The Biochemistry and Structural Biology of
Plant Cell Wall Deconstruction

Harry J. Gilbert

The Complex Carbohydrate research Center
The University of Georgia
315 Riverbend Road
Georgia 30602
USA
E.mail hgilbert@ccrc.uga.edu
Tel. No. +1-706-5830655
Abstract

The plant cell wall is the most abundant source of organic carbon in the biosphere. It is estimated that, annually, ~$10^{11}$ tons of plant structural polysaccharides are recycled by microbial enzymes. The saccharification of the plant cell wall is not only an important biological process but is of growing industrial significance, as lignocellulosic polymers represent an attractive renewable substrate for the biofuels and bioprocessing sectors. In the last 20 years there has been an explosion of structural, biochemical and phylogenetic information on plant cell wall degrading enzymes. In sharp contrast, there is limited detailed biochemical data and no structural information on the glycosyltransferases that synthesize the plant cell wall. One of the most significant features of plant cell wall degrading enzymes is their complex molecular architectures in which the catalytic modules are appended to non-catalytic carbohydrate binding modules (CBMs). While CBMs potentiate the activity of their catalytic partners against complex substrates, the mechanism by which this occurs remains a contentious issue. This article will review recent structural data revealing how subtle modifications to both the catalytic modules and CBMs of plant cell wall degrading enzymes can have a radical influence on specificity and activity. This report will also highlight the novel role CBMs can play in both substrate recognition and cellular location.
Introduction

The cell walls of plants are the most abundant source of organic carbon on the planet. This photosynthetically fixed carbon is recycled by microbial enzymes that convert cell wall polysaccharides to mono- and oligosaccharides, a process that is of biological and industrial importance. (Sticklen, 2008; Himmel and Bayer, 2009). Plant cell walls are recalcitrant to biological depolymerization as the extensive interactions between polysaccharides, and between polysaccharides and lignin, restrict access to the battery of microbial glycoside hydrolases, pectate lyases and esterases that break down these composite structures (see (Mohnen, 2008) for review). Since the early 1990s there has been an explosion of structural information on both the catalytic and non-catalytic components of these enzymes. This review will provide an overview/update of the structure-function relationships of the enzymes that catalyze plant cell wall deconstruction.

The plant cell wall

Plant cell walls are comprised predominantly of the polysaccharides cellulose, hemicellulose and pectin, although secondary walls are often rigidified by the impregnation of lignin, a heterogenous aromatic polymer. The structure of the plant cell has been extensively reviewed previously and will be described briefly here (see (Harris and Stone, 2008; Mohnen, 2008; Mohnen et al., 2008) for an overview of plant cell wall structure).

Cellulose is a β-1,4-linked glucose (Glc) molecule that is substantially crystalline. All hemicellulosic polysaccharides contain a β-linked sugar backbone. In xylans, mannans and xyloglucans the backbone sugars are β-1,4-D-Xyl, β-1,4-D-Man and β-1,4-D-Glc, respectively, while in glucomannan the backbone consists of randomly dispersed β-1,4-Glc and β-1,4-Man sugars. The backbones of hemicellulosic polysaccharides are decorated with a variety of sugars and acetyl groups explaining why these polymers are not crystalline. There are three major forms of pectin: homogalacturonan (HG), rhamnogalacturonan I (RG-I) and rhamnogalacturonan II (RG-II) (reviewed in (Mohnen, 2008)). HG consists of a polygalacturonic acid backbone (Mohnen, 2008). RG-I displays a backbone composed of an alternating disaccharide, (α-1,4)-d-GalA→(α-1,2)-l-Rha)ₙ that contains extensive decorations at the O4 of the rhamnose residues (Mohnen, 2008). RG-II is the most structurally complex of the three pectic polysaccharides,
consisting of 13 different sugars and over 20 different linkages (see O'Neill et al., 2004) for an extensive review).

**CAZy**

Enzymes that modify complex carbohydrates, together with their accessory non-catalytic carbohydrate binding modules (CBMs), have been grouped into sequence-based families on the continuously updated CAZy database ((Cantarel et al., 2009); http://www.cazy.org/). Members of the same enzyme family display a common fold, while the catalytic apparatus and mechanism are similarly conserved. Currently 44 of the 115 glycoside hydrolase families (GHs) contain enzymes that contribute to plant cell wall deconstruction. Crystal structures of relevant enzymes in 41 of these 44 GHs have been reported. With respect to polysaccharide lyase families (PLs) and carbohydrate esterase families (CEs) there are 6 out of 21 PLs and 11 out of 16 CEs that contain enzymes that play a role in plant cell wall metabolism. Of the 59 CBM families, around half of these modules bind to components of the plant cell wall, and structural information is available for all but three of these families.

While the structures of CEs, PLs and CBMs are dominated by the α/β hydrolase (Correia et al., 2008), parallel β-helix (Pickersgill et al., 1994) and jelly roll (or β-sandwich) (Czjzek et al., 2001) folds, respectively, there is a large number of different folds within the GHs, which are discussed below. Indeed, the same criteria used to include enzymes in the same GH have now been used to cluster a proportion of the GHs into 14 different clans (Cantarel et al., 2009). The structural biology of plant cell wall degrading systems provides elegant examples of both convergent and divergent evolution (Fig. 1).

