Correction

BIOCHEMISTRY

Correction for "Structural insights into a unique cellulase fold and mechanism of cellulose hydrolysis," by Joana L. A. Brás, Alan Cartmell, Ana Luísa M. Carvalho, Genny Verzé, Edward A. Bayer, Yael Vazana, Márcia A. S. Correia, José A. M. Prates, Supriya Ratnaparkhe, Alisdair B. Boraston, Maria J. Romão, Carlos M. G. A. Fontes, and Harry J. Gilbert, which appeared in issue 13, March 29, 2011, of *Proc Natl Acad Sci USA* (108:5237–5242; first published March 10, 2011; 10.1073/pnas.1015006108).

The authors note the following statements should be added to the Acknowledgments: "The authors thank Dr. Ethan Goddard-Borger for synthesizing the substrate 2,4-dinitrophenyl-β-D-cellotriose and potential inhibitors of CtCel124. The authors also thank Steve Withers and Gideon Davies for helpful discussions."

Additionally, the grant number 00562492 should instead appear as DE-FG02-09ER16076.

www.pnas.org/cgi/doi/10.1073/pnas.1104766108

Structural insights into a unique cellulase fold and mechanism of cellulose hydrolysis

Joana L. A. Brás[a,1], Alan Cartmell[b,1], Ana Luísa M. Carvalho[c,2], Genny Verzé[c,d], Edward A. Bayer[e], Yael Vazana[e], Márcia A. S. Correia[a], José A. M. Prates[a], Supriya Ratnaparkhe[b], Alisdair B. Boraston[f], Maria J. Romão[c], Carlos M. G. A. Fontes[a], and Harry J. Gilbert[b,2]

[a]Centro de Investigação Interdisciplinar em Sanidade Animal, Faculdade de Medicina Veterinária, Pólo Universitário do Alto da Ajuda, Avenida da Universidade Técnica, 1300-477 Lisbon, Portugal; [b]Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602-4712; [c]Rede de Química e Tecnologia, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal; [d]Biocrystallography Laboratory, Department of Biotechnology, University of Verona, 37129 Verona, Italy; [e]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100 Israel; and [f]Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada V8W 3P6

Edited by Arnold L. Demain, Drew University, Madison, NJ, and approved February 1, 2011 (received for review October 8, 2010)

Clostridium thermocellum **is a well-characterized cellulose-degrading microorganism. The genome sequence of *C. thermocellum* encodes a number of proteins that contain type I dockerin domains, which implies that they are components of the cellulose-degrading apparatus, but display no significant sequence similarity to known plant cell wall–degrading enzymes. Here, we report the biochemical properties and crystal structure of one of these proteins, designated *Ct*Cel124. The protein was shown to be an *endo*-acting cellulase that displays a single displacement mechanism and acts in synergy with Cel48S, the major cellulosomal *exo*-cellulase. The crystal structure of *Ct*Cel124 in complex with two cellotriose molecules, determined to 1.5 Å, displays a superhelical fold in which a constellation of α-helices encircle a central helix that houses the catalytic apparatus. The catalytic acid, Glu96, is located at the C-terminus of the central helix, but there is no candidate catalytic base. The substrate-binding cleft can be divided into two discrete topographical domains in which the bound cellotriose molecules display twisted and linear conformations, respectively, suggesting that the enzyme may target the interface between crystalline and disordered regions of cellulose.**

The plant cell wall is an important biological and industrial resource. Deconstruction of this composite structure provides nutrients that are utilized by microorganisms from a variety of ecosystems. Indeed, mammalian herbivores derive a significant proportion of their energy from the hydrolysis of plant cell wall polysaccharides by their symbiotic microbiota. The deconstruction of the plant cell wall is also of growing environmental and industrial significance as the demand for renewable sources for bioenergy and substrates for the chemical industry increases (1). The major plant cell wall polysaccharide is cellulose, a β-1,4-glucose polymer (2), which is hydrolyzed by a range of glycoside hydrolases (cellulases). These enzymes display *endo* (endo-β1, 4-glucanase), *exo* (cellobiohydrolases that release cellobiose from cellulose), or *endo*-processive (cleaves internally and then acts in a processive manner on the generated product) modes of action. The classical paradigm for cellulose hydrolysis is the endoglucanase-cellobiohydrolase synergy model, which is based, primarily, on aerobic fungal cellulase systems (see ref. 3 for review). This model, however, does not accurately reflect clostridial systems, where endo-processive GH9 enzymes are central to the degradative process (4) (see below), or *Cytophaga hutchinsonii*, which appears to lack classical cellobiohydrolases. Cellulases are currently grouped into 11 of the 120 glycoside hydrolase sequence-based families (GHs) within the Carbohydrate-Active enZymes (CAZy) database (5). Because there is limited conservation in the catalytic apparatus and the overall fold between the cellulase-containing families, these enzymes are generally thought to have evolved by convergent evolution (3).

Clostridium thermocellum is a well-characterized cellulose-degrading microorganism (6, 7). The bacterium synthesizes a large multienzyme complex, known as the "cellulosome," which catalyzes the degradation of the plant cell wall (6, 7). Enzymes are recruited into the *C. thermocellum* cellulosome through the interaction of their type I dockerin modules with the multiple cohesin domains present on the scaffoldin (defined as CipA) (reviewed in refs. 6 and 7). The genome of *C. thermocellum* encodes 72 proteins containing type I dockerins. These proteins, therefore, are likely to be components of the cellulosome and thus contribute to cellulose or, in a wider context, plant cell wall deconstruction. Synergy experiments have identified two cellulosomal enzymes, an *exo*-acting GH48 cellobiohydrolase that acts from the reducing end of cellulose chains and the cellotetraose-producing *endo*-processive GH9 endoglucanase, Cel9R (8), as central components of the *C. thermocellum* cellulase system (9). It has been suggested that, by generating cellotetraose as the major product from cellulose, likely through the action of Cel9R, *C. thermocellum* minimizes the utilization of ATP during import of glucose units (10). The cellulase activity obtained by combining cellulosomal enzymes in vitro, however, is considerably lower than that displayed by the cellulosome presented on the surface of *C. thermocellum*. It is possible, therefore, that proteins, currently of unknown function, either in the cellulosome or displayed on the surface of *C. thermocellum*, make a significant contribution to the cellulose-degrading capacity of the bacterium. Although most of the cellulosomal proteins can be assigned to glycoside hydrolase, esterase, or polysaccharide lyase families, 14 of the predicted type I dockerin containing proteins display little sequence similarity to enzymes in the CAZy database. Thus, some of these non-CAZy *C. thermocellum* proteins may comprise novel enzymes that target the hydrolysis of components of the plant cell wall such as cellulose. One of these hypothetical proteins, Cthe_0435 (hereafter designated as *Ct*Cel124), is upregulated when *C. thermocellum* is cultured on crystalline cellulose (11), suggesting that *Ct*Cel124 plays a role in the hydrolysis of the glucose polymer.

