Crystal Structure of the Catalytic Domain of the β-1,4-Glycanase Cex from *Cellulomonas fimi*[†,‡]

André White,[§] Stephen G. Withers,[∥] Neil R. Gilkes,[⊥] and David R. Rose*[,§]

Protein Engineering Network of Centres of Excellence, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 500 Sherbourne Street, Toronto, Canada M4X 1K9, and Departments of Chemistry and Microbiology, University of British Columbia, Vancouver, Canada V6T 1Z1

Received July 8, 1994; Revised Manuscript Received August 25, 1994[⊛]

ABSTRACT: β-1,4-Glycanases, principally cellulases and xylanases, are responsible for the hydrolysis of plant biomass. The bifunctional β-1,4-xylanase/glucanase Cex from the bacterium *Cellulomonas fimi*, one of a large family of cellulases/xylanases, depolymerizes oligosaccharides and releases a disaccharide unit from the substrate nonreducing end. Hydrolysis occurs with net retention of the anomeric configuration of the sugar through a double-displacement mechanism involving a covalent glycosyl-enzyme intermediate. The active site nucleophile, Glu233, has been unambiguously identified by trapping of such an intermediate [Tull et al. (1991) *J. Biol. Chem. 266*, 15621–15625] and the acid/base catalyst, Glu127, by detailed kinetic analysis of mutants [MacLeod et al. (1994) *Biochemistry 33*, 6371–6376]. However, little is known about the enzyme's overall folding and its active site architecture. We report here the high-resolution crystal structure of the catalytic domain of Cex. The atomic structure refinement results in a model that includes 2400 protein atoms and 45 water molecules, with an *R*-factor of 0.217 for data extending to 1.8-Å resolution. The protein forms an eight-parallel-stranded α/β-barrel, which is a novel folding pattern for a microbial β-glycanase. The active site, inferred from the location of Glu233, Glu 127, and other conserved residues, is an open cleft on the carboxy-terminal end of the α/β-barrel. An extensive hydrogen-bonding network stabilizes the ionization states of the key residues; in particular, the Asp235–His205–Glu233 hydrogen-bonding network may play a role in modulating the ionization state of Glu233 and in controlling local charge balance during the reaction.

Cellulases (β-1,4-glucanases) and β-1,4-xylanases are examples of a broad range of glycoside hydrolyzing enzymes produced by microorganisms that make use of plant biomass as a source of carbon and energy. They are usually modular enzymes composed of two or more discrete structural and functional units. Typically, in enzymes derived from both bacteria and fungi, a catalytic domain is joined to a substrate-binding domain by a short linker sequence; additional domains are sometimes involved (Gilkes et al., 1991b). The catalytic domains of over 150 known β-1,4-glucanases and β-1,4-xylanases can be classified into a few distinct families on the basis of amino acid sequence similarities (Henrissat & Bairoch, 1993). Comparative analysis suggests that, despite their apparent diversity, all these enzymes arose by mutation and shuffling of relatively few progenitor modules (Henrissat & Bairoch, 1993).

Cellulomonas fimi is a cellulolytic bacterium that synthesizes at least six β-1,4-glucanases and four β-1,4-xylanases (Meinke et al., 1994; Shen et al., 1994). One of these, Cex, also known as XynB (Hazlewood & Gilbert, 1993), is a 47.1-kDa enzyme comprised of an N-terminal catalytic domain (cex-cd) and a C-terminal cellulose-binding domain (CBD_{Cex}) joined by a linker polypeptide containing 20 alternating proline and threonine residues (O'Neill et al., 1986). Small-angle X-ray scattering analysis indicates that the structure of Cex resembles a tadpole in which the catalytic domain corresponds to the globular head and the cellulose-binding domain to the extended tail (N. R. Gilkes et al., unpublished data). Cex was first recognized as a β-1,4-glucanase, and analysis of (carboxymethyl)cellulose hydrolysis showed that its action was predominantly exohydrolytic (Gilkes et al., 1984). Subsequent analysis revealed that its catalytic efficiency on *p*-nitrophenyl xylobioside was about 50-fold higher than on the cellobioside analog (Gilkes et al., 1991a). This preference is a common feature of all the enzymes currently known in family 10 (formerly family F) of glycosyl hydrolases (Gilkes et al., 1991b; Henrissat & Bairoch, 1993). The specificity of cex-cd for both β-1,4-xylosidic and glucosidic bonds is of interest not only mechanistically but also because of the importance of xylanases in various biotechnological processes. Some applications, such as enzymatic prebleaching of kraft pulps for paper manufacture, require xylanase enzymes that are free of β-1,4-glucanase activity, while others would benefit from enzymes with dual specificities (Wong & Saddler, 1992). In addition to providing an atomic basis for its catalytic mechanism, the determination of the structure of cex-cd should yield clues as to how the specificity for each substrate is controlled; this would ultimately allow the engineering of such enzymes to change their substrate preference.