**Mechanism of plant cell wall deconstruction**

The vast majority of glycoside hydrolases cleave glycosidic bonds by either a single or double displacement mechanism, which leads to inversion and retention of anomeric configuration, respectively (see (Rye and Withers, 2000) for review). Polysaccharide lyases cleave their scissile bond through a β-elimination mechanism (Herron et al., 2000). While carbohydrate esterases generally hydrolyze ester linkages through a double displacement mechanism in which serine [(Schubot et al., 2001) or aspartate in CE8; (Fries et al., 2007)] functions as the catalytic
nucleophile, exceptions to this mode of action are apparent in CE4 where catalysis is metal dependent (Taylor et al., 2006).

**Catalytic modules of glycoside hydrolases**

Currently, the crystal structure of the catalytic modules of representatives of nearly all the relevant GHs, PLs, CEs and CBM families, which contribute to plant cell wall deconstruction, have been reported (Cantarel et al., 2009). Some structural folds have given rise to a myriad of enzymes that display significant differences in specificity, exemplified by the GHs located in clan GH-A. Members of this clan display a (β/α)$_8$-fold in which the catalytic residues are presented at the C-terminus of β strands 4 and 7 (Henrissat et al., 1995; Jenkins et al., 1995). While the enzymes all hydrolyse an equatorial glycosidic bond, their mode of action (exo and endo), specificity for the sugar at the catalytic -1 subsite and more distal regions of the substrate binding region (Xyl, Man, Glc, Ara, Gal), and the linkage cleaved (e.g. β-1,4 β-1,3) varies between enzymes (Fig. 1). The same enzyme activity can often be found in multiple GHs, located in distinct clans, as a consequence of convergent evolution (Fig. 1). For example, cellulases are located in 11GHs with seven of these families distributed across four different clans, while four of these GHs are, currently, not linked to a clan. There have been several reviews on the 3D structure of the catalytic modules of glycoside hydrolases, including plant cell wall degrading enzymes (Davies and Henrissat, 1995; Henrissat and Davies, 1997; Henrissat and Davies, 2000; Davies et al., 2005; Gilbert et al., 2008). Therefore, this review will provide a brief overview of the structures of these enzymes and a more detailed description of more recent structural information.

**Cellulases:** Cellulose utilization is believed to be mediated by endo-β-1,4-glucanases, cellulbiohydrolases (also called exo-β-1,4-glucanases) and β-glucosidases. Classically cellulose hydrolysis, of which the *Hypocrea jecorina* (formerly *Trichoderma reesei*) system is the archetype, is viewed as a synergistic process between endo-acting cellulases which create new ends from which the exo-acting cellulbiohydrolases can release cellobiose from either the reducing (GH7 and GH48) or non-reducing (GH6) end of the cellulose chains (reviewed in (Kleywegt et al., 1997; Teeri, 1997)). This model, however, is inconsistent with several features of cellulose degradative systems. Thus biochemical and structural data indicate that GH6
cellobiohydrolases are not, exclusively, exo acting (Amano et al., 1996; Armand et al., 1997; Varrot et al., 1999). Furthermore, some highly active cellulase systems lack a classic pair of cellobiohydrolases that act from the reducing and non-reducing ends of cellulose chains, respectively (Xie et al., 2007; Weiner et al., 2008). Indeed, one of the most distinctive features of the cellulose degrading bacterium, *Cytophaga hutchinsonii*, is the absence of GH6, GH48 or GH7 cellobiohydrolases (Xie et al., 2007), although it is possible the bacterium contains novel cellobiohydrolases. An intriguing report by Tolonen and colleagues showed that a single endo-processive GH9 cellulase was essential for cellulose degradation in *Clostridium phytofermentans* (Tolonen et al., 2009). Given the redundancy in cellulase systems, demonstration that a single enzyme is essential for a functional degradative system is rare, and questions the classical synergy model. While there now does not appear to be a single unifying model for cellulose hydrolysis, recent studies, deploying atomic force microscopy to visualize the movement of cellulase molecules on its crystalline substrate, will likely provide novel insights into the mechanism by which these enzymes function (Igarashi et al., 2009).

Plant cellulases are restricted to a very small number of families exemplified by *Arabidopsis thaliana*, whose genome encodes 25 cellulases (endoglucanases) all in GH9. Phylogenetic analysis of the *Arabidopsis* GH9s points to three distinct subfamilies; α, β and γ. Biochemical studies on γ-endoglucanases show that they are ~100-fold less active than the corresponding microbial enzymes, reflecting the loss of a critical aromatic residue at the -2 subsite (Master et al., 2004). This reinforces a remodelling, rather than a degradative role for these membrane-associated cellulases. A cohort of the α-endoglucanases contains a C-terminal module that binds to cellulose (Urbanowicz et al., 2007), which has functional implications discussed below.

**Xyloglucan:** The β1,4-glucan backbone of xyloglucan is hydrolyzed by specific endoglucanases, i.e. endo-xyloglucanase or xyloglucan endo-hydrolases, from GH5, GH7, GH12, GH16, GH44 and GH74. GH12 enzymes can tolerate the side chains in xyloglucan. Indeed, GH5 and GH74 endoxyloglucanases can make productive interactions with the α1,6-Xyl decorations and, in the case of the GH5 enzymes, Gal pendants of the Xyl residues (Martinez-Fleites et al., 2006; Gloster et al., 2007). Maybe the most interesting aspect of xyloglucan modification is found in GH16 where enzymes may display endoxyloglucanase activity or, in the case of XETs, remodel...
the structure of the polysaccharide through transglycosylation reactions (Baumann et al., 2007).

This article will not discuss these GH16 enzymes, which are covered in detail in the review by Ekläf and Brumer in this issue of *Plant Physiology*.