Here, we show that *Ct*Cel124 is an endoglucanase that acts in synergy with the major exocellulase of the *C. thermocellum* cel-

Author contributions: E.A.B., M.J.R., C.F., and H.J.G. designed research; J.L.A.B., A.C., A.L.M.C., G.V., Y.V., M.C., J.A.M.P., and S.R. performed research; A.C., A.L.M.C., E.A.B., Y.V., A.B.B., M.J.R., C.M.G.A.F., and H.J.G. analyzed data; and A.L.M.C., C.M.G.A.F., and H.J.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Coordinates and observed structure factor amplitudes for the *Ct*Cel124-cellotriose (2) complex, to 1.5 Å resolution, have been deposited in the Protein Data Bank in Europe (PDBe), www.ebi.ac.uk/pdbe (PDB ID code 2XQO).

[1]J.B. and A.C. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: hgilbert@ccrc.uga.edu or alcarvalho@dq.fct.unl.pt.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1015006108/-/DCSupplemental.

lulosome to degrade crystalline cellulose. The crystal structure of *Ct*Cel124 reveals a substrate-binding cleft in which the bound cellooligosaccharides adopt two distinct conformations, indicating that the enzyme targets the interface between crystalline and amorphous regions of cellulose. The active site of the cellulase displays structural conservation to GH23 enzymes, a family that contains inverting lysozymes and lytic transglycosylases.

Results and Discussion

Catalytic Properties of CtCel124. *Ct*Cel124 is highly upregulated when *C. thermocellum* is cultured on crystalline cellulose (11), suggesting the protein may contribute to the metabolism of the poylsaccharide. To test this hypothesis, the biochemical properties of the 220 residue C-terminal module of the protein (designated $Ct\text{Cel124}_{CD}$) was assessed. The data, summarized in Table 1, show that the enzyme hydrolyzed barley β-glucan, a β-1,3-β-1,4 mixed linked glucan, phosphoric acid swollen cellulose (PASC), and carboxymethylcellulose. The specific activity of the enzyme against β-glucan was only fourfold higher than the value for PASC. The initial reaction products released from PASC ranged from cellotriose to cellohexaose (Fig. 1). Such a profile is typical of *endo*-acting enzymes, and thus *Ct*Cel124 appears to be an endo-β-1,4-glucanase. The difference in activity between the soluble and insoluble polysaccharides is relatively modest compared to, for example, GH5 endoglucanases and GH9 *endo*-processive endoglucanases, which generally display a much stronger preference for β-glucan (12).

To explore the capacity of *Ct*Cel124 to disrupt the plant cell wall, the catalytic module was incubated with sections of *Arabidopsis* stem, which were subsequently stained by Calcofluor White that binds predominantly to cellulose, and the family 9 carbohydrate binding module (CBM9) fused to green fluorescent protein. CBM9 binds to the reducing end of cellulose and xylan chains and thus provides a direct readout of cellulose hydrolysis (13). The Calcofluor White staining data (Fig. 1) show that the primary cell walls were considerably thinner, and significantly disrupted, after cellulase treatment. Although CBM9 did not bind to untreated cell walls, after incubation with $Ct\text{Cel124}_{CD}$, the protein stained the secondary and primary cell walls (Fig. 1). These data indicate that *Ct*Cel124 is able to attack cellulose embedded in cell walls. These promising data suggest that combining *Ct*Cel124, with other *exo*- and other *endo*-acting cellulases, may have a significant effect on cellulose degradation within intact cell walls.

$Ct\text{Cel124}_{CD}$ was approximately 20-fold more active against cellohexaose than cellopentaose (Table 1) but displayed no activity against cellotetraose. $Ct\text{Cel124}_{CD}$ hydrolyzed cellohexaose predominantly to cellotriose, whereas cellopentaose was converted exclusively to cellobiose and cellotriose (Fig. S1). These data indicate that *Ct*Cel124 contains six dominant subsites extending from −3 to +3 (defined using standard nomenclature ref. 14). The small amount of cellotetraose and cellobiose, released from cellohexaose, is consistent with the weak −4 subsite (binding of cellohexaose from subsites −4 to +2 will generate cellotetraose and cellobiose), identified through the crystallization of the enzyme in complex with substrate (see below).

Affinity of $Ct\text{Cel124}_{CD}$ for Cellulose and Cellooligosaccharides. The affinity of the inactive catalytic acid mutant (E96A) of *Ct*Cel124 for cellohexaose and regenerated cellulose (RC) was determined by isothermal titration calorimetry (ITC) and depletion isotherms (Fig. S2 and Table S1). The affinity (association constant, K_A) for cellohexaose is $1.5 \pm 0.07 \times 10^4$ M^{-1} at 40 °C. Depletion binding isotherms showed that E96A had a K_A for RC of 3.9 $(\pm 0.5) \times 10^5$ M^{-1} at 40 °C. Thus, the affinity of the enzyme for RC is approximately 10-fold higher than for cellohexaose, which spans the substrate-binding cleft, suggesting that the enzyme is tailored to the conformation adopted by, at least, some regions of RC, and is not optimized to bind to the twisted structure adopted by cellooligosaccharides in solution (discussed within a structural context below).

