A Tomato Endo-\(\beta\)-1,4-glucanase, SlCel9C1, Represents a Distinct Subclass with a New Family of Carbohydrate Binding Modules (CBM49)*

Received for publication, August 18, 2006, and in revised form, January 26, 2007 Published, JBC Papers in Press, February 23, 2007 DOI 10.1074/jbc.M607925200

Breeanna R. Urbanowicz\textsuperscript{1,1}, Carmen Catalá\textsuperscript{1}, Diana Irwin\textsuperscript{5}, David B. Wilson\textsuperscript{1}, Daniel R. Ripoll\textsuperscript{1}, and Jocelyn K. C. Rose\textsuperscript{2,2}

From the Departments of \textsuperscript{1}Plant Biology and \textsuperscript{2}Molecular Biology and Genetics and the \textsuperscript{2}Computational Biology Service Unit at the Cornell Theory Center, Cornell University, Ithaca, New York 14853

A critical structural feature of many microbial endo-\(\beta\)-1,4-glucanases (EGases, or cellulases) is a carbohydrate binding module (CBM), which is required for effective crystalline cellulose degradation. However, CBMs are absent from plant EGases that have been biochemically characterized to date, and accordingly, plant EGases are not generally thought to have the capacity to degrade crystalline cellulose. We report the biochemical characterization of a tomato EGase, \textit{Solanum lycopersicum} Cel8 (SlCel9C1), with a distinct C-terminal noncatalytic module that represents a previously uncharacterized family of CBMs. \textit{In vitro} binding studies demonstrated that this module indeed binds to crystalline cellulose and can similarly bind as part of a recombinant chimeric fusion protein containing an EGase catalytic domain from the bacterium \textit{Thermobifida fusca}. Site-directed mutagenesis studies show that tryptophans 559 and 573 play a role in crystalline cellulose binding. The SlCel9C1 CBM, which represents a new CBM family (CBM49), is a defining feature of a new structural subclass (Class C) of plant EGases, with members present throughout the plant kingdom. In addition, the SlCel9C1 catalytic domain was shown to hydrolyze artificial cellulolytic polymers, cellulose oligosaccharides, and a variety of plant cell wall polysaccharides.

The hydrolysis of cellulose, the most abundant biopolymer on earth, occupies a central position in the global carbon cycle, and a broad range of organisms secrete sets of cellulolytic enzymes to degrade this complex insoluble substrate. The best studied of these are endo-\(\beta\)-1,4-glucanases (also termed EGases\textsuperscript{3}, or cellulases; EC 3.2.1.4), which have been identified and characterized in bacteria, fungi, plants, and animals (1–3). Particular attention has been paid to microbial EGases due to their industrial importance in textile modification and their potential use in the processing of lignocellulosic biomass (4), resulting in detailed insights into their expression, regulation, and enzymatic properties (4–6). Moreover, exhaustive structure-function studies have identified key structural features that contribute to cellulose binding and hydrolysis.

As with many glycosyl hydrolases, microbial EGases typically have a modular structure, involving at least one catalytic domain (CD) joined by a flexible linker region to a single or multiple carbohydrate-binding modules (CBMs) (7). CBMs are structurally diverse noncatalytic domains that typically target proteins to polysaccharide substrates, and they collectively exhibit a range of binding specificities (reviewed in Ref. 8). CBMs attach the enzyme to the substrate surface, potentiating the catalytic activity by increasing the local enzyme concentration and possibly disrupting the surface structure for more efficient catalysis (9). It has also been shown that CBMs can target the enzyme to specific substrates and even substrate microdomains (10, 11). The binding of EGases to cellulose is considered to be a limiting step in cellulose hydrolysis, and CBMs are thus critical components of these modular cellulolytic proteins (12).

In contrast to the detailed biochemical analyses of these microbial enzymes, remarkably little is known about the \textit{in vivo} substrates and mechanism of action of plant EGases. Most activities have been reported using artificial soluble cellulose derivatives, such as carboxymethylcellulose (CMC), and the few more detailed studies of substrate specificity have failed to reveal a common pattern (reviewed in Refs. 3 and 13–15), with various isoymes showing preferential activities against different classes of soluble glucans. However, an important and consistent conclusion is that plant EGases cannot degrade crystalline cellulose, a characteristic that has long been attributed to a distinct structural feature of plant EGases: the absence of a CBM.

\textsuperscript{*} This work was supported in part by United States Department of Agriculture NRI Award 2002-35304-12680 (to J. K. C. R. and C. C.) and by grants from the United States National Science Foundation Plant Genome Program (DBI-0606595) and New York State Office of Science, Technology, and Academic Research (NYSTAR). This research was conducted in part with the resources of the Cornell Theory Center, which receives funding from Cornell University, New York State, Federal agencies, foundations, and corporate partners. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

\textsuperscript{1} The on-line version of this article (available at http://www.jbc.org) contains supplemental Tables S1 and S2.

\textsuperscript{2} Supported by United States Department of Agriculture Initiative for Future Agricultural and Food Systems Multidisciplinary Graduate Education Traineeship Fellowship 2001-52014-11484.

\textsuperscript{3} To whom correspondence should be addressed: Dept. of Plant Biology, 228 Plant Science Bldg., Cornell University, Ithaca, NY 14853. Tel.: 607-255-4781; Fax: 607-255-5407; E-mail: jr286@cornell.edu.