**β-Mannanases:** β-mannanases display a (β/α)8 barrel-fold and are located within GH5 and GH26, while β-mannosidases are GH2 enzymes; all three GHs are within clan GH-A. The crystal structures of β-mannanases generally reveal an open active-site cleft with at least four subsites. An unusual feature of β-mannanases is that substrate specificity is not conferred by the recognition of Man in its relaxed chair conformation (\(4C_1\)) at the critical -1 subsite (glycosidic bond cleavage occurs between the sugars bound at the -1 and +1 subsites; (Davies et al., 1997)), but through the \(B_{2,5}\) topology displayed by the oxocarbonium transition state (Ducros et al., 2002; Tailford et al., 2007; Cartmell et al., 2008).

In addition to mannan, β-mannanases hydrolyze glucomannan, a heterogenous β-1,4-linked polymer of Glc and Man. β-mannanases are defined by their capacity to hydrolyze mannosidic bonds, which requires that Man is positioned in the -1 subsite. Recognition of Man and Glc at subsites distal to -1 is highly variable, although some general trends are emerging, which point to a divergence in specificity between GH5 and GH26 mannanases. GH5 mannanases are able to accommodate Glc at the -2 and +1 subsites (Tailford et al., 2009), and are thus able to hydrolyze mannosidic linkages flanked by Man or Glc. Indeed one of these enzymes, *Baking*5A, does not recognize O2 as a specificity determinant at any subsite distal to -1. Thus, while *Baking*5A hydrolyzes only mannosidic bonds, the topographical features of the substrate binding cleft of this enzyme are optimized to utilize glucomannan as its preferred substrate (Tailford et al., 2009). The relaxed specificity for Glc or Man, apart from the critical -1 subsite, is a feature shared with the other GH5 mannanases, where structural information is available.

In contrast, the GH26 mannanases, characterized to date, generally display tight specificity for Man at both the -2 subsite and -1 subsites. Indeed, a cohort of GH26 mannanases contain an arginine at the -2 substrate that makes extensive interactions with the substrate, and appears to confer unusually high activity against small mannooligosaccharides (Ducros et al., 2002; Cartmell et al., 2008). Screening genomic databases for other GH26 enzymes that retain this
arginine may facilitate the identification of novel mannooligosaccharidases. Currently, the two
Cellvibrio enzymes that contain a high-affinity -2 subsite do not possess additional negative
binding subsites, which may explain why the high activity displayed against mannotriose and
mannotetraose is not translated to the hydrolysis of polysaccharides (Hogg et al., 2001; Cartmell
et al., 2008).

Xylan degradation: The xylan backbone is hydrolysed primarily by GH10 and GH11 xylanases,
while the Araf side chains are removed by arabinofuranosidases from GH43, GH51, GH54 and
GH62 (see (Gilbert et al., 2008) for review of xylan degradation). The uronic side chains are
released from the non-reducing end of xyloooligosaccharides by GH67 α-glucuronidases (Nurizzo
et al., 2002), although recent data showed that GH115 α-glucuronidases remove the uronic acid
decorations from the internal regions of xylan (Ryabova et al., 2009). Each of these families
contains at least one structural representative, with the exception of GH62 and GH115
(http://www.cazy.org). GH43 enzymes may display the highest level of substrate diversity
exemplified by the activity of two arabinofuranosidases from this family that remove the O3-side
chain from Xyl residues that are decorated at both O2 and O3 with Araf (van den Broek et al.,
2005; Sorensen et al., 2006). The crystal structure of this enzyme (HJG unpublished data) reveals
an extended substrate binding pocket that interacts with both O2 and O3 linked Araf. By contrast
an arabinoxylan-specific GH43 arabinofuranosidase, which removes O2 or O3-linked Araf side
chains from singularly substituted Xyl residues, contains a small substrate binding pocket
embedded in a shallow cleft that is optimized to bind the 3-fold helical structure of the xylan
backbone (Vandermarliere et al., 2009). Recent protein crystallographic studies have shown that
xylan side chains can be accommodated, and can actually be exploited as specificity determinants
(Pell et al., 2004; Vardakou et al., 2005), while a GH5 xylanase displays an absolute requirement
for 4-O-methyl-D-glucuronic acid (MeGlcA) appended to the Xyl positioned at the -2 subsite
(Vrsanska et al., 2007). There are two structures of this enzyme (Larson et al., 2003; St John et
al., 2009), however, the mechanism by which the enzyme recognizes the uronic acid side chain
remains unclear.

Pectin degradation: The structures of pectinases (polygalacturonases), pectate lyases and pectin
methylesterases have been extensively described and reviewed previously (Jenkins and
Pickersgill, 2001). In general these enzymes display a right-handed parallel \( \beta \)-helix topology. Exceptions include PL10 pectate lyases, which adopt an \((\alpha/\alpha)_6\) toroid conformation (Charnock et al., 2002), and PL2 lyases that display a \((\alpha/\alpha)_7\) barrel and utilize manganese rather than calcium in the active site (Abbott and Boraston, 2007). The catalytic apparatus in PL10, and those displaying a \( \beta \)-helix fold, is conserved, providing an example of convergent evolution (Charnock et al., 2002). An arginine is the most likely candidate catalytic base in these PLs. The basic residue abstracts the C5 proton which, in several PL families (PL2, PL9 and PL10), results in the formation of an enolate-enolate intermediate in which the two negatively charged oxygens are stabilized by calcium and hydrogen bonds. The collapse of the intermediate results in the cleavage of the scissile bond, although the mechanism by which the leaving group (glycosidic oxygen) is protonated remains unclear. An interesting variation of this catalytic mechanism has been proposed for PL1 lyases. It was suggested that the PL1 lyase generates an enol-enolate through donation of a proton by a nearby lysine to one of the oxygen atoms of the carboxylate. The authors suggest that through this intermediate, PL1 lyases are more active than PL10 and PL9 enzymes that can only generate the enolate-enolate intermediate (Seyedarabi et al., 2010).