Synergy Between GH48S and CtCel124. It is well established that *exo*- and *endo*-acting cellulases act in synergy to hydrolyze cellu-

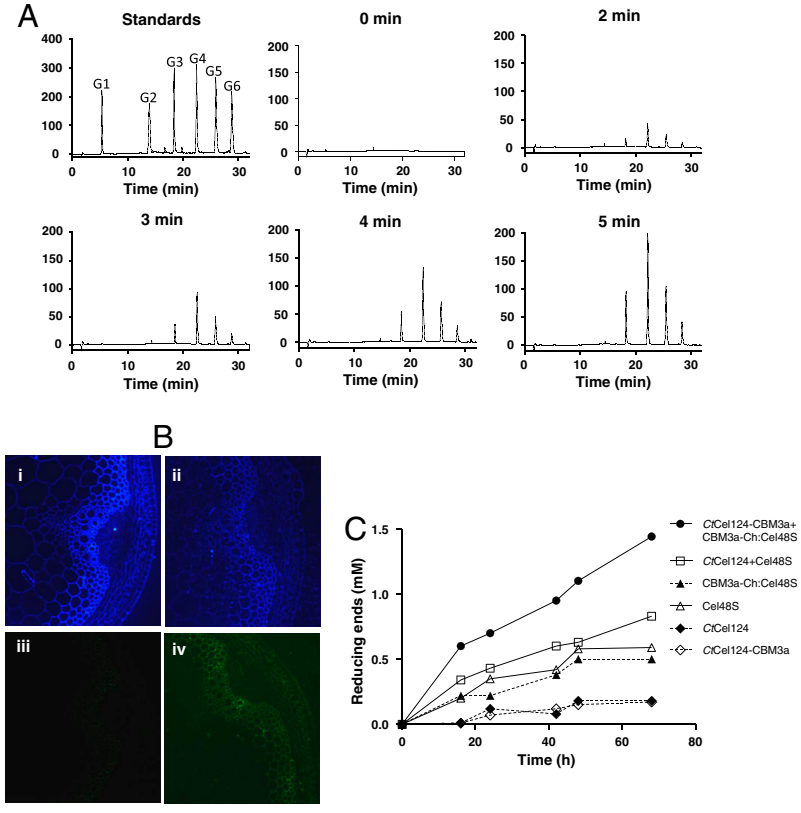

Fig. 1. Catalytic activity of $Ct\text{Cel124}_{CD}$. (*A*) The HPAEC analysis of the reaction products generated by $Ct\text{Cel124}_{CD}$ from phosphoric acid swollen cellulose. The cellulose at 5% (wt/vol) was incubated with 2 μM of $Ct\text{Cel124}_{CD}$, and at various time points the release of cellulooligosaccharides was analyzed by HPAEC. (*B*) The hydrolysis of sections of Arabidopsis stem tissue. Sections *i* and *iii* are untreated, whereas *ii* and *iv* were incubated with 2 μM $Ct\text{Cel124}_{CD}$ for 16 h. Sections *i* and *ii* were stained with Calcofluor White, whereas sections *iii* and *iv* were probed with CBM9 fused to GFP. (*C*) The kinetics of Avicel hydrolysis by $Ct\text{Cel124}_{CD}$ and Cel48S. Dashed lines indicate a single enzyme activity, and solid lines indicate activity of the two enzymes in combination. In Cel48S : CBM3a-Coh, Cel48S (which contains a type I dockerin) was preincubated with CBM3a-Ch (CBM3a fused to a type I cohesin) to generate Cel48S attached to CBM3a through the dockerin-cohesin interaction.

Table 1. Catalytic activity of wild type and mutants of Cel124$_{CD}$

Enzyme	Substrate	Specific activity *	k_{cat}/K_M, min^{-1} M^{-1}
Cel124$_{CD}$	β-glucan	$1.3 \times 10^4 \pm 4.0 \times 10^3$	—[†]
Cel124$_{CD}$	PASC	$3.2 \times 10^3 \pm 5.3 \times 10^2$	—
Cel124$_{CD}$	carboxymethylcellulose	$9.6 \times 10^2 \pm 1.4 \times 10^2$	—
Cel124$_{CD}$	Avicel	$7.4 \times 10^{-1} \pm 2.8 \times 10^{-2}$	
Cel124$_{CD}$	lichenan	$6.1 \times 10^3 \pm 2.1 \times 10^2$	
Cel124$_{CD}$	chitin	NA[‡]	
Cel124$_{CD}$	chitosan	NA	
Cel124$_{CD}$	cellohexaose[§]	—	$1.0 \times 10^4 \pm 2.0 \times 10^2$
Cel124$_{CD}$	cellopentaose	—	$4.5 \times 10^2 \pm 8.5 \times 10^1$
Cel124$_{CD}$	cellotetraose	—	NA
Cel124$_{CD}$	chitohexaose	—	NA
Cel124$_{CD}$	2, 4 DNP-cellotriose	—	$1.1 \times 10^2 \pm 3.6 \times 10^0$
Cel124$_{CD}$ E96A	β-glucan	NA	
Cel124$_{CD}$ E96A	cellohexaose	—	NA
Cel124$_{CD}$ E96A	2, 4 DNP-cellotriose	—	$7.9 \times 10^1 \pm 1.2 \times 10^1$
Cel124$_{CD}$ N188A	cellohexaose	—	$1.4 \times 10^2 \pm 3.1 \times 10^1$
S110A	cellohexaose	—	$4.1 \times 10^3 \pm 1.0 \times 10^2$
S110E	cellohexaose	—	$5.5 \times 10^3 \pm 6.6 \times 10^2$
S110D	cellohexaose	—	$3.2 \times 10^3 \pm 3.7 \times 10^2$

*Specific activity is expressed as molecules of reducing sugar produced per molecule of enzyme per minute. Assays were carried out at 50 °C using 0.5% substrate for soluble polysaccharides and 2% substrate for insoluble polysaccharides.

[†]Dash (—) indicates activity not assessed

[‡]NA, no activity detected.

