natl Acad Sci U S A* 97(19): 10342-7.

Chothia, C. & A.M. Lesk (1986). The Relation between the Divergence of Sequence and Structure in Proteins. *Embo J* 5(4): 823-6.

Clarke, A.J. (1997). *Biodegradation of Cellulose Enzymology and Biotechnology*. Technomic Publishing Company, Inc., Lancaster, 1-21.

Collaborative Computational Project, Number 4 (1994). The Ccp4 Suite: Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography* 50: 760-763.

Davies, G., M.L. Sinnott & S.G. Withers (1998). Glycosyl Transfer. *Comprehensive Biological Catalysis a Mechanistic Reference*. M. Sinnott. Academic Press Limited. San Diego, I**:** 119 - 209.

Davies, G.J., K.S. Wilson & B. Henrissat (1997). Nomenclature for Sugar-Binding Subsites in Glycosyl Hydrolases. *Biochemical Journal* 321: 557-559.

Divne, C., J. Stahlberg, T.T. Teeri & T.A. Jones (1998). High-Resolution Crystal Structures Reveal How a Cellulose Chain Is Bound in the 50 Angstrom Long Tunnel of Cellobiohydrolase I from Trichoderma Reesei. *Journal of Molecular Biology* 275(2): 309-325.

Domard, A. (1996). *Some Physicochemical and Structural Basis for Applicability of Chitin and Chitosan.* In: Chitin and Chitosan Environmental friendly and versatile biomaterials; 2nd Asia Pacific Chitin Symposium, (W.F. Stevens, M.S. Rao & S. Chandrkrachang, City, 1-12.

(http://www.fibersource.com/f-tutor/cellulose.htm#chemistry; 2005.01.16).

Fung, K.L., K.J. Zhao, Z.M. He & M.L. Chye (2002). Tobacco-Expressed Brassica Juncea Chitinase Bjchi1 Shows Antifungal Activity in Vitro. *Plant Mol Biol* 50(2): 283-94.

Hamel, F., R. Boivin, C. Tremblay & G. Bellemare (1997). Structural and Evolutionary Relationships among Chitinases of Flowering Plants. *J Mol Evol* 44(6): 614-24.

Harhangi, H.R., A.C. Freelove, W. Ubhayasekera, M. van Dinther, P.J. Steenbakkers, A. Akhmanova, C. van der Drift, M.S. Jetten, S.L. Mowbray, H.J. Gilbert & H.J. Op den Camp (2003). Cel6a, a Major Exoglucanase from the Cellulosome of the Anaerobic Fungi Piromyces Sp. E2 and Piromyces Equi. *Biochim Biophys Acta* 1628(1): 30-9.

Hart, P.J., H.D. Pfluger, A.F. Monzingo, T. Hollis & J.D. Robertus (1995). The Refined Crystal Structure of an Endochitinase from Hordeum Vulgare L. Seeds at 1.8 a Resolution. *J Mol Biol* 248(2): 402-13.

Heiner, A.P. & O. Teleman (1998). On the Structure of Crystalline Cellulose I. *Carbohydrases from Trichoderma Reesei and Other Microorganisms Structures, Biochemistry, Genetics and Applications*. M. Claeyssens, W. Nerinckx & K. Piens. The Royal Society of Chemistry. 204-213.

Hemmi, H., J. Ishibashi, T. Tomie & M. Yamakawa (2003). Structural Basis for New Pattern of Conserved Amino Acid Residues Related to Chitin-Binding in the Antifungal Peptide from the Coconut Rhinoceros Beetle Oryctes Rhinoceros. *J Biol Chem* 278(25): 22820-7.

Henrissat, B. (1991). A Classification of Glycosyl Hydrolases Based on Amino-Acid-Sequence Similarities. *Biochemical Journal* 280: 309-316.

Henrissat, B. (1999). Classification of Chitinases Modules. *Chitin and Chitinases*. P. Jolles & R.A.A. Muzzarelli. Birkhauser. Basel**:** 137-156.