Kinetic studies on Cex have provided good evidence for a double-displacement catalytic mechanism in which a glycosyl-enzyme intermediate is formed *via* the attack of a nucleophilic side chain and is then hydrolyzed to release the product with the same anomeric configuration as the substrate. Both the

[†] This work was supported by the Protein Engineering Network of Centres of Excellence, one of twelve Network of Centres of Excellence funded by the Government of Canada, and by a National Cancer Institute of Canada grant to D.R.R.

[‡] Final model atomic coordinates have been deposited with the Brookhaven Protein Data Bank (Accession Number 1EXO).

* Corresponding author: telephone (416) 924-0671 ext. 5446; fax (416) 926-6529; email drose@oci.utoronto.ca.

[§] University of Toronto.

[∥] Department of Chemistry, University of British Columbia.

[⊥] Department of Microbiology, University of British Columbia.

[⊛] Abstract published in *Advance ACS Abstracts,* October 1, 1994.

0006-2960/94/0433-12546$04.50/0

formation of the intermediate (glycosylation) and its hydrolysis (deglycosylation) require acid/base catalytic assistance, and both proceed *via* oxocarbenium ion-like transition states (Withers et al., 1986; Tull & Withers, 1994). The nucleophilic residue was identified as Glu233 by trapping of the glycosyl-enzyme intermediate using a 2-deoxy-2-fluoroglycoside substrate, followed by proteolysis of the complex and sequencing of the labeled peptide (Tull et al., 1991). The intermediate that was stably trapped in these experiments was shown to be catalytically competent, as the addition of sugar acceptors such as cellobiose resulted in turnover *via* transglycosylation, the reverse reaction of the enzyme (Tull et al., 1991). Site-directed mutagenesis at Glu233 has yielded mutants with kinetic properties fully consistent with this role (MacLeod, 1994). Mutations at Glu127, followed by detailed kinetic analysis, confirmed its role as an acid/base catalyst (MacLeod et al., 1994). Determination of the structure of the Cex catalytic domain at atomic resolution will provide an overall picture of the folding of the enzyme and details of the atomic structure of the active site, including the positions of Glu233, Glu127, and other residues involved in substrate binding and enzyme catalysis. This will provide a molecular basis for future mutagenesis studies and will add to our understanding of the mechanism of this family of retaining β-glucanases, of which no structures have yet been published.

Three-dimensional structures of known bacterial and fungal β-glucanases and xylanases show a remarkable variety of different folding motifs. The *Trichoderma reseii* CbhII exoglucanase forms a parallel seven-stranded α/β-barrel (Rouvinen et al., 1990), which is also seen in the thermophilic endocellulase E2 catalytic domain from *Thermomonospora fusca* (Spezio et al., 1993); the *Humicola insolens* endoglucanase V folds into a six stranded β-barrel (Davies et al., 1993); the endoglucanase CelD from *Clostridium thermocellum* forms an $(\alpha/\alpha)_6$-barrel (Juy et al., 1992); the *Bacillus subtilis* and *Trichoderma harzianum* 20-kDa xylanases fold into a sandwich of two β-pleated sheets (Campbell et al., 1993; D. R. Rose and R. Campbell, in preparation), a folding similar to that to the hybrid *Bacillus* 1,3–1,4-β-glucanase (Keitel et al., 1993). To date, however, there is no reported three-dimensional structure of a family 10 β-1,4-glycanase, though two other crystallizations have been published (Pikersgill et al., 1993; Souchon et al., 1994). Here we report the high resolution crystal structure of the catalytic domain of the β-1,4-glycanase Cex.

MATERIALS AND METHODS

Crystallization and Data Collection. The cex-cd was obtained by proteolysis of recombinant Cex protein and then crystallized as previously described (Bedarkar et al., 1992). Two useful heavy atom derivatives were obtained by soaking protein crystals in 15% (w/v) PEG 4000, 100 mM sodium acetate, pH 4.6, and either 20 mM trimethyllead acetate (TMLA) for 30 h or 0.5 mM $(COOCH_3)_2Hg$ for 3 h followed by a 21-h back-soak. Each data set was measured from a single crystal at room temperature on a Xuong-Hamlin area detector connected to a Rigaku RU-200 rotating anode X-ray generator operated at 6 kW. Oscillation frames of 0.1° on ω were measured, from which data to 1.8-Å resolution were obtained and then reduced using the software by Howard and Neilsen (Howard et al., 1985). Bijvoet pairs for both derivative data sets were measured in order to obtain anomalous scattering data.

Isomorphous Replacements. The positions of the heavy atoms were determined from difference Patterson maps and subsequently refined using the software package PHASES (Furey & Swaminathan, 1990); the single site for the lead derivative was solved first and then used along with its anomalous scattering signal to calculate phases, in order to locate the positions of the mercury atoms by difference Fourier synthesis. The crystallographic tetragonal space group ambiguity was resolved by observing the lowest residual obtained for the combined refinement of both heavy atoms, in either $P4_12_12$ or $P4_32_12$. The crystals belong to the space group $P4_12_12$. The heavy atom positions of both derivatives were refined together and their phases combined at 2.3-Å resolution using the package PHASES. An automated solvent flattening procedure (Wang, 1985) as implemented in the program PHASES was carried out to improve the combined protein phases. The heavy atom positions were then refined again, this time against the phases obtained following the solvent flattening algorithm. Protein phases to 2.3-Å resolution were determined and used to generate an electron density map.