\textsuperscript{4} The abbreviations used are: EGase, endo-\(\beta\)-1,4-glucanase; ASC, acid-swollen cellulose; BMCC, bacterial microcrystalline cellulose; CBM, carbohydrate binding module; CD, catalytic domain; CMC, carboxymethyl cellulose; FP, fusion protein; GH9, glycosyl hydrolase family 9; MLG, mixed linkage glucan; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; MES, 4-morpholineethanesulfonic acid; GST, glutathione S-transferase.
Plant EGases belong to glycosyl hydrolase family 9 (GH9) and comprise large multigene families (16, 17) that group into three distinct subfamilies (3). α- and β-EGases all have a predicted N-terminal signal sequence for secretion to the cell wall, whereas γ-EGases have a GH9 catalytic core coupled to a long N-terminal extension, with a membrane-spanning domain that anchors the protein to the plasma membrane or intracellular organelles (14, 18). We previously identified a tomato EGase, originally named TomCel8 (19) and now termed *Solanum lycopersicum* Cel9C1 (SlCel9C1), which represents a new divergent structural subclass within the α-EGases, and orthologs have now been identified in several plant species (3, 19–22). The members of this subclass exhibit a distinctive modular architecture, with a conventional N-terminal signal peptide and GH9 catalytic core but with an additional discrete C-terminal extension connected to the CD by a proline and hydroxyamino acid-rich linker region (Fig. 1A). This C-terminal module has features that are reminiscent of microbial CBMs, suggesting that this domain might confer binding to cellulose, although no biochemical evidence has been presented to support this hypothesis.

Repeated attempts to generate recombinant SlCel9C1 have revealed its susceptibility to hydrolysis, preventing characterization of the full-length protein. However, in this paper, we describe a dual strategy to demonstrate that the C-terminal module of SlCel9C1 binds to crystalline cellulose, the first such example in plants. Our results indicate that SlCel9C1 and orthologs comprise a distinct subclass of plant EGases, characterized by a distinct C-terminal domain that represents a new family of CBMs (designated CBM49). Data are also presented showing that the SlCel9C1 CD can hydrolyze a variety of cellulosic and noncellulosic plant cell wall substrates, and potential roles of this new structural subclass of EGase are discussed.

**EXPERIMENTAL PROCEDURES**

**Expression of the TfCel6A CD–SlCel9C1 CBM Fusion Protein (Cel6/Cel9C1 FP) in *Escherichia coli*—To create a *T. fusca* TfCel6A CD–SlCel9C1 CBM fusion protein construct, the SlCel9C1 CBM46 DNA sequence (amino acids 500–607) was amplified by PCR (Table S1), followed by digestion with PstI. The resulting SlCel9C1 CBM fusion protein construct, the TfCel6A CD–SlCel9C1 CBM46 DNA sequence (amino acids 500–607) was amplified by PCR (Table S1) and digested with PstI and XhoI. The cDNA encoding the TfCel6A CD (amino acids 1–312), described in Ref. 23, that contains TfCel6A in the pET and XhoI. The cDNA encoding the TfCel6A CD (amino acids 1–312), described in Ref. 23, that contains TfCel6A in the pET

**Construction of Glutathione S-Transferase-SlCel9C1 CBM Fusion Protein and Site-directed Mutagenesis—The pGEX expression system was used for site-directed mutagenized proteins were expressed using the pGEX expression system. The region of the SlCel9C1 DNA sequence containing the CBM (amino acids 526–625) was amplified by PCR (Table S1) and ligated into EcoRI/Sall-digested pGEX-5X-1 (GE Healthcare) to generate GST-SlCel9C1 CBM (GST-CBM).

Site-directed mutagenesis of GST-CBM was performed using a QuikChange site-directed mutagenesis kit (Stratagene). The associated PCR primers are listed in Table S2. The presence of the individual mutations was verified by DNA sequencing (Cornell Biotechnology Resource Center, Ithaca, NY), and positive clones were further designated as GST-CBM W522A, GST-CBM Y529A, GST-CBM W559A, and GST-CBM W573A, with number designations representing amino acids in the mature SelCel9C1 protein.

**Protein expression of the GST-CBM and its mutants in BL21-CodonPlus (DE3)-RIPL cells (Stratagene) was induced with 0.2 mM isopropyl-1-thio-β-D-galactopyranoside for 4 h at 28 °C according to the pGEX system manual (GE Healthcare). Cell pellets were resuspended in 20 mM Tris, pH 8, 150 mM NaCl, 5 mM dithiothreitol, and 1 mM phenylmethylsulfonyl fluoride and lysed with a French press following by high speed centrifugation and filtration to remove cell debris. The cell-free extracts were loaded onto GSTrap FF columns (GE Healthcare), and bound proteins were eluted with 50 mM MES, pH 6.5, 100 mM NaCl, 5 mM dithiothreitol, 25 mM reduced glutathione.