Recent advances have also been made in understanding the processive mechanism displayed by pectin methyl esterases, which yield blocks of non-methylated galacturonic acid (GalA). Structural and biochemical data show that the enzyme demethylates the sugar at the +1 subsite and uses the negative charge of the carboxylate as a specificity determinant at the -1 subsite, and to some extent at -2, while +3 makes hydrophobic contact with the methyl group of the esterified uronic acid (Fries et al., 2007). Thus, after removing the methyl group the GalA generated then slides along the substrate binding cleft to occupy the -1 site and thus a new methylated GalA is presented in the crucial +1 subsite. This progressive sliding of pectin along the substrate binding cleft is encouraged further by the specificity displayed by the +3 and -2 subsites.

**Structural changes that modulate the mode of enzyme action**

The structural basis for the GH6 and GH7 cellobiohydrolases and endoglucanases is well established and has been extensively reviewed (Kleywegt et al., 1997; Teeri, 1997; Varrot et al., 1999). Recent structural data have also provided insight into how subtle structural changes can convert endo-acting glycoside hydrolases and polysaccharide lyases into exo-acting enzymes.
Thus small loop extensions, surrounding the distal subsite that accommodates the non-reducing end of the substrate create steric constraints that prevent extension of the substrate beyond this subsite. Variants of these enzymes, in which the loops extensions have been removed, display an endo mode of action (Proctor et al., 2005; Cartmell et al., 2008; Ochiai et al., 2009). Fig. 2.

**Plant cell wall glycosyltransferases:**
The crystal structure of several glycosyltransferases in numerous GT families have been reported in the last decade. The data have revealed only two major folds for these enzymes, while also providing insights into the likely catalytic mechanisms displayed by inverting and retaining glycosyltransferases (reviewed by (Lairson et al., 2008)). These studies, however, have focused, almost exclusively, on enzymes that are not membrane-associated and, currently, there is no high resolution crystal structural information on glycosyltransferases that contribute to plant cell wall synthesis. Cellulose synthase, however, can be visualized by freeze fracture techniques, in conjunction with immunological methods. The data revealed six globular complexes ~25 nm in diameter. Each of the six subunits (each subunit contains multiple cellulose synthase molecules) of these rosettes synthesize multiple $\beta$-1,4-glucan chains, which co-crystallize to form microfibrils (see (Somerville, 2006) for review). Although issues remain concerning the nature of the primer and the direction of chain growth elongation (although elongation from the nonreducing end is the preferred model), the lack of detailed structural information on these enzymes precludes further discussion of this enzyme system here. It is evident, however, that using genetic approaches, *Arabidopsis* plant cell wall glycosyltransferases have been identified and, in some instances, predicted activities have been verified by detecting appropriate transfer reactions in heterologous hosts (See (Liepman et al., 2010) for a review). It is evident that in the next few years there will be rapid advances in the identification of glycosyltransferase genes that encode plant cell wall synthesizing enzymes. It is highly likely that the resultant data will underpin the much needed detailed structural and biochemical information of these plant cell wall glycosyltransferases.

**CBMs**
Microbial plant cell wall hydrolases display complex molecular architectures in which the catalytic module is appended, by flexible linker sequences, to one or more CBMs (reviewed in
(Boraston et al., 2004)). In some of the 59 CBM families, exemplified by CBM1, CBM10 and
CBM20, ligand specificity is invariant (Linder and Teeri, 1997; Southall et al., 1999;
Raghothama et al., 2000), while in some families, such as CBM6 (Czjzek et al., 2001; Pires et
al., 2004), CBM4 (Boraston et al., 2002) and CBM35 (Tunnicliffe et al., 2005; Montanier et al.,
2009a), carbohydrate recognition is highly variable. In addition to defining a phylogenetic
relationship between CBMs by clustering these modules into sequence-based families, they have
also been classified into three categories (type A, B and C) based on the topology of their ligand
binding sites and their mode of ligand recognition (reviewed by (Boraston et al., 2004)). Fig 3.

CBMs have now been described that bind to the major polysaccharides found in plant cell wall
(see (Boraston et al., 2004) for review), while modules that recognize the side chains of these
polymers, and the products released through their deconstruction, have also been identified
(Notenboom et al., 2001; Miyanaga et al., 2004; Montanier et al., 2009a). In general, the ligand
specificity of CBMs reflects the substrate cleaved by the cognate enzyme (discussed further
below). Many of these enzymes, however, also contain a CBM that binds to crystalline cellulose
(Kellett et al., 1990; McKie et al., 2001; Hogg et al., 2003). It has been suggested that once
bound these type A modules are able to slide across the surface of cellulose (Jervis et al., 1997)
enabling the substrate-specific type B and type C CBMs to lock onto its ligand, and thus direct
the enzyme to its target glycosidic bonds (Kellett et al., 1990).