[§]The substrate concentration used was 30 μM, which is $\ggg K_M$ as the increase in rate was directly proportional to substrate concentration up to 100 μM and thus provides a direct readout of k_{cat}/K_M.

lose (see ref. 3 for review). In addition, these enzymes normally contain cellulose-specific CBMs that potentiate catalysis by recruiting the cellulases to the surface of the insoluble substrate (15). In the *C. thermocellum* cellulosome, the most abundant *exo*-acting cellulase is Cel48S, whereas the crystalline cellulose-specific CBM (CBM3a) is supplied by the noncatalytic scaffoldin CipA (6, 7). To explore the possible synergy between *Ct*Cel124 and Cel48S, the capacity of the enzymes, individually and in combination, to release reducing sugar from Avicel was assessed. The data (Fig. 1) showed that 1.3-fold more reducing sugar was released when the two enzymes were used in combination, compared to the additive value when the two enzymes were used in isolation. These data indicate that *Ct*Cel124 and Cel48S exhibit a degree of synergy when acting on highly crystalline cellulose. When both enzymes were appended to CBM3a, which binds to crystalline cellulose, more extensive synergy (1.9-fold) was observed between the two cellulases. Thus, it is possible that the CBM may target Cel48S and *Ct*Cel124 to similar regions of Avicel and, by so doing, potentiate the synergy between the two enzymes.

The observed synergy between Cel48S and *Ct*Cel124 is consistent with previous studies showing similar potentiation in cellulose hydrolysis when *endo*- or *endo*-processive cellulases were combined with the GH48 *exo*-acting cellulase (9, 16). The mechanisms by which *endo*- and *exo*-acting cellulases act in synergy have been extensively explored. The favored model, at least for fungal systems, proposes that the *endo*-acting enzymes target amorphous regions of cellulose creating new termini from which *exo*-acting cellobiohydrolases can extend substrate hydrolysis into the crystalline regions of the polysaccharide (3). Such a model, however, does not explain the low degree of synergy observed between some enzyme combinations, suggesting that new termini generated by endoglucanases are not always available to the cellobiohydrolases. Indeed, the significance of the *C. thermocellum* GH48 enzymes, in the capacity of the cellulosome to solubilize crystalline cellulose, has been questioned by recent studies on mutants of the bacterium lacking these enzymes. The growth rate of the GH48 knockout mutants on cellulose, the cell yield of the variants, and the activity of the cellulosome against Avicel were reduced by 40, 60, and 35%, respectively, compared to wild-type *C. thermocellum* (17). These data suggest that Cel48S certainly contributes to cellulose degradation, but the classical *endo–exo* synergy model does not fully explain the capacity of the *C. thermocellum* cellulosome to completely solubilize crystalline cellulose.

It should be emphasized that although *Ct*Cel124 contains a type I dockerin, the module displays a preference for the cohesin in the cell-envelope protein OlpC, rather than the cohesins in the cellulosome scaffold protein, CipA (18). These data indicate that the enzyme is predominantly located at the cell surface of the bacterium. Lynd and coworkers showed that the cellulosome, when appended to the surface of *C. thermocellum*, is more efficient at cellulose degradation than when the complex is released into the culture media (19). Thus, it is possible that this increased activity reflects synergistic interactions between catalytic components of the cellulosome and enzymes, such as *Ct*Cel124, which are predicted to be directly appended to the surface of the bacterium. For example, *Ct*Cel124 may create the chain ends at the amorphous-crystalline interface (see below) that are required by the cellulosomal *exo*-acting enzymes to hydrolyze cellulose.

Crystal Structure of CtCel124. The X-ray crystal structure of *Ct*Cel124$_{CD}$ was determined in complex with two cellotriose (which arose through the hydrolysis of cellohexaose during crystallization). The crystal structure revealed a 210 amino acid α-helical protein containing eight α-helices and a small β-sheet comprising three antiparallel β-strands. α-Helix-4 (α-H4) forms the hydrophobic core of the protein, and the other seven helices encircle the core helix (Fig. 2). Thus, *Ct*Cel124$_{CD}$ appears to display an α_8 superhelical fold. Such a structure has not previously been observed in cellulase families, which display the following folds: $(\beta/\alpha)_8$-barrel (GH5, GH51, and GH44), distorted α/β-barrel (GH6), β-jelly roll fold (GH7 and GH12), $(\alpha/\alpha)_6$-barrel (GH8, GH9, and GH48), β_6-barrel (GH45), and sevenfold β-propeller (GH74). It would appear, therefore, that the different cellulase families are in general the result of (functional) convergent evolution, a view reinforced by the superhelical fold displayed by the endoglucanase *Ct*Cel124.

Structural Similarity of CtCel124 to Other Glycoside Hydrolases. A BLAST search reveals four proteins in the UNIPROT database that display significant sequence identitites (>44%) with *e* values $<e^{-36}$ (Fig. S3). We therefore propose that *Ct*Cel124 comprises the founding member of a CAZy family designated GH124. These homologous enzymes are derived from highly cellulolytic