**Polysaccharide Substrates—**Stock suspensions of bacterial microcrystalline cellulose (BMCC; Monsanto Cellulon, Monsanto Co.) and phosphoryc acid swollen cellulose were prepared as in Ref. 26. Insoluble oat-spelled xylan was prepared as in Ref. 27, and low viscosity (degree of substitution = 0.65–0.9, degree of polymerization = 400) and medium viscosity (degree of substitution = 0.7, degree of polymerization = 1100) CMC were purchased from Sigma. The following polysaccharide substrates were obtained from Megazyme International (Wicklow, Ireland): low viscosity carob galactomannan (Gal/Man = 22:78), sugar beet arabinan (arabinose/Gal/rhamnose/GalUA = 88:3:2:7), amyloid xylglucan (arabinose/Gal/xylene/Glc = 3:16:36:45), low viscosity wheat arabinoxylan (arabinose/xylene = 41:59; Glc, Gal, and Man < 1%), and medium viscosity barley β-glucan (purity > 97% with <0.3% arabinoxylan contamination).

**Binding Assays—**The protocol was adapted from Ref. 28, and cellulosic substrates were prepared as in Ref. 26. Binding assays were carried out at room temperature in siliconized 2.0-mL
New Modular Plant EGases with a CBM

A

Class A

Class B

Class C

B

FIGURE 1. Structural and sequence variation among plant family 9 glycosyl hydrolases. A, schematic representation of modular structure: cytoplasmic domain (dark gray), transmembrane domain (white), signal sequence (black), GH9 catalytic domain (light gray), linker region (thick black line), carbohydrate-binding module (hexagon). Structural subclasses are represented by TomCel9 (Class A, U78526), TomCel1 (Class B, U13054), and SICel9C1 (Class C, AAD08699). B, the SICel9C1 amino acid sequence alignment of the C-terminal 110 amino acids of SICel9C1 with selected orthologs from other plant species and a family 2 CBM from C. fimi xylanase 10A. Three conserved surface-exposed Trp residues (corresponding to Trp17, Trp54, and Trp72 in AtCel9C1) are marked with asterisks. The CBMs comprise the following. SI, SICel9C1, AAD08699; At, Arabidopsis thaliana, At1g64390; Os, Oryza sativa, Os10g06120; Ba, Brassica rapa, Ba09G220; Pt, Populus tomentosa, Pt01G09970; Pp, Physcomitrella patens, Pp25G04650; La, Lycopersicon esculentum, La05G09010; Sz, Solanum tuberosum, Sz06G09010; Sl, S. lycopersicum, Sl05G09010.

RESULTS AND DISCUSSION

Modular Architecture of Plant EGases—EGases from tomato have historically been referred to as TomCel1 to -8; however, we have renamed TomCel8 as SICel9C1, in accordance with the designation of tomato as S. lycopersicum and to conform to the standardized naming scheme used for bacterial EGases (32). This nomenclature provides important information, since, in the case of SICel9C1, the name indicates that this protein is a tomato (Sl) cellulase (Cel) from GH family 9 (9) with a Class C catalytic domain structure (Fig. 1A). Within the plant EGase superfamily, classes A–C correspond to the membrane-anchored, secreted GH9 catalytic module alone and the group with the

Characterization of Enzyme Activity—Hydrolytic activities of the Cel6/Cel9C1 FP, TfCel6A, and the Cel6A CD were assayed as in Refs. 26 and 29, with bacterial microcrystalline cellulose (BMCC, 2.5 mg/ml), low viscosity carboxymethyl cellulose (CMC, 1% w/v), and phosphoric acid-swollen cellulose (ASC, 0.2% w/v) in 0.4 ml of Buffer A at 30 °C for 20, 4, and 2 h, respectively, with 0.4 nmol of protein assay for BMCC and 0.067 nmol for CMC and ASC. Hydrolytic activity of the SICel9C1 CD was quantified as in Ref. 30 in a total volume of 100 μl, containing a final concentration of 0.2% (w/v) of each glycan substrate (Megazyme) in Buffer A, unless otherwise noted, for 4 h at 37 °C. The optimum temperature for SICel9C1 CD activity was determined with 1% (w/v) low viscosity CMC (Sigma) in Buffer A over a range of 25–72 °C for 4 h. The pH profile of SICel9C1 CD activity was determined with 1% (w/v) low viscosity CMC (Sigma) in Buffer A (pH 4–8) for 4 h at 37 °C. To investigate the effect of calcium on activity, 5 mM CaCl2 plus or minus 10 mM EDTA was included in the reaction mixture for 4 h at 37 °C. The substrate specificity of the SICel9C1 CD was assayed (substrates listed in Fig. 6) in 100-μl reactions containing 0.2% (w/v) glycan substrate in Buffer A, unless otherwise noted, for 4 h at 37 °C.

The ability of the SICel9C1 CD to degrade cello-oligosaccharides (cellobiose, G2; -triose, G3; -tetraose, G4; -pentaose, G5; -hexaose, G6; Seikagaku America, Falmouth, MA) and the resulting reaction products was analyzed by TLC on Whatman LKSD 150-A silica gel plates as in Ref. 31 with the exception that the oligosaccharides were separated by two ascents of ethyl acetate/water/methanol (40:15:20, v/v/v).
additional C-terminal domain, respectively (Fig. 1A). Libertini et al. (3) proposed that the class with a putative CBM (Class C; Fig. 1A) is a subgroup nested within the larger group containing just a CD (Class B; Fig. 1A). However, their phylogenetic study was primarily focused on DNA sequences and provided a more evolutionary perspective, taking into account intron/exon organization. The cognate protein sequences clearly show that plant GH9 ligases have a modular organization with three distinct subgroups (Fig. 1A). EGases are probably derived from an ancient eukaryotic ancestor that predates the divergence of eukaryotic kingdoms (33) and are thus ubiquitous. Accordingly, GH9 genes, including members of both Classes A and B, have been identified in many primitive plant taxa, such as mosses, ferns, and cycads (3). The additional presence of an expressed sequence tag encoding a predicted EGase with a similar putative CBM in the moss Physcomitrella patens (accession number BJ591253) further indicates that all three subclasses are present throughout the plant kingdom.