Henrissat, B., I. Callebaut, S. Fabrega, P. Lehn, J.P. Mornon & G. Davies (1995). Conserved Catalytic Machinery and the Prediction of a Common Fold for Several Families of Glycosyl Hydrolases. *Proceedings of the National Academy of Sciences of the United States of America* 92(15): 7090-7094.

Holm, L. & C. Sander (1994). Structural Similarity of Plant Chitinase and Lysozymes from Animals and Phage. An Evolutionary Connection. *FEBS Lett* 340(1-2): 129-32.

Igarashi, K., I. Momohara, T. Nishino & M. Samejima (2002). Kinetics of Inter-Domain Electron Transfer in Flavocytochrome Cellobiose Dehydrogenase from the White-Rot Fungus Phanerochaete Chrysosporium. *Biochem J* 365(Pt 2): 521-6.

Irwin, D., D.H. Shin, S. Zhang, B.K. Barr, J. Sakon, P.A. Karplus & D.B. Wilson (1998). Roles of the Catalytic Domain and Two Cellulose Binding

Domains of Thermomonospora Fusca E4 in Cellulose Hydrolysis. *J Bacteriol* 180(7): 1709-14.
Iseli, B., T. Boller & J.M. Neuhaus (1993). The N-Terminal Cysteine-Rich Domain of Tobacco Class I Chitinase Is Essential for Chitin Binding but Not for Catalytic or Antifungal Activity. *Plant Physiol* 103(1): 221-6.
Jones, T.A., J.Y. Zou, S.W. Cowan & Kjeldgaard (1991). Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models. *Acta Crystallogr A* 47 (Pt 2): 110-9.
Juy, M., A.G. Amit, P.M. Alzari, R.J. Poljak, M. Claeyssens, P. Beguin & J.P. Aubert (1992). 3-Dimensional Structure of a Thermostable Bacterial Cellulase. *Nature* 357(6373): 89-91.
Karasuda, S., S. Tanaka, H. Kajihara, Y. Yamamoto & D. Koga (2003). Plant Chitinase as a Possible Biocontrol Agent for Use Instead of Chemical Fungicides. *Biosci Biotechnol Biochem* 67(1): 221-4.
Kasprzewska, A. (2003). Plant Chitinases--Regulation and Function. *Cell Mol Biol Lett* 8(3): 809-24.
Kim, K.J., Y.J. Yang & J.G. Kim (2003). Purification and Characterization of Chitinase from Streptomyces Sp. M-20. *J Biochem Mol Biol* 36(2): 185-9.
Kleywegt, G.J., J.Y. Zou, M. Kjeldgaard & T.A. Jones (2001). Around O. *International Tables for Crystallography, Vol. F. Crystallography of Biological Macromolecules*. M.G. Rossmann & E. Arnold. Kluwer Academic. Dordrecht**:** 353-356,
366-367.
Koga, D., M. Mitsutomi, M. Kono & M. Matsumiya (1999). Biochemistry of Chitinases. *Chitin and Chitinases*. P. Jolles & R.A.A. Muzzarelli. Birkhauser. Basel**:** 111-123.
Koga, D., K. Yoshioka & Y. Arakane (1998). Hplc Analysis of Anomeric Formation and Cleavage Pattern by Chitinolytic Enzyme. *Bioscience Biotechnology and Biochemistry* 62(8): 1643-1646.
Leslie, A.G.W. (1992). Mosflm. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*(26).
Li, J. & J.-Y. Liu (2003). A Novel Cotton Gene Encoding a New Class of Chitinase. *Acta Botanica Sinica* 45(12): 1489-1496.
Li, W.L., J.D. Faris, S. Muthukrishnan, D.J. Liu, P.D. Chen & B.S. Gill (2001). Isolation and Characterization of Novel Cdna Clones of Acidic Chitinases and Beta-1,3-Glucanases from Wheat Spikes Infected by Fusarium Graminearum. *Theoretical and Applied Genetics* 102(2-3): 353-362.
McCarter, J.D. & S.G. Withers (1994). Mechanisms of Enzymatic Glycoside Hydrolysis. *Curr Opin Struct Biol* 4(6): 885-92.
Meins, F., B. Fritig, H.J.M. Linthorst, J.D. Mikkelsen, J.M. Neuhaus & J. Ryals (1994). Plant Chitinase Genes. *Plant Molecular Biology Reporter* 12(2): S22-S28.
Mitsunaga, T., M. Iwase, W. Ubhayasekera, S.L. Mowbray & D. Koga (2004). Molecular Cloning of a Genomic DNA Encoding Yam Class Iv Chitinase. *Biosci Biotechnol Biochem* 68(7): 1508-17.
Muñoz, I.G., W. Ubhayasekera, H. Henriksson, I. Szabo, G. Pettersson, G. Johansson, S.L. Mowbray & J. Ståhlberg (2001). Family 7 Cellobiohydrolases from Phanerochaete Chrysosporium: Crystal Structure of the Catalytic Module