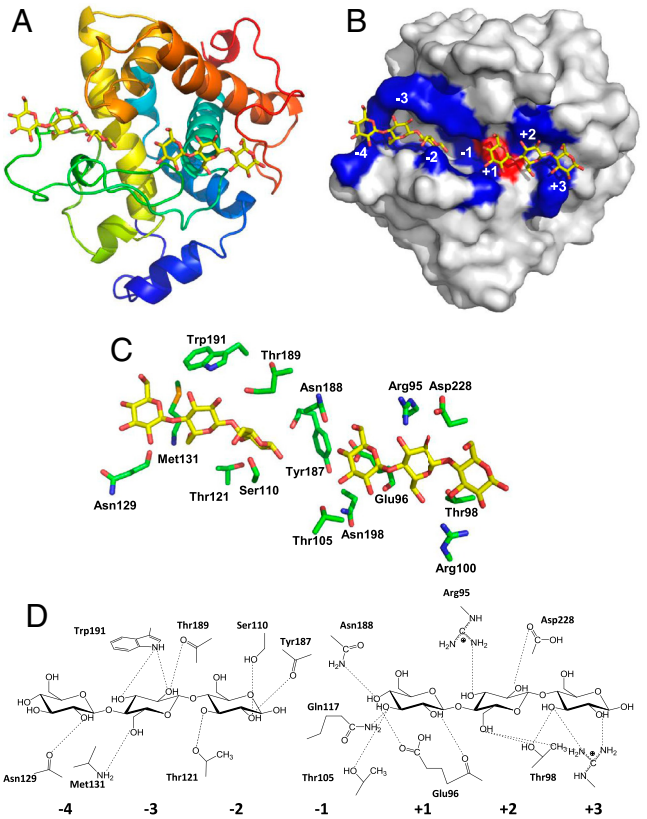

Fig. 2. Crystal structure of *Ct*Cel124. (*A*) The crystal structure of *Ct*Cel124$_{CD}$ color ramped from the N-terminus (blue) to the C-terminus (red) with cellotriose shown in stick format. (*B*) The solvent accessible surface of *Ct*Cel124$_{CD}$ in complex with two cellotriose molecules occupying the subsites indicated. Amino acids that make direct interactions with the ligands are shaded blue, and the catalytic acid Glu96 is in red. (*C*) The location of the residues that make direct polar contact with the two cellotriose molecules. The backbone carbonyl (Glu96, Asn126, Thr187, Thr189) and amides (Met131) that make hydrogen bonds with the substrate are included. Amino acids (carbons in green) and ligand (carbons in yellow) are shown in stick format. (*D*) A schematic of the direct polar interactions between the cellulase and the two ligand molecules. This figure and subsequent structural figures were constructed using PyMol (http://www.pymol.org/).

organisms that also assemble its plant cell wall degrading apparatus into cellulosomal structures. 3D structural comparisons with the Protein Data Bank (PDB) reveal that black swan lysozyme G (Lyz23), a member of GH23, displays the closest structural similarity to *Ct*Cel124$_{CD}$. The proteins have a root-mean-square deviation of 2.1 Å for 115 aligned residues, which display 21% sequence identity (7% identity when the complete catalytic modules were compared). The secondary structural elements of *Ct*Cel124, apart from α-H1 and α-H3, are conserved in Lyz23 (Fig. 3). The conformation and sequence of the critical loop connecting α-H4 and α-H5, which extends along one face of the substrate-binding cleft of *Ct*Cel124 (see below) is not conserved in Lyz23. Furthermore, an additional helix, present in Lyz23 but not in *Ct*Cel124, would make steric clashes with the Glc at the +3 subsite. It is evident, therefore, that significant differences in the topology and the residues in the substrate-binding cleft of *Ct*Cel124 and Lyz23 explains why these two enzymes display very different substrate specificities.

Active Site and Catalytic Mechanism of CtCel124. The active site of *Ct*Cel124 displays a high degree of structural similarity with Lyz23, which hydrolyzes glycosidic bonds through a single displacement (inverting) mechanism, and the *Escherichia coli* GH23 lytic transglycosylase Slt70, which cleaves glycosidic bonds through a water-independent substrate-assisted mechanism that

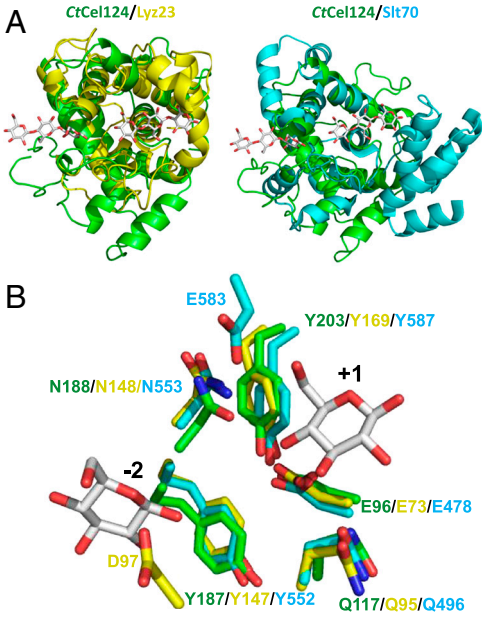

Fig. 3. Overlay of the structural fold and active site of *Ct*Cel124 and GH23 enzymes. (*A*) An overlay of *Ct*Cel124$_{CD}$ (green) and the GH23 black swan type G-lysozyme (yellow; PDB 1GBS). (*B*) An overlay of the structural fold of *Ct*Cel124$_{CD}$ (green) and the GH23 *E. coli* lytic transglycosylase Slt70 (cyan; PDB 1QTE). (*C*) An overlap of the three enzymes at the −1 active site. The silver-colored sugars are derived from the two cellotriose molecules.

ultimately leads to the formation of a 1,6-anhydro product (Fig. 3) (see Fig. S4 for details of the mechanism). Thus, Glu96, Gln117, Tyr187, Asn188, and Tyr203, which compose the central core of the active site of *Ct*Cel124, are structurally conserved in the two GH23 enzymes. It is evident, however, that the active sites of both Lyz23 and Slt70 contain an additional, but distinct, acidic residue, Asp97 and Glu583, respectively, which are lacking in the cellulase. The structural similarities and differences between the three enzymes present hypotheses regarding the catalytic mechanism of *Ct*Cel124. For example, it is possible that the cellulase cleaves glycosidic bonds through a lytic transglycosylase mechanism in which the 1,6-anhydro bond is hydrolyzed on enzyme to generate the reducing sugar products. This mechanism, however, is not supported by the observation that chemically synthesized 1,6-anhydro-cellotriose is not hydrolyzed by the cellulase, and that neither chitin nor chitosan are substrates for the enzyme (Table 1 and Fig. S5). Although the capacity of Slt70 to display a lytic activity is likely conferred, in part, by the catalytic acid-base residue Glu478 (equivalent to Glu96 in *Ct*Cel124), there is no obvious acidic residue in the cellulase or, indeed, in Slt70 capable of stabilizing the negative charge of the C2 N-acetyl group during substrate-assisted catalysis (Fig. S4). Although the structural similarity of *Ct*Cel124 and Lyz23 may indicate that the cellulase also acts through an inverting mechanism, the enzyme lacks a residue equivalent to Asp97 in Lyz23, the likely catalytic base (20), which is required to activate the water molecule that attacks C1 (Fig. S4). HPLC analysis of cellotriose, generated from cellopentaose by *Ct*Cel124, however, shows that the trisaccharide was primarily in its α configuration (Fig. S6), confirming that the cellulase cleaves glycosidic bonds by an acid-base single displacement mechanism leading to anomeric inversion. The lack of a candidate Asp or Glu catalytic base is evident in some inverting glycoside hydrolases. In GH6 enzymes, the identity of the Brønsted base has remained particularly elusive, and a "Grotthus"-style mechanism, in which a remote amino acid activates an active site water via a string of solvent molecules, remains the likely mechanism (21). It is likely, therefore, that *Ct*Cel124 also hydrolyzes glycosidic bonds through a Grotthus-style mechanism, although small nucleophilic ions such as thiols, acetate,

or phosphate ions may fulfill the role of the catalytic base by activating the nucleophilic water.