The putative CBM domain of Class C EGases typically has 100–110 amino acids, and BLAST searches of the data bases indicate that these domains are most similar to microbial family 2 CBMs. The amino acid sequences of the putative CBM domain from SlCel9C1 and selected plant orthologs were aligned with the family 2a CBM from Cellulomonas fimii xylanase 10A (Fig. 1B), revealing the conservation of specific residues that have been experimentally determined to be critical for the binding of family 2a CBMs to cellulose (Trp17, Trp54, and Trp72 in CBM2a) (34), as indicated in Fig. 1B by asterisks. However, the low overall degree of amino acid sequence identity (~18%) is below the threshold, estimated to be at least 35% (35), necessary to make conclusions regarding its structure or potential function. Consequently, a biochemical approach was taken to determine whether the putative CBM domain plays a role in carbohydrate binding.

SlCel9C1 CBM Substrate Binding Studies—Numerous attempts to express the full-length SlCel9C1 protein in E. coli or P. pastoris consistently generated two polypeptides with the predicted size of the CD and the CBM but none with the expected size of the native protein. This probably reflects the high susceptibility of the linker region to proteolysis, which can be prevalent in cell cultures (36). Many attempts were made to circumvent this problem, such as varying culture pH, temperature, and media components and the inclusion of various protease inhibitor mixtures without success. Therefore, two alternative strategies were taken to determine whether the C-terminal domain is a functional CBM.

To establish that the SlCel9C1 CBM can potentiate cellulose binding as part of a modular EGase enzyme, a chimeric fusion protein (Cel6/Cel9C1 FP) was generated, comprising the CD of TfCel6A, a well characterized EGase from T. fusca (37) that was engineered to replace its own family 2 CD with the SlCel9C1 CBM. The binding of the Cel6/Cel9C1 FP to two crystalline cellulose substrates, BMCC and Avicel, was compared with that of both the intact TfCel6A and the TfCel6A CD alone. TfCel6A showed the greatest binding to BMCC, with ~80% of the protein bound to the substrate (Fig. 2A). The TfCel6A CD was used in this experiment as a negative control and, as expected, did not bind to BMCC, since it lacks a CBM, whereas at high substrate concentrations, the Cel6/Cel9C1 FP bound to BMCC almost as well as TfCel6A. Thus, under these conditions, the SlCel9C1 CBM conferred equivalent binding to that of the TfCel6A CBM2 and functioned as a discrete cellulose binding module, the first reported example from plant EGases. Similar results were obtained using a gel-based qualitative assay with Avicel as a binding substrate (Fig. 2B).

Effect of SlCel9C1 CBM on Cellulolytic Activity—A key function of EGase CBMs is believed to be the potentiation of cellulase hydrolysis, by increasing the duration and degree of localized association between the CD and its substrate. In order to determine whether this is the case for the SlCel9C1 CBM, the hydrolytic activity of the Cel6/Cel9C1 FP on three cellulose substrates was compared with that of the TfCel6A and the TfCel6A CD alone (Table 1). All three proteins hydrolyzed crystalline BMCC, but the Cel6/Cel9C1 FP and the TfCel6A CD alone had only 29 and 56%, respectively, of the TfCel6A activity. In contrast, TfCel6A and TfCel6A CD had the same activity against ASC, an insoluble, noncrystalline cellulose substrate. Although a CBM is not required for activity on noncrys-

![FIGURE 2. Binding of the purified Cel6/Cel9C1 fusion protein to cellulosic substrates. A, the Cel6/Cel9C1 FP (▲), T. fusca Cel6A (TfCel6A; ○), and T. fusca Cel6A CD (TfCel6A CD; ■) were incubated with different concentrations of BMCC. B, the Cel6/Cel9C1 FP, T. fusca Cel6A, and T. fusca Cel6A CD were incubated with different concentrations of Avicel, after which bound or unbound proteins were separated by SDS-PAGE. Molecular mass markers are shown (kDa).](image-url)
talline substrates, the Cel6/Cel9C1 FP still only had approximately half the specific activity of the other enzymes. One possible explanation for this reduced activity is the charge difference between the two domains of the Cel6/Cel9C1 FP, since the predicted pH values of the TfCel6A CD and CBM are 5.9 and 4.2, respectively, whereas those of the SlCel9C1 CD and CBM domain are 8.1 and 10.1. This large charge difference (4.2 pH units) between the two domains of the FP, which are connected by a flexible linker region, could promote an interdomain association that might hinder substrate accessibility to the active site cleft.

This was investigated by examining the activities of the three proteins with CMC, a soluble, noncrystalline cellulosic polymer (Table 1), since it was reasoned that this single chain soluble polysaccharide would enter more readily into the active site, resulting in greater activities than with BMCC or ASC. This proved to be the case for all three proteins (Table 1), but the activity of the Cel6/Cel9C1 FP was still less than that of TfCel6A or TfCel6A CD, lending support to the idea that steric hindrance at the active site may be responsible for the reduced activities. However, based on the results of the binding data with the Cel6/Cel9C1 FP, the cellulose substrates seem to be fully accessible to the CBM. Another explanation is that the two modules are in a configuration that spatially separates the catalytic domain from the substrate, causing reduced substrate accessibility and, consequently, activity.