In general, the affinity of CBMs for their target plant-derived ligands is low ($K_{D}$ ~100 μM)
(Boraston et al., 2004). Some enzymes, however, contain multiple copies of CBMs that display
the same specificity, and in these proteins avidity effects between these modules has led to
increased affinity for polysaccharides (Bolam et al., 2001; Freelove et al., 2001; Boraston et al.,
2002). It is interesting that nature has deployed CBM duplication as a mechanism for increased
affinity rather than increasing the interactions between ligand and a single CBM module. It is
possible that as CBMs generally bind to ligands that are in intimate contact with other
components of the plant cell wall, steric constraints prevent extensive interactions between the
protein and target carbohydrate.
**Plant CBMs:** CBMs are less prevalent among plant glycoside hydrolases that cleave structural polysaccharides, however, several CBM49 and CBM22 modules are present in plant cellulases and xylanases, respectively. The CBM49 modules are located in a subfamily of GH9 plant endoglucanases, and one of these modules was shown to bind tightly to crystalline cellulose (Urbanowicz et al., 2007). It is possible that the GH9 CBM49-containing endoglucanases play a role in modulating the structure of crystalline cellulose.

**How do CBMs potentiate catalysis?**

The mechanism by which CBMs potentiate catalysis remains unclear. It has been hypothesised that cellulose-specific CBMs may play a key role in disrupting the ordered hydrogen bonding network in crystalline cellulose, making the surface chains accessible to the appended cellulase (Knowles et al., 1987; Teeri, 1997). There is biochemical, biophysical and microscopic data (Din et al., 1994; Wang et al., 2008) indicating that CBMs mediate changes to the surface structure of cellulose. Furthermore, the addition of CBMs *in trans* to the cognate catalytic module has led to a modest potentiation (0.2 to 1.5-fold) in catalytic activity against insoluble substrates (Din et al., 1994; Moser et al., 2008). Cellulases, typically endoglucanases, however, are often 3-orders of magnitude more active against soluble forms of cellulose than the crystalline polysaccharide (Durrant et al., 1991; Irwin et al., 1993). Thus, CBMs acting *in trans* have only a minor influence on the access problem, although this might reflect the dissociation of the targeting and (possible) disrupting function of these modules. Of potential significance is the location of crystalline cellulose-specific CBMs in many enzymes that display no cellulase activity (Kellett et al., 1990; McKie et al., 2001; Hogg et al., 2003; Vincent et al., 2010), which argues against these modules having a specialized function in cellulose degradation. A more likely explanation for the capacity of CBMs to increase the activity of glycoside hydrolases against insoluble substrates is that they reduce the “accessibility problem” by bringing the appended catalytic modules into intimate and prolonged association with their target substrate thereby enhancing catalytic efficiency.

In several organisms, however, there are populations of CBMs, which are not components of enzymes, and these modules may destabilize the crystalline structure of some polysaccharides. Thus, CBM33, which is highly expressed in chitin degrading bacteria such as *Serratia marcescens*, potentiates the chitinases from this organism, particularly during the latter stages of...
the degradative process when the glycoside hydrolases are attacking highly crystalline forms of
the polysaccharide (Vaaje-Kolstad et al., 2005). More modest potentiation of cellulases by “non-
catalytic” bacterial CBMs has also been reported (Moser et al., 2008), while it has been
suggested that several non-catalytic fungal proteins may play a role in plant cell wall disruption.
It is believed that the primary function of GH61s (now established as fungal non-catalytic
carbohydrate binding proteins) is to disrupt plant cell wall structure and thus increase the access
of degradative enzymes to their substrates (Rosgaard et al., 2006; Karkehabadi et al., 2008;
Harris et al.). Indeed, fungi often contain multiple copies of this protein and they are co-
expressed with a range of cellulases (Vanden Wymelenberg et al., 2009). Another potential
fungal CBM33 analogue is Swollenin from *Trichoderma reesei*, which appears to have
disruptive effects on cellulose, although recent studies suggest that the protein may display
endoglycanase activities (Yao et al., 2008). In addition to microorganisms, plants (and plant cell
wall degrading nematodes) also produce proteins, referred to as expansins, which mediate a
relaxation in the structure of the cell wall (see (Cosgrove, 2000) for review). Expansins
mechanically weaken plant cell walls (McQueen-Mason and Cosgrove, 1994), and their use in
improving cellulase efficiency has been reported (Han and Chen, 2007). Currently, GH61s and
expansins appear to disrupt the cellulose-hemicellulose interface, while the functional
importance of Swollenin remains opaque.

The structure of a CBM33 reveals a binding surface that contains several conserved polar
residues that are pivotal to the synergistic effects of this protein with chitinases (Vaaje-Kolstad et
al., 2005). Significantly, mutations that prevented the CBM33 from potentiating chitinase
activity had little effect on the affinity of the protein for chitin. This led to the proposal that the
specific polar interactions between chitin and the protein disrupt the hydrogen-binding network
between individual polysaccharide chains (Vaaje-Kolstad et al., 2005). This could also explain
the specificities displayed by CBM33 modules. The crystal structures of GH61s from *Hypocrea
jecorina* (Karkehabadi et al., 2008) and *Thielavia terrestris* (Harris et al., 2010) reveal a similar
surface to CBM33, again pointing to a disruptive function for these proteins.

While the effects of non-catalytic proteins on cellulose hydrolysis have, to date, been
disappointing, continued efforts at identifying the functional significance of these molecules is
merited. For example, it is possible that specific combinations of these proteins are required to disrupt the structure of cellulose, while, currently, the influence of these accessory proteins have been explored only in isolation.