A central feature of the catalytic apparatus of *Ct*Cel124 is Glu96, which makes a hydrogen bond with *O4* of the Glc at +1 (Figs. 2 and 3). The glutamate is underneath the β face of the +1 Glc and is thus in an ideal orientation to promote leaving group departure by donating a proton to the scissile glycosidic *O*. This is consistent with the observation that mutation of Glu96 (E96A) completely inactivates the enzyme against polysaccharides and oligosaccharides where the leaving group is poor. Against 2,4-nitrophenyl-cellotriose (2,4 DNP-cellotriose), in which the 2,4-dinitrophenolate leaving group does not require protonation (pKa approximately 3.5) (22), the E96A mutation does not decrease activity (Table 1). This is again consistent with the view that Glu96 is the catalytic acid of *Ct*Cel124. The descending limb of the pH curve reports a single ionizing group, the catalytic acid, with a pKa of 6.8. The required modulation of the pKa of the Glu96 carboxylic acid (pKa of carboxylic acid groups in solution are approximately 4.0) is likely contributed in part through a polar interaction with the OH of Tyr203, although its apolar environment (Glu96 is in close proximity to Leu119, Tyr187, Tyr203, and the aliphatic chain of Gln117) will ensure that the carboxylic acid group of this residue is mostly protonated at pH 5.5. The equivalent residue in Lyz23 and Slt70, Glu73, and Glu478, respectively, are also likely to function as the catalytic acid during glycosidic bond cleavage.

To summarize, the active site of *Ct*Cel124 comprises a basic structural platform that is capable of mediating glycosidic bond cleavage of *gluco*-configured substrates, likely through a Grotthus-style mechanism. Adornment of this catalytic platform with Asp97 in Lys70 (Fig. 3) enables the enzyme to exhibit a classical single displacement (inverting) acid-base mechanism. However, it is unclear how the C2 N-acetyl group in Slt70 is activated, which is believed to be an essential feature of its lytic transglycosylase activity.

The Substrate-Binding Cleft of *Ct*Cel124. The superhelical fold provides a platform for the substrate-binding cleft that extends across the top of the protein (Fig. 2). The cleft houses the two molecules of cellotriose that bind to subsites −4 to −2 and +1 to +3, respectively. No ligand was evident in the −1 subsite. The extended loop, connecting α-H4 and α-H5, forms one wall of the binding cleft, and the other wall consists of several different structural elements that include the C-terminal end of α-H7, the N-terminal region of α-H8, and the N-terminal region of the loop connecting these helices (Fig. 2). Unlike the majority of glycanases, and carbohydrate-binding proteins in general, the substrate-binding cleft of *Ct*Cel124 does not contain a significant hydrophobic platform; the cellulase makes numerous direct polar contacts with the two cellotriose molecules (Fig. 2).

A striking feature of the substrate-binding cleft is the different topologies displayed by its positive and negative subsites, respectively (Fig. 2). Subsites −4 to −1 form a deep narrow cleft in which the bound trisaccharide is significantly twisted (Fig. S7). In contrast, subsites +1 to +3 display a more open topology (Fig. 2), and the conformation of the bound trisaccharide adopts an approximate twofold screw axis (Fig. S7). It is not possible to obtain the crystal structure of an enzyme in complex with insoluble polysaccharides such as cellulose. However, the conformation adopted by cellooligosaccharides such as cellotriose, on enzyme, provides insight into the likely topological features of cellulose bound to CtCel124$_{CD}$. Thus, the helical structure of cellotriose bound to the negative subsites is distinct from the twofold screw axis displayed by glucan chains in crystalline cellulose. Indeed, the twisted structure of cellotriose is adopted by cellooligosaccharides in solution (23). By contrast, the linear conformation adopted by cellotriose bound to the distal positive subsites is similar to the structure of the glucan chains in crystalline cellulose (2). Thus, it is likely that the substrate-binding cleft of *Ct*Cel124 is tailored to recognize specific substructures of cellulose, which are at the interface between crystalline and paracrystalline (or amorphous) regions of cellulose. Indeed, competition experiments between cellulose-specific CBMs indicate that these proteins recognize specific topological features of the polysaccharide. Thus, CBMs belonging to families 4, 17, and 28 recognize distinct amorphous or paracrystalline regions of cellulose (24, 25), and CBM2a, CBM3a, and CBM1, which bind to crystalline cellulose, also display distinct specificities (26). Many cellulose-degrading bacteria express a large number of *endo*-β-1,4-glucanases (27, 28), exemplified by *C. thermocellum* that has the potential to synthesize approximately 30 endoglucanases. The biological rationale for the expansion in this enzyme activity in *C. thermocellum* and other organisms is currently unclear. Cellulose, although chemically invariant, displays very different topologies ranging from highly crystalline structures to isolated highly twisted glucan chains (amorphous cellulose). It is possible that at least some of the endoglucanases expressed by a single organism are tailored to recognize specific cellulose substructures found in nature. *Ct*Cel124, by targeting the boundary between crystalline and amorphous regions of cellulose, may generate reaction products that comprise substrates for *exo*-acting cellulases that act on the nonreducing end of crystalline cellulose and the reducing end of isolated cellulose chains, respectively.