Site-directed Mutagenesis of SlCel9C1 CBM—To further examine the nature of the SlCel9C1CBM and to gain important structure-function information, computational modeling was used to identify residues that potentially contribute to cellulose binding. The “3-D Jury” scoring function of the Structure Prediction Meta Server was used to identify probable fold architecture of the SlCel9C1 CBM (38, 39). This method identified two alternative immunoglobulin-like β-sandwich folds, and the structures with scores ranked as the most “significant” were as follows: the family 2 CBM of an exo-1,4-β-D-glycanase from C. fimim (Protein Data Bank code 1EXG) and human ADP-ribosylation factor-binding protein GGA1 (Protein Data Bank code 1NA8). These results suggested that the structure of the SlCel9C1 CBM is distinct from that of known microbial CBMs, but the degree of similarity with the 1EXG microbial CBM allowed general topological features of this domain to be predicted and three-dimensional models to be generated.

A refined model of the SlCel9C1 CBM domain (Fig. 3A), based on the template from the CBM2 of C. fimini xylanase 10A (1EXG), closely matched the features of the β-barrel fold of the parent structure (i.e. only a few short insertions/deletions are present in the final alignment). CBM2 from C. fimii is a member of a larger group of CBMs termed Type A, that bind to surfaces of crystalline substrates via a hydrophobic stacking interaction with ligands mediated by aromatic residues on a flat binding plane (8, 34). The computational model was then used as a guide to identify residues with potentially important roles in cellulose binding, prior to confirmatory site-directed mutagenesis studies. As with the 1EXG template, the model contains a well-defined hydrophobic core, composed of more than five aromatic residues. These included Trp522 of SlCel9C1, which the sequence alignment in Fig. 1B originally suggested might represent one of the cellulose-binding residues (Trp17) of C. fimii CBM2 (1EXG); however, in the predictive model, it corresponds to Trp12 within the hydrophobic core of the CBM. The inferred functionally important residues of SlCel9C1 Trp559 and Trp573 are proposed to align with Trp574 and Trp572 in

### Table 1

| Protein       | Activity (µmol cellobiose/min/µmol protein) |
|---------------|-------------------------------------------|
| T. fusca Cel6A | 0.34 ± 0.01                               |
| T. fusca Cel6A CD | 0.19 ± 0.01                               |
| Cel6A/Cel9C1 FP | 0.11 ± 0.01                               |
| T. fusca Cel9C1 FP | 0.27 ± 0.27                               |
| BMCC ASC CMC  |                                           |
| T. fusca Cel6A | 23.79 ± 1.96                               |
| T. fusca Cel6A CD | 23.52 ± 2.77                               |
| Cel6A/Cel9C1 FP | 12.68 ± 2.90                               |
| T. fusca Cel9C1 FP | 42.90 ± 2.89                               |

FIGURE 3. Site-directed mutagenesis of the SlCel9C1 CBM. A, molecular model of SlCel9C1 CBM highlighting the proposed functionally important residues that were mutated. The image comprises a calcium atom superposition of the best SlCel9C1 CBM model (cyan) on the NMR template, 1EXG, (red). B, binding of the GST-CBM to BMCC. The binding efficiency of the GST-CBM (•) to 0–3 mg/ml BMCC for 3 h at 25 °C was compared with that of GST alone (A). Error bars represent the S.D. of triplicate reactions. C, relative binding efficiency of mutants with single amino acid substitutions (Fig. 3A) to 2 mg/ml BMCC for 3 h at 25 °C compared with GST-CBM (WT).
the template (Fig. 3A), which is consistent with the features of known CBMs (40). The model further suggests that Trp\textsuperscript{529} of SlCel9C1 may be spatially similar to Trp\textsuperscript{17} from 1EXG, thereby representing a third potential binding site (Fig. 3A). It has been shown previously with the C.\textit{fimi} CBM2a that this binding site can be occupied by a Trp or Tyr residue without compromising cellulose binding (34). Interestingly, the Trp\textsuperscript{529} is conserved between CBMs from other plant EGases in Class C, further suggesting an important functional role (Fig. 1B).

To facilitate protein expression and purification for site-directed mutation, the CBM of SlCel9A and related mutated variants were expressed as C-terminal fusion proteins joined to glutathione S-transferase by a 10-amino acid linker (GST-CBM). In a co-incubation assay using affinity-purified proteins, GST-CBM bound to BMCC, whereas GST alone, the negative control, showed no binding (Fig. 3B), demonstrating that the SlCel9C1 CBM also acts as functional cellulose binding module when fused to GST and expressed in \textit{E. coli}.

To determine whether any of the conserved aromatic residues discussed above (Fig. 3A) contribute to the interaction between the SlCel9C1 CBM and cellulose, the following residues were all individually mutated to alanine: Trp\textsuperscript{522}, Tyr\textsuperscript{529}, Trp\textsuperscript{559}, and Trp\textsuperscript{573}. The last three are predicted by the model to be surface-exposed and thus potentially mediate the stacking interaction with crystalline cellulose, whereas Trp\textsuperscript{522} is predicted to be enclosed in the hydrophobic core of the module (Fig. 3A).