of Cel7d (Cbh58) at 1.32 Angstrom Resolution and Homology Models of the Isozymes. *Journal of Molecular Biology* 314(5): 1097-1111.
Murshudov, G.N., A.A. Vagin & E.J. Dodson (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D-Biological Crystallography* 53: 240-255.
Muzzarelli, R.A.A. (1999). Native, Industrial and Fossil Chitins. *Chitin and Chitinases*. P. Jolles & R.A.A. Muzzarelli. Birkhauser. Basel**:** 1-6.
Nakazaki, T., Y. Tomimoto, H. Ikehashi, Y. Kowyama, M. Yano, K. Yamamoto & T. Sasaki (1997). A Novel Chitinase in Rice (Oryza Sativa L.) Detected from Husk Proteins and Its Gene Locus. *Breeding Science* 47(4): 363-369.
Navaza, J. (1994). Amore - an Automated Package for Molecular Replacement. *Acta Crystallographica Section A* 50: 157-163.
Neuhaus, J.M., B. Fritig, H.J.M. Linthorst, F. Meins, J.D. Mikkelsen & J. Ryals (1996). A Revised Nomenclature for Chitinase Genes. *Plant Molecular Biology Reporter* 14(2): 102-104.
Raghothama, S., R.Y. Eberhardt, P. Simpson, D. Wigelsworth, P. White, G.P. Hazlewood, T. Nagy, H.J. Gilbert & M.P. Williamson (2001). Characterization of a Cellulosome Dockerin Domain from the Anaerobic Fungus Piromyces Equi. *Nat Struct Biol* 8(9): 775-8.
Raghothama, S., P.J. Simpson, L. Szabo, T. Nagy, H.J. Gilbert & M.P. Williamson (2000). Solution Structure of the Cbm10 Cellulose Binding Module from Pseudomonas Xylanase A. *Biochemistry* 39(5): 978-84.
Read, R.J. (2001). Pushing the Boundaries of Molecular Replacement with Maximum Likelihood. *Acta Crystallogr D Biol Crystallogr* 57(Pt 10): 1373-82.
Robertus, J.D. & A.F. Monzingo (1999). The Structure and Action of Chitinases. *Exs* 87: 125-35.
Sakon, J., D. Irwin, D.B. Wilson & P.A. Karplus (1997). Structure and Mechanism of Endo/Exocellulase E4 from Thermomonospora Fusca. *Nat Struct Biol* 4(10): 810-8.
Sasaki, C., Y. Itoh, H. Takehara, S. Kuhara & T. Fukamizo (2003). Family 19 Chitinase from Rice (Oryza Sativa L.): Substrate-Binding Subsites Demonstrated by Kinetic and Molecular Modeling Studies. *Plant Mol Biol* 52(1): 43-52.
Sinnot, M.L. (1990). Catalytic Mechanisms of Enzymic Glycosyl Transfer. *Chemical Reviews* 90(7): 1171-1202.
Steenbakkers, P.J., X.L. Li, E.A. Ximenes, J.G. Arts, H. Chen, L.G. Ljungdahl & H.J. Op Den Camp (2001). Noncatalytic Docking Domains of Cellulosomes of Anaerobic Fungi. *J Bacteriol* 183(18): 5325-33.
Steenbakkers, P.J., W. Ubhayasekera, H.J. Goossen, E.M. van Lierop, C. van der Drift, G.D. Vogels, S.L. Mowbray & H.J. Op den Camp (2002). An Intron-Containing Glycoside Hydrolase Family 9 Cellulase Gene Encodes the Dominant 90 Kda Component of the Cellulosome of the Anaerobic Fungus Piromyces Sp. Strain E2. *Biochem J* 365(Pt 1): 193-204.
Stevens, W.F. (1996). *Chitosan: A Key Compound in Biology and Bioprocess Technology.* In: Chitin and Chitosan Environmental friendly and versatile biomaterials; 2nd Asia Pacific Chitin Symposium, (W.F. Stevens, M.S. Rao & S. Chandrkrachang, City, 13-21.