The Structural basis for CBM specificity:
The three dimensional structures of representatives of 21 of the 24 CBM families that target the plant cell wall have been determined (Reviewed by (Boraston et al., 2004)). The vast majority of these modules display a jelly roll fold comprising two anti-parallel \( \beta \)-sheets that form the two surfaces of these proteins. Ligand binding occurs on the concave surface presented by one of the \( \beta \)-sheets (Boraston et al., 2002), or in the loops that connect the two \( \beta \)-sheets (defined hereafter as site 1) (Czjzek et al., 2001; Montanier et al., 2009a).

Crystalline Cellulose recognition: CBMs that bind crystalline cellulose contain three aromatic amino acids that adopt a planar topology with respect to each other (Kraulis et al., 1989; Xu et al., 1995; Tormo et al., 1996; Raghothama et al., 2000). These residues make extensive hydrophobic contacts with fully exposed sugar rings presented at the 110 face of cellulose crystals (Lehtio et al., 2003). Ligand recognition by these CBMs is driven primarily through an increase in entropy resulting in the desolvation of the interacting macromolecules (Creagh et al., 1996). By contrast, enthalpy drives the binding of CBMs to discreet polysaccharide chains, where both polar and apolar interactions occur, while entropy has a negative impact on overall affinity (Charnock et al., 2000; Bolam et al., 2001; Boraston et al., 2002; Boraston et al., 2002; Charnock et al., 2002). The negative entropy may reflect conformational restriction of the ligand bound to the protein, which is not entirely offset by the release of tightly bound water molecules (see (Boraston et al., 2004) for review). It should be recognized, however, that the energetic freedom of the solvating molecules of the protein is a controversial issue and thus the molecular basis for the thermodynamic forces that drive ligand recognition in type B CBMs remains unclear.

Xylan versus cellulose recognition: Subtle changes in structure can lead to significant changes in ligand specificity, exemplified by xylan and cellulose specificity within CBM2, which is defined by the conformation adopted by the surface aromatic residues, which are perpendicular in xylan
and planar in cellulose binding modules (Simpson et al., 1999; Simpson et al., 2000), consistent with the conformation adopted by the two polysaccharides. The perpendicular arrangement of surface tryptophans in xylan-binding CBM2s is mediated by an arginine, while in cellulose binding CBM2s the basic residue is replaced by a glycine enabling the tryptophan to collapse onto the surface of the protein and adopt a planar orientation with respect to the other aromatic residues (Simpson et al., 2000). While the perpendicular arrangement of aromatic residues in xylan-binding CBMs is a common feature, modules that recognize the hemicellulosic polymer can adopt different ligand binding strategies. Thus, in CBM4, CBM6 and CBM22, xylan recognition is dominated by a single Xyl residue that is sandwiched between a pair of planar aromatic residues within a deep ligand binding cleft (Czjzek et al., 2001; Charnock et al., 2002; Simpson et al., 2002). While this binding mode confers higher affinity for isolated xylan chains, CBMs that recognize xylan through the asymmetric distribution of aromatic residues display more versatile ligand recognition; they are able to bind to the hemicellulose within in a variety of terrestrial plant cell walls, a specificity that is not displayed by the modules from CBM4, CBM6 and CBM22 (McCartney et al., 2006).

The topology of the ligand binding cleft influences specificity: While CBMs that bind to internal regions of polysaccharides display an open cleft topology, the shape of the cleft influences specificity. This is exemplified in CBM4 where structurally related modules bind to linear \( \beta_1,4 \) polysaccharides, such as cellulose, or highly curved structures such as \( \beta_1,3 \)-glucan (Boraston et al., 2002). In \( C/\text{CBM4-1} \), both ends of the cleft are open enabling the protein to bind linear glucan chains such as cellulose. However, insertions in two loops confers a U shape topology on the longitudinal axis of the binding cleft \( Tm\text{CBM4-2} \), which is complementary to the curved conformation adopted by its ligand, \( \beta_1,3 \)-glucan. A more extreme example of how topological changes can cause a dramatic change in ligand specificity is evident in site 1 in CBM6 modules. This site may adopt a pocket-like topology, and thus recognize the termini of polysaccharide chains (Pires et al., 2004), or display an open cleft and bind to the internal regions of xylan (Czjzek et al., 2001). From the discussion above it is apparent that CBMs, in common with lectins, display preformed carbohydrate-recognition sites which mirror the solution conformations of their target ligands, thereby minimizing the energetic penalty paid upon binding.
Recognition of heterogeneous polymers: Glucomannan (contains a random distribution of β1,4 linked D-Man and D-Glc residues) presents a significant challenge with respect to CBM specificity. While mannan-specific CBMs recognize the regions of glucomannan containing successive Man residues (Tunnicliffe et al., 2005), two CBM families, CBM29 and CBM16, contain proteins that display specificity for the heterogeneous polymer, in addition to cellulose and mannan. In these modules the aromatic residues in the binding cleft make planar contacts with sugars at n and n+2 and thus avoid steric clashes with the axial O2 in the Man residues. In addition, several polar residues are capable of making hydrogen bonds with the axial or equatorial O2 of Man and Glc, respectively; while at other sugar binding subsites O2 is not a specificity determinant (Charnock et al., 2002; Bae et al., 2008). CBMs that have also been shown to harness both the backbone and side chain of decorated glucans such as xyloglucan (Najmudin et al., 2006). Specificity for this polymer has also been engineered into a xylan-specific module (Gunnarsson et al., 2006), while recent structural information on this protein provides insight into how the observed change in specificity was achieved (Gullfot et al., 2010).