The nonconservative substitution of the selected aromatic residues to alanine supported some, but not all, of the predictions based on the structural model. The W573A mutation had the most dramatic effect on binding (Fig. 3C), resulting in less than 10\% of the binding capacity of the unmutated GST-CBM (wild type). Similarly, the W522A and W559A mutants displayed 25 and 30\% reduced binding, respectively. However, the Y529A mutation had no significant effect on binding when compared with wild type (W7; Fig. 3C), indicating that it does not contribute to the interaction with cellulose. The results with the W559A and W573A mutants therefore support the predictions derived from the model. In the case of Trp\textsuperscript{522}, the observed decrease in binding could be due to a loss in stability of the domain due to disruption of the hydrophobic core, or it may be modeled incorrectly and is actually surface-exposed.

Characterization of the SlCel9C1 CD—The \textit{in vivo} substrates of plant EGases have still not been established, and the few \textit{in vitro} studies using various purified native or recombinant isozymes have not shown a consistent pattern of substrate specificity. Most biochemically characterized plant EGases belong to Class B, comprising the secreted GH9 CD, and whereas they typically all have CMCase activity and no activity against crystalline cellulose, different activities have been reported against potential cell wall substrates with internal $\beta$-1,4-Glc linkages, including mixed linkage (1,3),(1,4)-$\beta$-d-glucan (MLG), glucomannan, and xyloglucan (15, 41). The activities of two Class A EGases (\textit{Brassica napus} BnCel16 and poplar PttCel9A) have also been examined with various substrates, and again dissimilarities were identified (42, 43). Both showed high activity on the noncrystalline substrates CMC and ASC but little to none on crystalline cellulose, xyloglucan, MLG, or xylan (42, 43), and only PttCel9A hydrolyzed cello-oligosaccharides.

To date, nothing has been reported regarding the substrate specificity of plant Class C EGases, so the optimum temperature and pH for recombinant SlCel9C1 CD activity was assayed, prior to examining activity against various cell wall glycans and cellulosic substrates. Hydrolysis of low viscosity CMC by SlCel9C1 was optimal at 37 °C (Fig. 4A), so this temperature was used for all further experiments. Many published reports describing plant EGase activity used assay conditions of 25–30 °C (42, 44), and it is interesting to note that the SlCel9C1 CD is less than half as active at these temperatures. The pH profile of SlCel9C1 was also characterized, and optimal activity was seen between pH 4.5 and 6.0 (Fig. 4B), which is similar to results obtained with previously characterized Class A plant EGases (42, 43). We also observed that calcium was required for activity and that, conversely, a calcium chelator inhibited activity (data not shown). When substrate specificity was assayed under optimal pH and temperature conditions, the highest activity was seen with barley MLG, followed by arabinoxylan, medium viscosity CMC, and low viscosity CMC, whereas there
was negligible activity with arabinan and tamarind xyloglucan (Fig. 5). No activity was detectable with BMCC or xyloglucan from tomato fruits or tomato suspension-cultured cells (data not shown).

The hydrolysis of cello-oligosaccharides (G2–G6) by SlCel9C1 CD was assessed by TLC (Fig. 6). The highest activity was seen with cellohexaose (G6), followed by markedly less activity on cellopentaose (G5) and cellotetraose (G4). The hydrolysis products were as follows: G6 digestion generated G3, G4, and G2; G5 was cleaved to G3 and G2; and hydrolysis of G4 produced G2 and G3 (Fig. 6). The results are consistent with previous studies of plant GH9 EGases from Classes A and B that appeared to have CD binding subsites with a higher affinity for at least six consecutive 1,4-β-linked Glc units (43). Plant Class A EGases have also been shown only to cleave G5 and G6 (42, 45). However, the additional activity we observed with the Class C SlCel9C1 CD on G4 has not previously been reported. This result confirms the previous suggestion (42) that the presence of Trp<sup>316</sup> in the catalytic cleft of Class C plant EGases, which is the only class that retains a Trp in this position, might facilitate cleavage of G4. To further corroborate the TLC data, MALDI-TOF mass spectrometry was used to characterize the products resulting from G5 digestion. This confirmed that G3 and G2, but no additional saccharides, were generated (data not shown). We also noted that the G6 commercial substrate contained a small amount of G7, which therefore did not result from transglycosylation activity (Fig. 6, sample 6).