Materials and Methods

Cloning, Expression, and Purification of *Ct*Cel124$_{CD}$. DNA encoding the C-terminal module of *Ct*Cel124 (designated *Ct*Cel124$_{CD}$; residues 131–350) was amplified from *C. thermocellum* strain ATCC 2745 genomic DNA by PCR, using primers listed in Table S2, and the resultant DNA was cloned into pET28a to generate pCel124. *Ct*Cel124$_{CD}$ encoded by pCel124 contains a His$_6$-tag. To generate *Ct*Cel124$_{CD}$ fused to the noncatalytic carbohydrate (cellulose)-binding module CBM3a (*Ct*Cel124$_{CD}$-CBM3a), overlapping PCR was used deploying primers that amplified the DNA sequences encoding the two modules. Expression of the proteins was achieved by adding isopropyl β-D-thiogalactopyranoside (1 mM final concentration) to midexponential phase cultures of *E. coli* BL21(DE3) harboring pCel124 with incubation for a further 16 h at 37 °C. The His$_6$-tagged recombinant protein was purified from cell-free extracts by immobilized metal ion affinity chromatography using standard methodology (29). To produce seleno-L-methionine *Ct*Cel124$_{CD}$, the *E. coli* methionine auxotroph B834(DE3) containing pCel124 was cultured as described by Charnock et al. (30). The recombinant protein was purified as described above except all buffers were supplemented with 5 mm β-mercaptoethanol. The production of recombinant Cel48S and the CBM3a-Coh construct (CBM3a fused to the type I cohesin3 from CipA, the primary scaffoldin of *C. thermocellum*) were described previously (9, 31, 32).

Enzyme Assays. In enzyme assays using soluble polysaccharides, the reactions were carried out in 50 mM MES buffer, pH 5.5, containing 1 mg/mL BSA and 0.5% of the target polysaccharide. Reactions, which were incubated at 50 °C, contained, typically, 15 nM of enzyme. At regular time points, aliquots were removed and reducing sugar present was determined (33), using glucose to construct the standard curve. In enzyme assays using Avicel (Sigma Chem. Co.) and PASC [prepared as described previously (15)] as substrates, the reaction mixture consisted of 500 nM enzyme in 100 mM acetate buffer, pH 5.0, 24 mM $CaCl_2$, 4 mM EDTA, and insoluble cellulose at a concentration of 2% (wt/vol). The reactions, which were carried out at 50 °C, were terminated by immersing the sample tubes in ice water. After centrifugation to remove insoluble substrate, the supernatant was assayed for reducing sugar. When assaying Cel48S in the presence of CBM3a-Coh, equimolar quantities of the protein partners were incubated for 2 h at 37 °C (without the substrate) preceding the assay. This ensured that Cel48S was appended to the CBM3a. The hydrolysis of cellooligosaccharides (30 μM), conducted in 50 mM MES buffer, was assessed by the rate of substrate depletion using a previously described Dionex–high-performance anion-exchange chromatography (HPAEC) method to detect cellooligosaccharides (15). HPAEC was also used to identify the reaction products released from PASC. The α and β configuration of the anomeric carbon of cellotriose, generated by the hydrolysis of cellopentaose by *Ct*Cel124$_{CD}$, was determined as described previously (15, 34).

Binding of the *Ct*Cel124$_{CD}$ Mutant E96A to Ligands. Binding was determined by ITC and by depletion binding isotherms. ITC measurements were made

BIOCHEMISTRY

at 40 °C following standard procedures (30) using a Microcal iTC$_{200}$ calorimeter in 50 mM MES buffer, pH 5.5. During a titration experiment, the protein sample at 88 μM was injected with 20 × 2 μL aliquots of 2 mM cellohexaose. For experiments with RC, prepared as described previously (13), the ligand was in the cell at 29.4 mg/mL, and the protein (835 μM) was the titrant. Integrated heat effects, after correction for heats of dilution, were analyzed by nonlinear regression using a single site-binding model (Microcal Origin, version 2.9). Thermodynamic parameters were calculated using the standard thermodynamic equation: $-RT\ln K_a = \Delta G = \Delta H - T\Delta S$. Depletion binding isotherms were performed in 50 mM MES buffer, pH 7.0, at protein concentrations ranging from 1 to 30 μM. Protein was added to 1 mg of RC in a final volume of 1 mL and incubated for 1 h with gentle mixing at 40 °C. The polysaccharide was centrifuged at 13,000 × *g* for 1 min, and the A_{280} of the supernatant was measured to quantify the amount of free protein remaining after binding. Bound protein was calculated from the total minus the free protein. The data were analyzed by nonlinear regression using a standard one-site binding model (GraphPad Prism, v5), and the N_0 and K_a values were obtained from the regressed isotherm data.

Crystallization of *Ct*Cel124$_{CD}$ and Data Collection. Structure determination was by the single-wavelength anomalous dispersion (SAD) method, using the anomalous diffraction of the selenium atoms, incorporated in the protein. *Ct*Cel124$_{CD}$ was crystallized with an equal volume (1 μL) of protein (60 mg/mL in solution with 10 mM cellohexaose) and reservoir solution [8% (vol/vol) tacsimate pH 5.0 and 20% (wt/vol) PEG 3350). Glycerol (30% wt/vol) was added as the cryoprotectant. The SAD experiment was conducted on beamline ID14-EH4 at the European Synchrotron Radiation Facility at Grenoble, France, using an ADSC Quantum-4 CCD detector. Data were collected at 0.954 Å wavelength. A total of 120 images with 1° oscillation for 5 s were collected. The data were processed with DENZO and the HKL2000 package and scaled with SCALA (35).

Phasing, Model Building, and Refinement. Patterson maps were deconvoluted by calculation of anomalous Patterson maps. The positions of the seleniums were refined, and phases were calculated using SHARP/autoSHARP (36). Subsequently, a cycle of phase improvement was applied using the program DM (37), where phases were modified and extended to 1.5 Å resolution yielding a figure of merit of 0.91 and an interpretable electron-density map. A model comprising 243 amino acids was built from the initial map with program COOT (38). Water molecules and alternative conformers were added using ARP/WARP (35), and refinement with REFMAC5 (39) was performed as deemed appropriate from the behavior of the cross-validation (R_{free}) subset of reflections (10%). Solvent molecules were added in the final stages of refinement according to hydrogen bond criteria and only if their B factors refined to reasonable values and if they improved the R_{free}. The statistics for structure refinement are shown in Table S3.

ACKNOWLEDGMENTS. The authors acknowledge financial support from Fundação para a Ciência e a Tecnologia, Portugal, through grants PTDC/BIA-PRO/69732/2006, PTDC/QUI-BIQ/100359/2008, and the individual fellowship SFRH/BD/38667/2007 (to J.L.A.B.); the U.K. Biological and Biotechnological Sciences Research Council (studentship to A.C.); the US Department of Energy (Grant 00562492); and the Israel Science Foundation (Grant 966/09). The authors also acknowledge the European Synchrotron Radiation Facility, Grenoble, France (beamline ID14-EH4) for access and technical support during data collection.