Calcium: CBMs, which display a jelly roll fold, contain a highly conserved structural calcium (see (Boraston et al., 2004) for review). Furthermore, there are increasing examples of CBMs where calcium plays a direct or indirect role in ligand recognition in site 1. Thus, in Aga16B-CBM6-2, calcium orientates a tyrosine such that it can interact with the neoagarose ligand (Henshaw et al., 2006), while in CBM36 (and a second xylan-specific CBM; HJG unpublished data) the metal ion coordinates with the O2 and O3 of Xyl residues within xylan (Jamal-Talabani et al., 2004). A cohort of six CBM35s (three of these modules are identical but are located in different xylan degrading enzymes) were recently shown to bind to uronic acids, where calcium makes critical electrostatic interactions with the C6 carboxylate (Montanier et al., 2009a). Subtle differences in the ligand binding site in this cohort of CBM35s confer differences in ligand specificity. The modules derived from the pectin metabolizing enzymes bind only to Δ4,5-anhydrogalacturonic acid, while the other CBM35s recognize both the pectin degradation product and glucuronic acid (Montanier et al., 2009a).

Dual catalytic and non-catalytic binding functions for an esterase
Within the context of plant cell wall degradation, the catalytic and CBM function is conferred by discrete regions of the multi-modular enzymes that catalyze this process. Recently, however, a CE2 esterase, \textit{CtCE2}, which is appended to a GH5 endoglucanase, was shown to have a non-catalytic cellulose binding function in addition to displaying esterase activity; other modules in this esterase family (which are not appended to other catalytic modules) do not recognize cellulose. The crystal structure of \textit{CtCE2} shows that cellulose binds to the active site of the esterase through hydrophobic interactions with three aromatic residues, and by hydrogen bonds with components of the catalytic apparatus. The crystal structure of other CE2 esterases showed that these enzymes lack all three aromatic residues explaining why they do not recognize cellulose. The CE2 family encapsulates the requirement for multiple activities by biocatalysts that attack challenging macromolecular substrates such as the plant cell wall, including the grafting of a second, powerful and discrete non-catalytic binding functionality into the active site of an enzyme. This report provides a rare example of “gene sharing (Montanier et al., 2009b),” where the introduction of a second functionality into the active site of an enzyme does not compromise the original activity of the biocatalyst, \textbf{Fig. 4}.

**Functional significance for CBM ligand recognition**

As discussed above, type B CBMs generally bind to substrates of the catalytic modules. Exceptions to this rule include a xylanase-derived CBM9, which binds the reducing end of xylan or cellulose (Notenboom et al., 2001), while CBM35s, located in pectin metabolizing enzymes, bind to the reaction products generated by pectate lyases (Montanier et al., 2009a). It would appear, therefore, that these CBMs are recruiting enzymes to regions of the cell wall where the target substrates for the appended enzymes are undergoing degradation and thus, in a sense, the modules are directing the catalytic apparatus to areas of the wall that are susceptible to degradation. The CBM35 appended to three xylan degrading enzymes binds to both glucuronic acid (GlcA) and the unsaturated product released by pectate lyases, but not to 4-\textit{O}-methyl-D-glucuronic acid (MeGlcA), the more common uronic acid found in xylans. The biological rationale for this dual specificity is unclear. It has been proposed that in model plants, such as \textit{Arabidopsis}, in rapidly dividing cells there are significant quantities of unmethylated GlcA (Pena et al., 2007). This has led to the hypothesis that, by targeting unmethylated GlcA, the CBM is directing enzymes to more open structures that are particularly susceptible to enzyme
degradation. It is possible that this cohort of CBM35s initially direct the xylan degrading apparatus to regions of cell walls that are being actively degraded, for which anhydrogalacturonic is a marker, but, as xylan structures are revealed, the enzyme is shuttled onto the hemicellulosic polysaccharide affording the enzyme access to its target substrate (Montanier et al., 2009a).

Recent studies have also shown that CBMs can display a bacterial anchoring function. Thus, the CBM35 from the Amycolatopsis orientalis exo-β-D-glucosaminidase (Chi-CBM35) tethers the enzyme to the cell wall of the bacterium (Montanier et al., 2009a). Similarly, a family of CBMs, unique to Ruminococcus albus, which bind to a wide spectrum of β-linked plant structural polysaccharides (CBM37), anchor their cognate enzymes to the surface of the bacterium (Ezer et al., 2008). The biological rationale for this CBM function appears to be to keep the enzymes in close proximity to the bacterium. However, Ezer et al. also proposed a model in which the CBM37 acts as a shuttle which transfers the appended enzymes from the bacterial surface to the plant cell wall (Ezer et al., 2008). In any event these recent reports of the cell adhesion role of CBMs, which was previously unconsidered, may prove to factor prominently in the function of these protein modules in the future.