The SlCel9C1 CD has a broad substrate specificity when compared with those of previously studied Class A or B plant EGases. A wide substrate range is not uncommon for microbial GH9 enzymes (42, 46), and xylanase activity has previously been detected among members of the GH9 family in microbes. We originally detected some hydrolytic activity on commercially obtained carob galactomannan, as determined by measuring reducing groups (data not shown). However, no depolymerization of galactomannan was observed by subsequent viscometric analysis, and the enzyme generated no reaction products when incubated with pure 6<sup>α</sup>,6<sup>β</sup>-D-galactosyl-man-nopentaose and assayed by MALDI-TOF mass spectrometry (data not shown). The hydrolytic activity may therefore have resulted from contamination of the commercial galactomannan with a small amount of an unknown polysaccharide. The high activity with barley MLG contrasts with the previously reported low activity exhibited by poplar Class A EGase on lichenan, another MLG substrate (42). However, barley β-glucan MLG has longer stretches of β-1,4-glucan between the β-1,3-glucosidic bonds, which may allow it to serve as a better substrate. Another Class A enzyme, B. napus Cel16, was also reported to have negligible activity on barley MLG (43). The minimal activity seen with xyloglucan agrees with previous studies of plant EGases (41–43) and probably reflects the infrequency of sufficiently contiguous stretches of unsubstituted 1,4-β-linked Glc residues, although it is interesting that tamarind xyloglucan was a slightly better substrate than tomato xyloglucan (data not shown), although the former shows a greater degree of side chain branching (47). The structurally similar TfCel9A also lacks activity on xyloglucan, suggesting that the high level of branching may interfere with access to the catalytic cleft (42).
This paper provides the first report of a plant EGase (SlCel9C1) with a functional, modular CBM that confers binding to crystalline cellulose. By analogy with microbial studies, this suggests that Class C plant EGases play a role in facilitating cellulose degradation. One possibility is that they function in processes associated with irreversible wall disassembly, such as fruit softening and organ abscission. This idea is supported by the observation that SlCel9C1 transcript abundance increases in ripening fruit coincident with rapid wall degradation (data not shown). However, it is notable that the SlCel9C1 substrate specificity in vitro appears to be broader than most known GH9 enzymes. Alternatively, Class C EGases might function to hydrolyze polysaccharide chains at the cellulose microfibril periphery, including amorphous or paracrystalline cellulose chains and other associating polymers. Indeed, it was reported that a subset of xyloglucan polymers is tightly bound to the microfibril surface and is thus inaccessible to a xyloglucanase that does not have a CBM (48). Although the balance of evidence suggests that most plant GH9 EGases do not hydrolyze xyloglucans (15, 41), this conclusion is based almost exclusively on in vitro assays with nonnative substrates. Furthermore, xyloglucan may adopt conformations in muro that are more susceptible to attack. One study using transgenic plants also suggested that plant EGases do not hydrolyze xyloglucans in vivo; however, this involved a Class B EGase without a CBM (49). The conformation and orientation of glycans is likely to be profoundly influenced by their interaction with cellulose (49, 50), so the results of in vitro analyses should be interpreted carefully.

A third scenario is that the CBM may function principally to target the CD to the substrate of interest to facilitate modification of cell wall microdomains following proteolytic separation of the CD and CBM modules. This type of hydrolase targeting mechanism has been proposed for a modular xylanase (51), and post-translational proteolysis has been suggested as an activation mechanism for another plant wall-loosening protein, β-expansin (52).

Last, Class C EGases might be involved in wall assembly (e.g. by regulating cellulose crystallinity during biosynthesis) and thus play a role in cell expansion. It has been shown that the application of exogenous bacterial CBMs to plant tissue can lead to increased growth (53), and transgenic tobacco plants expressing a bacterial CBM were reported to grow more rapidly and produce more biomass than their wild type counterparts (54). This phenomenon was attributed to the CBM interfering with microfibril biosynthesis and crystallization.

The expression of plant Class C EGase genes has been associated with both degradative processes, such as fruit softening and abscission (20, 21), and cell elongation (22), so these proteins may have multiple physiological functions. Studies are now in progress to address this question.

Acknowledgment—We thank Dr. L. Walker for helpful advice.