Suetake, T., S. Tsuda, S. Kawabata, K. Miura, S. Iwanaga, K. Hikichi, K. Nitta & K. Kawano (2000). Chitin-Binding Proteins in Invertebrates and Plants Comprise a Common Chitin-Binding Structural Motif. *J Biol Chem* 275(24): 17929-32.

Tang, C.M., M.L. Chye, S. Ramalingam, S.W. Ouyang, K.J. Zhao, W. Ubhayasekera & S.L. Mowbray (2004). Functional Analyses of the Chitin-Binding Domains and the Catalytic Domain of Brassica Juncea Chitinase Bjchi1. *Plant Mol Biol* 56(2): 285-98.

Truong, N.H., S.M. Park, Y. Nishizawa, T. Watanabe, T. Sasaki & Y. Itoh (2003). Structure, Heterologous Expression, and Properties of Rice (Oryza Sativa L.) Family 19 Chitinases. *Biosci Biotechnol Biochem* 67(5): 1063-70.

Uzcategui, E., G. Johansson, B. Ek & G. Pettersson (1991). The 1,4-Beta-D-Glucan Glucanohydrolases from Phanerochaete Chrysosporium. Re-Assessment of Their Significance in Cellulose Degradation Mechanisms. *J Biotechnol* 21(1-2): 143-59.

Vagin, A. & A. Teplyakov (1997). Molrep: An Automated Program for Molecular Replacement. *Journal of Applied Crystallography* 30: 1022-1025.

Vallim, M.A., B.J. Janse, J. Gaskell, A.A. Pizzirani-Kleiner & D. Cullen (1998). Phanerochaete Chrysosporium Cellobiohydrolase and Cellobiose Dehydrogenase Transcripts in Wood. *Appl Environ Microbiol* 64(5): 1924-8.

van Loon, L.C., W.S. Pierpoint, T. Boller & V. Conejero (1994). Recommendations for Naming Plant Pathogenesis-Related Proteins. *Plant Molecular Biology Reporter* 12(3): 245-264.

Wang, Q.P., R.W. Graham, D. Trimbur, R.A.J. Warren & S.G. Withers (1994). Changing Enzymatic-Reaction Mechanisms by Mutagenesis - Conversion of a Retaining Glucosidase to an Inverting Enzyme. *Journal of the American Chemical Society* 116(25): 11594-11595.

Wen, C.M., C.S. Tseng, C.Y. Cheng & Y.K. Li (2002). Purification, Characterization and Cloning of a Chitinase from Bacillus Sp. Nctu2. *Biotechnol Appl Biochem* 35(Pt 3): 213-9.

Withers, S.G., K. Rupitz, D. Trimbur & R.A. Warren (1992). Mechanistic Consequences of Mutation of the Active Site Nucleophile Glu 358 in Agrobacterium Beta-Glucosidase. *Biochemistry* 31(41): 9979-85.

Zhao, K.J. & M.L. Chye (1999). Methyl Jasmonate Induces Expression of a Novel Brassica Juncea Chitinase with Two Chitin-Binding Domains. *Plant Mol Biol* 40(6): 1009-18.