Future perspectives

In the last 15 years there have been significant advances in the 3D structural analysis of plant cell wall degrading enzymes. The data have informed our understanding of the mechanism of both catalysis and substrate recognition, which has led to the identification of numerous “specificity motifs”, some of which are described in this article. It is evident, however, that the explosion of genomic and metagenomic information is resulting in an exponential increase in the identification of CAZy enzymes. This has resulted in a significant imbalance between the number of enzymes in CAZy families and the biochemical/structural analysis of these proteins. Indeed only around 3% of the proteins in CAZy have a characterized biochemical activity, while 3D structural information is only available for 0.3% of these enzymes (Cantarel et al., 2009). It is estimated that we can safely predict the activities of <20% of the proteins within CAZy. The situation is compounded further by the difficulties in determining the biochemical properties of plant cell wall degrading enzymes, where the chemical complexity and requirement for a
hierarchical degradative process create significant functional barriers. Notwithstanding these problems, continued biochemical and structural information is urgently required if we are to fully integrate information obtained from the “Omics” technologies to understand the biology of plant cell wall deconstruction. Indeed, integrating structure, function and phylogenetics, to develop predictive models for ligand/substrate specificity, is an important goal for structural biologists working on plant cell wall modifying enzymes. An example of such an analysis was developed recently by Abbott. Deploying CBM6 as a model system, Abbott was able to identify two regions that appear to be “hot spots” of primary and tertiary structure variation, which confer functional specificity in these modules, a view supported by the recent characterization of a xylose-specific CBM6 (Abbott et al., 2009). A more general phylogenetic analysis of endoglucanases belonging to several GHs was also insightful in providing a predictive platform for glycoside hydrolase activities (Vlasenko et al., 2010). As discussed above the characterization of glycosyltransferases that catalytic the synthesis of plant structural polysaccharides represents the biggest challenge in the cell wall field. It is evident that a significant investment is required to develop our understanding of the structure/function relationships of these enzymes, which is essential if we are to fully understand the mechanism for the biogenesis of the plant cell wall.

REFERENCES


Creagh AL, Ong E, Jervis E, Kilburn DG, Haynes CA (1996) Binding of the cellulose-binding domain of exoglucanase Cex from Cellulomonas fimii to insoluble microcrystalline cellulose is entropically driven. Proc Natl Acad Sci U S A 93: 12229-12234

21


Han YJ, Chen HZ (2007) Synergism between corn stover protein and cellulase. Enzyme Microb Technol 41: 638-645


Downloaded from on October 31, 2017 - Published by www.plantphysiol.org
Copyright © 2010 American Society of Plant Biologists. All rights reserved.


McQueen-Mason S, Cosgrove DJ (1994) Disruption of hydrogen bonding between plant cell wall polymers by proteins that induce wall extension. Proc Natl Acad Sci U S A 91: 6574-6578

exo/endo preference of Cellvibrio japonicus arabinanase 43A. Proc Natl Acad Sci U S A 102: 2697-2702


Simpson PJ, Bolam DN, Cooper A, Ciruela A, Hazlewood GP, Gilbert HJ, Williamson MP (1999) A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIA cellulose-binding domain but different ligand specificity. Structure 7: 853-864


Tolonen AC, Chilaka AC, Church GM (2006) Targeted gene inactivation in Clostridium phytofermentans shows that cellulose degradation requires the family 9 hydrolase Cphy3367. Mol Microbiol 74: 1300-1313


Fig. 1  Structural convergence and divergence in plant cell wall hydrolases. In the examples shown a GH28 polygalacturonase (green; PDB 1BHE) and GH49 dextranase (pink; PDB 1OGM) are the clan GH-N representative enzymes. The clan GH-C enzymes are a GH11 xylanase (green; PDB 1BCX) and a GH12 endoglucanase (light blue; PDB 1OA4). The Clan GH-A enzymes are a GH5 endoglucanase (magenta, PDB 1A3H), GH26 mannanase (blue; PDB 2BVT) and a GH53 endo-β1,4-galactanase (green; PDB 1R8L). The catalytic residues are shown in stick format in a darker form of the respective color of the protein fold. Convergent evolution is evident by the observation that xylanases are found in three glycoside hydrolase families that display very different folds.

Fig. 2  An overlay of an endo (CjMan26A; green) and exo (CjMan26C; cyan) β-mannanase from Cellvibrio japonicas. A small extension of the loop at the distal -2 subsite presents two residues, D130 and L129, shown in stick format in purple. These residues present a steric block that prevents extension of substrate distal to the sugar bound at the -2 subsite. The residues shown in dark blue, in the surface representations of the two enzymes, are D130 in Cj Man26C and the equivalent amino acid (E121) in CjMan26A.

Fig. 3  Examples of type A, type B and type C carbohydrate binding modules (CBMs). CBM2a is derived from the Cellulomonas fimi xylanase Xyn10A (PDB 1XG), CBM15 is a component of the C. japonicas xylanase Xyn10C (PDB 1GNY) and the CBM9 is from a Thermotoga maritima GH10 xylanase (PDB 1I82). The folds are ramped from blue (N-terminus) to red (C-terminus). The three aromatic residues that form a ligand binding apolar surface in the CBM2a module are colored red and are shown in stick format in the respective surface and fold depictions of the protein.
The structural features of the active site of CtCE2 that displays esterase and a CBM function. The catalytic module of CtCE2 contains two discrete domains that display a typical α/β hydrolase fold (blue), evident in serine esterases, and a jelly roll fold (red), respectively. The surface representation of CtCE2 reveals a cleft that accommodates cellopentaose (shown in yellow) and houses the active site of the esterase. The three aromatic residues that play a key role in binding cellulose are shown in red. The active site of three CE2 esterases show cellopentaose (grey) and the three aromatic residues (green) from CtCE2, the two aromatic residues (magenta) in CjCE2B and the single aromatic amino acid (yellow) in CjCE2A. All three enzymes are serine esterases and the catalytic triad (S160, H335, D333) of CE2A is displayed in yellow.
Overlay of Man26A/C

Man26A

Man26C

Figure 2
Figure 4