1. Himmel ME, Bayer EA (2009) Lignocellulose conversion to biofuels: Current challenges, global perspectives. *Curr Opin Biotechnol* 20:316–317.
2. Nishiyama Y, Johnson GP, French AD, Forsyth VT, Langan P (2008) Neutron crystallography, molecular dynamics, and quantum mechanics studies of the nature of hydrogen bonding in cellulose I beta. *Biomacromolecules* 9:3133–3140.
3. Gilbert HJ (2010) The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol* 153:444–455.
4. Tolonen AC, Chilaka AC, Church GM (2009) Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Mol Microbiol* 74:1300–1313.
5. Cantarel BL, et al. (2009) The Carbohydrate-Active enZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Res* 37:D233–238.
6. Bayer EA, Belaich JP, Shoham Y, Lamed R (2004) The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol* 58:521–554.
7. Fontes CM, Gilbert HJ (2010) Cellulosomes: Highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu Rev Biochem* 79:655–681.
8. Zverlov VV, Schantz N, Schwarz WH (2005) A major new component in the cellulosome of *Clostridium thermocellum* is a processive endo-beta-1,4-glucanase producing cellotetraose. *FEMS Microbiol Lett* 249:353–358.
9. Fierobe HP, et al. (2002) Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. *J Biol Chem* 277: 49621–49630.
10. Zhang YH, Lynd LR (2005) Cellulose utilization by *Clostridium thermocellum*: Bioenergetics and hydrolysis product assimilation. *Proc Natl Acad Sci USA* 102:7321–7325.
11. Raman B, et al. (2009) Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: A quantitative proteomic analysis. *PLoS One* 4:e5271.
12. Hazlewood GP, Davidson K, Laurie JI, Romaniec MP, Gilbert HJ (1990) Cloning and sequencing of the celA gene encoding endoglucanase A of *Butyrivibrio fibrisolvens* strain A46. *J Gen Microbiol* 136:2089–2097.
13. Boraston AB, et al. (2001) Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. *Biochemistry* 40:6240–6247.
14. Davies GJ, Wilson KS, Henrissat B (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* 321:557–559.
15. Hall J, et al. (1995) The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas fluorescens* subsp. *cellulosa* is important for the efficient hydrolysis of Avicel. *Biochem J* 309:749–756.
16. Zhang XZ, Sathitsuksanoh N, Zhang YH (2010) Glycoside hydrolase family 9 processive endoglucanase from *Clostridium phytofermentans*: Heterologous expression, characterization, and synergy with family 48 cellobiohydrolase. *Bioresource Technol* 101:5534–5538.
17. Olson DG, et al. (2010) Deletion of the Cel48S cellulase from *Clostridium thermocellum*. *Proc Natl Acad Sci USA* 107:17727–17732.
18. Pinheiro BA, et al. (2009) Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*. *Biochem J* 424:375–384.
19. Lu YP, Zhang YHP, Lynd LR (2006) Enzyme-microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. *Proc Natl Acad Sci USA* 103:16165–16169.
20. Karlsen S, Hough E, Rao ZH, Isaacs NW (1996) Structure of a bulgecin-inhibited g-type lysozyme from the egg white of the Australian black swan. A comparison of the binding of bulgecin to three muramidases. *Acta Crystallogr* D52:105–114.
21. Koivula A, et al. (2002) The active site of cellobiohydrolase Cel6A from *Trichoderma reesei*: The roles of aspartic acids D221 and D175. *J Am Chem Soc* 124:10015–10024.
22. Damude HG, Withers SG, Kilburn DG, Miller RC, Warren RAJ (1995) Site-directed mutation of the putative catalytic residues of endoglucanase CenA from *Cellulomonas fimi*. *Biochemistry* 34:2220–2224.
23. Sugiyama H, Hisamichi K, Usui T, Sakai K, Ishiyama J (2000) A study of the conformation of beta-1,4-linked glucose oligomers, cellobiose to cellohexaose, in solution. *J Mol Struct* 556:173–177.
24. Boraston AB, Kwan E, Chiu P, Warren RA, Kilburn DG (2003) Recognition and hydrolysis of noncrystalline cellulose. *J Biol Chem* 278:6120–6127.
25. McLean BW, et al. (2002) Carbohydrate-binding modules recognize fine substructures of cellulose. *J Biol Chem* 277:50245–50254.
26. Blake AW, et al. (2006) Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *J Biol Chem* 281:29321–29329.
27. Lykidis A, et al. (2007) Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX. *J Bacteriol* 189:2477–2486.
28. Weiner RM, et al. (2008) Complete genome sequence of the complex carbohydrate-degrading marine bacterium, *Saccharophagus degradans* strain 2–40 T. *PLoS Genet* 4:e1000087.
29. Pell G, et al. (2004) Structural and biochemical analysis of *Cellvibrio japonicus* xylanase 10C: How variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases. *J Biol Chem* 279:11777–11788.
30. Charnock SJ, et al. (2000) The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: Structure and biochemistry of the *Clostridium thermocellum* X6b domain. *Biochemistry* 39:5013–5021.
31. Barak Y, et al. (2005) Matching fusion protein systems for affinity analysis of two interacting families of proteins: The cohesin-dockerin interaction. *J Mol Recognit* 18:491–501.
32. Haimovitz R, et al. (2008) Cohesin-dockerin microarray: Diverse specificities between two complementary families of interacting protein modules. *Proteomics* 8:968–979.
33. Miller GL (1959) The use of dinitrosalicylic acid reagent for the determination of reducing sugar. *Anal Chem* 31:426–428.
34. Braun C, Meinke A, Ziser L, Withers SG (1993) Simultaneous high-performance liquid chromatographic determination of both the cleavage pattern and the stereochemical outcome of the hydrolysis reactions catalyzed by various glycosidases. *Anal Biochem* 212:259–262.
35. Collaborative Computational Project (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr* D50:760–763.
36. La Fortelle E, Bricogne G (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* 276:472–494.
37. Cowtan KD, Main P (1993) Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. *Acta Crystallogr* D49:148–157.
38. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr* D60:2126–2132.
39. Winn MD, Isupov MN, Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr* D57:122–133.