Zou, J., G.J. Kleywegt, J. Stahlberg, H. Driguez, W. Nerinckx, M. Claeyssens, A. Koivula, T.T. Teeri & T.A. Jones (1999). Crystallographic Evidence for Substrate Ring Distortion and Protein Conformational Changes During Catalysis in Cellobiohydrolase Ce16a from Trichoderma Reesei. *Structure Fold Des* 7(9): 1035-45.

6 Acknowledgements

This Thesis is by far the most significant scientific accomplishment in my life and it would be impossible without people who supported me and believed in me.

First of all I would like to express my heartfelt gratitude towards my supervisor Sherry Mowbray for accepting me as a student. Your efforts, guidance and enthusiasm expand and strengthen my scientific horizons. Thank you very much for the friendliness, kindness, and above all, the faith in me, which made my stay in Uppsala comfortable.

I sincerely thank my second supervisor Jerry Ståhlberg for his help with his deep knowledge on carbohydrate degradation. I admire your kind and friendly support and encouragements.

I extend my appreciation to Alwyn Jones for all help and scientific input during these years.

I wish to express my gratefulness to

Terese Bergfors for all help and most importantly directing me to Sherry. Without your protein crystallization course, I would not end up here.

Mats Sandgren, Inés Muñoz, Sabah Mahdi, Jinyu Zou, Mark Harris, Gunner Berglund, Martin Svenda, Lars Liljas, Torsten Unge, Gerard Kleywegt, Emma Jakobsson, Jenny Berglund, Eva-Lena Andersson, Kaspars Tars, Anke Terwisscha van Scheltinga, Saeid Karkehabadi, Jimmy Lindberg, Alasdair MacKenzie-Hose, Rosmarie Friemann, Malin Uppsten, Margareta Ingelman, Patrik Johansson and Seved Löwgren for all the help.

my past and present colleagues for their scientific and non-scientific discussions, interesting company during conferences and synchrotron trips.

Elleonor Uhlin, Ulla Uhlin and Stefan Knight for all the administrative help and Christer Andersson, Erling Wikman, Remco Wouts and David van der Spoel for keeping the computers running.

my collaborators Peter Steenbakkers, Harry Harhangi and Huub Op den Camp of Radbound University, Nijmegen, The Netherlands, Jessica Tang and Mee-Len Chye of University of Hong Kong, Daizo Koga of Yamaguchi University, Japan and Nina Grantcharova and Klas Flärdh of Uppsala University.

Kiyohiko Igarashi and Masahiro Samejima of University of Tokyo, Japan, Charles Specht of Boston University, USA, Karl Kramer, ARS-USDA,

Manhattan, Kansas, USA, Marie-France Versali, University of Liege, Belgium, for all the scientific discussions

W.F. Stevens of Asian Institute of Technology, Thailand for introducing me to the world of chitin-chitosan
Didier Montet of CIRAD, Montpelier, France for his help and friendship.

my high school botany teacher E.S.U. De Silva (St. Thomas' College, Gurutalawa, Sri Lanka) for his initial stimulus and contributions in my scientific life.

It is my pleasure to thank my friends Ananda, Adikari, Nanayakkara and Palitha for their encouragements in everyway. I appreciate your friendship, which makes me comfortable always.

I wish to express my admiration to
Chinthaka Sananth and Thilanka for their affection.
Jayani, Amarasinghe, Britt, Kjell, Chandra and Claus for their help and friendship.
Jayantha Welihinda for his friendship, scientific and non-scientific discussions.
Wimal Pathmasiri and Jagath Weerasena for all the help.

I am thankful to Barbro Lowisin and Ulf Lindahl of Uppsala University for all the supports and GLIBS meetings.

I gratefully acknowledge the financial support of the Swedish Foundation for Strategic Research via the Glycoconjugates in Biological Systems (GLIBS) network and Swedish Research Council (VR).

I am indebted to my mother, father, Udani, Samantha, Wasantha, Manju and their families and Surani for their encouragement, help and love. The great affection of my mother-in-law is thankfully recalled.