REFERENCES
1. Lynd, L. R., Van Zyl, W. H., and Pretorius, I. S. (2002) Microbiol. Mol. Biol. Rev. 66, 506–515
2. Hildén, L., and Johansson, G. (2004) Biotech. Lett. 26, 1683–1693
3. Libertini, E., Li, Y., and McQueen-Mason, S. J. (2004) J. Mol. Evol. 58, 506–515
4. Lynd, L. R., Van Zyl, W. H., McBride, J. E., and Laser, M. (2005) Curr. Opin. Biotechnol. 16, 577–583
5. Rabinovich, M. L., Melnik, M. S., and Bolobova, A. V. (2002) Biochemistry (Mosc.) 67, 1026–1050
6. Bayer, E., Belaich, J. P., Shoham, Y., and Lamed, R. (2004) Annu. Rev. Microbiol. 58, 521–554
7. Wilson, D. B., and Irwin, D. C. (1999) Adv. Biochem. Eng. Biotechnol. 65, 1–21
8. Boraston, A. D., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Biochem. J. 382, 769–781
9. Linder, M., and Teeri, T. T. (1997) J. Bacteriol. 179, 15–28
10. Boraston, A. B., Kwan, E., Chiu, P., Warren, A. J., and Kilburn, D. G. (2002) J. Biol. Chem. 278, 6120–6127
11. Carrard, G., Koivula, A., Soderlund, H., and Beguin, P. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 10342–10347
12. Jung, H., Wilson, D. B., and Walker, L. P. (2003) Biotech. Bioeng. 84, 151–159
13. Brummell, D. A., Lashbrook, C. C., and Bennett, A. B. (1994) An. Chem. Soc. Symp. Ser. 566, 100–129
14. Malhoj, M., Pagant, S., and Höfte, H. (2002) Plant Cell Physiol. 43, 1399–1406
15. Rose, J. K. C., Catalá, C., Gonzalez-Carranza, Z. H., and Roberts, J. A. (2003) in The Plant Cell Wall (Rose, J. K. C., ed) pp. 264–324, Blackwell Publishing, Oxford, UK
16. Coutinho, P. M., and Henrisrat, B. (1999) in Recent Advances in Carbohydrate Bioengineering (Gilbert, H. J., Davies, G., Henrisrat, B., and Svensson, B., eds) pp. 3–122, Royal Society of Chemistry, Cambridge, UK
17. Henrisrat, B., Coutinho, P. M., and Davies, G. J. (2001) Plant Mol. Biol. 47, 55–72
18. Robert, S., Bichet, A., Grandjean, O., Kierzukowski, D., Satiat-Jeunemaitre, B., Pelletier, S., Hauser, M.-T., Höfte, H., and Vernhettes, S. (2005) Plant Cell 17, 3378–3389
19. Catalá, C., and Bennett, A. B. (1998) Plant Physiol. 118, 1535
20. Trainotti, L., Spolacco, S., Pavanello, A., Baldan, B., and Casadoro, G. (1999) Plant Mol. Biol. 40, 323–332
21. Trainotti, L., Pavanello, A., and Zanin, D. (2006) J. Exp. Bot. 57, 589–598
22. Arpat, A. B., Waugh, M., Sullivan, J. P., González-M, F., Frisch, D., Main, D., Wood, T., Leslie, A., Wing, R., and Wilkins, T. A. (2004) Plant Mol. Biol. 54, 911–929
23. Salminen, O. (2002) Development of an Overexpression System for Production of Recombinant “Thermobifida fusca” Cellulose Cel6A and Its Multi-dimensional NMR Analysis. Ph.D. thesis, Cornell University, Ithaca, NY
24. Sali, A., and Blundell, T. L. (1993) J. Mol. Biol. 234, 779–815
25. Sali, A., Potterton, L., Yuan, F., Van Vlijmen, H., and Karplus, M. (1995) Proteins 23, 318–326
26. Irwin, D., Walker, L., Spezio, M., and Wilson, D. B. (1993) Biotechnol. Bioeng. 42, 1002–1013
27. Kim, J. H., Irwin, D., and Wilson, D. B. (2004) Can. J. Microbiol. 50, 835–843
28. Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (1998) J. Bacteriol. 180, 1709–1714
29. Ghose, T. K. (1987) Pure Appl. Chem. 59, 257–268
30. Lever, M. (1972) Anal. Biochem. 47, 272–279
31. Jung, E. D., Lao, G., Irwin, D., Barr, B. K., Benjamin, A., and Wilson, D. B. (1993) Appl. Environ. Microbiol. 59, 3032–3043
32. Henrisrat, B., Teeri, T. T., and Warren, R. A. J. (1998) FEBS Lett. 425, 352–354
33. Davison, A., and Blaxter, M. (2005) Mol. Biol. Evol. 22, 1273–1284
34. McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A., and Kilburn, D. G. (2000) Protein Eng. 13, 801–809
35. Sánchez, R., and Sá, S. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 13597–13602
36. Jahic, M., Gustavsson, M., Jansen, A.-K., Martinelle, M., and Enfors, S.-O. (2003) J. Biotechnol. 102, 45–53
37. Wolfgang, D. E., and Wilson, D. B. (1999) Biochemistry 38, 9746–9751
38. Bujnicki, J. M., Fleoﬀson, A., Fischer, D., and Rychlewski, L. (2001) Bioinformatics 17, 750–751

New Modular Plant EGases with a CBM
New Modular Plant EGases with a CBM

39. Ginalsiki, K., Elofsson, A., Fischer, D., and Rychlewski, L. (2003) Bioinformatics 19, 1015–1018
40. Xu, G.-Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E., and Harvey, T. S. (1995) Biochemistry 34, 6993–7009
41. Brummell, D. A., and Harpster, M. H. (2001) Plant Mol. Biol. 47, 311–340
42. Master, E. M., Rudsander, U. J., Zhou, W., Henriksson, H., Divine, C., Denman, S., Wilson, D. B., and Teeri, T. T. (2004) Biochemistry 43, 10080–10089
43. Mølhøj, M., Ulvskov, P., and Degan, F. D. (2001) Plant Physiol. 127, 674–684
44. Woolley, L. C., James, D. J., and Manning, K. (2001) Planta 214, 11–2145
45. Maclachlan, G. (1988) Methods Enzymol. 160, 382–391
46. Eckert, K., Zielinski, F., Lo Leggio, L., and Schneider, E. (2002) Appl. Microbiol. Biocatal. 60, 428–436
47. York, W. S., Kolli, V. S. K., Orlando, R., Albersheim, P., and Darvill, A. G. (1996) Carbohydr. Res. 285, 99–128
48. Pauly, M., Albersheim, P., Darvill, A., and York, W. S. (1999) Plant J. 20, 629–639
49. Harpster, M. H., Brummell, D. A., and Dunsmuir, P. (2002) Plant Mol. Biol. 50, 345–355
50. Rose, J. K. C., and Bennett, A. B. (1999) Trends Plant Sci. 4, 176–183
51. Chen, N. J., and Paull, R. E. (2003) Funct. Plant Biol. 30, 433–441
52. Downes, B. P., Steinbaker, C. R., and Crowell, D. N. (2001) Plant Physiol. 126, 244–252
53. Shpigel, E., Roiz, L., Goren, R., and Shoseyov, O. (1998) Plant Physiol. 117, 1185–1194
54. Shoseyov, O., Shami, Z., and Shpigel, E. (February 6, 2001) U. S. Patent 6,184,440