Kumari .... I extremely grateful to you for your love, dedication, help, patience, encouragement and taking care of everything at home throughout this long period.

My loving Chetan and Sayu, you were my only relief when I was exhausted after my long days.

Wimal

Appendix A

Table of data collection and refinement statistics for *Brassica juncea* chitinase structures discussed in section 3.4.2

Data collection	**BjCHI3**	**BjCHI3 E234A**	**BjCHI3 E234A soaked cellobiose, Hg, glycol chitin**	**BjCHI2 native**	**BjCHI2 Hg soak**	**BjCHI2 Hg & glycol chitin soak**
Space group	P1	P21	P21	P21	P212121	P21
Environment	ESRF ID14:2	ESRF ID14:2	ESRF ID 14:2	ESRF ID 14:3	ESRF ID 14:2	ESRF ID 14:2
Wavelength	0.933	0.934	0.934	1.009	0.934	0.934
Cell dimensions (Å, °)	a=51.2 b=61.9 c=75.4 α=100.8 β=89.4 γ=91.1	a=62.6 b=51.8 c=75.1 α=90 β=101.6 γ=90	a=61 b=47.2 c=77.8 α=90 β=100.9 γ=90	a=61.9 b=51.7 c=76.3 α=90 β=101.3 γ=90	a=48.4 b=49.2 c=100.8 α=90 β=90 γ= 90	a=61.5 b=51.5 c=76.5 α=90 β=101.4 γ=90
Resolution (Å)	74.54-1.5	73.5 – 2.7	76.7 – 2.6	74.80-2.6	50.41-1.8	75 – 2.58
Unique reflections	87332	12821	13735	14586	22996	14665
Average multiplicity (High resolution shell)	2.1 (1.5)	3.6 (3.7)	7 (5.7)	3.9 (4.0)	6.9 (7.2)	3.6 (3.7)
Completeness (%) (High resolution shell)	63.6 (40.0)	97.7	99.4	98.9	99.8	98.1
R_{merge} (High resolution shell)	15.6 (20.4)	20.3(34.2)	16.6(31.7)	12.2(42.8)	13.3(27.2)	13.7(36.4)
$\langle I/\sigma I \rangle$ (High resolution shell)	8.7 (3.1)	7.1 (3.6)	17 (10.2)	11.9 (3.1)	15 (7.3)	13.3 (4.4)
Refinement						
Number of reflections (completeness %)	88638 (63.6%)	12190 (97.1)	12887 (99.3)	13681 (100)	21722 (99.7)	13914 (97.6)
Resolution range (Å)	74.54 – 1.5	73.52-2.7	76.25-2.6	60.75 - 2.6	50.64 -1.8	74.95 - 2.6
R-factor / R-free (%)	21.6 (25.6)	20.49 (28.9)	20.59 (28.1)	21.58 (28.7)	18.46 (24.3)	23.65 (31.0)
Number of protein atoms	1900	1895	1895	1901	1935	1901
Average B (Å2) in each molecule	A 12.7 B 12.3 C 11.1 D 10.9	A 21.2 B 20.5	A 11.4 B 11.8	A 31.3 B 41.4	17.3	A 27.9 B 36.7
Number of water molecules (Average B, $Å^2$)	653 (23.6)	59 (13.6)	74 (5.6)	57 (20.3)	253 9 (31.5)	56 (19.4)
Number and name of ion (Average B, $Å^2$)[e]					6 Cl^- ions (20.2)	
r.m.s bond length (Å)	0.009	0.008	0.008	0.008	0.012	0.008
r.m.s. bond angle (°)	1.15	1.01	1.02	1.07	1.32	1.072
No. Ramachandran plot outliers (%)	(a) 10(4.6) (b) 7 (3.2) (c) 5 (2.3) (d) 5 (2.3)	(a)8 (3.7) (b) 5 (2.3)	(a)13(6.0) (b)13(6.0)	16 (3.7)	4 (1.8)	20 (4.6)