Structural Analysis and Model Building. A primary tracing of the protein was obtained using the program BONES, as implemented in the program MAPMAN (Kleywegt & Jones, 1994). This trace was edited and used as a guide to build a poly-Cα into the electron density map. All the model building and graphical rebuilding were carried out on a Silicon Graphics Indigo, using the program O version 5.9.1 (Jones et al., 1990). An atomic model of residues 3–312 of cex-cd was then interpreted from the electron density map, using the poly-Cα as a guide and assisted by a library of substructures (Jones & Thirup, 1986). Residues 1 and 2 were located at a later stage.

Structure Refinement. All the protein structure refinements were performed using the program X-PLOR version 3.1 (Brünger et al., 1987), in which 10% of the diffraction data were set aside to compute an R_{free} so as to follow the refinement progress objectively (Brünger, 1992). The initial protein atomic model was refined by simulated annealing and Powell minimization against all 15–2.3 Å data with $F > 2\sigma$ as follows. Prior to 120 cycles of restrained energy refinement, each atom was given an arbitrary constant temperature factor of 20 Å.^{2} This was followed by molecular dynamics in which the structure was slowly cooled from 3000 to 300 K in steps of 25 K. An additional positional refinement of 120 cycles was applied, followed by temperature factor refinement of first an overall and then individual atomic B values.

An electron density map was computed using the $2mF_o - DF_c$ coefficients and phases, combined from the model and the experimental MIR, as implemented in the program SIGMAA (Read, 1986). Incorrectly interpreted regions were located by a "residue real space correlation" (Jones et al., 1991) and corrected with the use of omit maps ($F_o - F_c$, ϕ_c computed with the omission of less than 10% of the residues). A model rebuilding was done using this electron density map, followed by 120 cycles of positional refinement at 300 K.

Peaks above 3σ in a $F_o - F_c$ map were located using the command PICKPEAK, as implemented in the program MAPMAN (Kleywegt & Jones, 1994). Each water molecule was visually examined and had to meet the following criteria: acceptable hydrogen bond geometry; within hydrogen bond distance of 2.5–3.5 Å to at least two protein atoms; electron density peak $\geq 3\sigma$ in a $F_o - F_c$ map; temperature factor $B \leq 60$ Å.^{2}

Data to 1.8-Å resolution were then included in rounds of positional refinement and model rebuilding. The stereochemistry of the model was verified using the software package

Table 1: Diffraction Data and MIRAS Phasing

Data Collection

	native	TMLA	$(CH_3COO)_2Hg$		native	TMLA	$(CH_3COO)_2Hg$
a, *b* (Å)	88.15	88.42	88.21	no. of unique reflections	29059	31654	24920
c (Å)	81.11	81.27	81.16	data completeness (%)	94	93	89
resolution (Å)	1.8	2.2	2.3	R_{sym} (%)[a]	6.1	6.0	4.2
no. of observations	128888	151941	200753				

MIRAS Phasing

						phasing power[d]	
heavy atom	concn (mM)	soaking (h)	no. of sites	R_{der}[b] (%)	R_{Cullis}[c] (%)	isomorphous	anomalous
TMLA	15	30	1	8.6	60	1.70	1.42
$(CH_3COO)_2Hg$	0.5	3	3	5.0	61	1.89	1.53

Heavy Atom Locations

type	site	*x* (Å)	*y* (Å)	*z* (Å)	*q*	*B* (Å^2)	close protein atoms
Pb		50.70	18.98	6.93	1.02	20.0[e]	E76 $O^{\epsilon1,\epsilon2}$; D14 $O^{\delta1,\delta2}$
Hg	1	11.12	75.44	7.95	1.02	39.22	H85 N^{ϵ}; D138 $O^{\delta2}$, O
Hg	2	18.91	67.56	1.23	1.00	13.45	H80 N^{ϵ}; E233 $O^{\epsilon1,\epsilon2}$
Hg	3	21.27	67.26	3.63	0.95	14.44	H205 N^{ϵ}; E233 $O^{\epsilon1}$

[a] $R_{sym} = \sum|I_i - \langle I\rangle|/\sum\langle I\rangle$, where I_i is the scaled intensity of the *i*th measurement and $\langle I\rangle$ is the mean intensity for that reflection. [b] $R_{der} = \sum|F_{PH} - F_P|/\sum(F_P + F_{PH})$, where F_P and F_{PH} are the native and derivative structure factor amplitudes, respectively. [c] $R_{Cullis} = \sum\|F_{PH} \pm F_P| - F_H|/\sum|F_{PH} \pm F_P|$ for centric reflections. [d] Phasing power is $\langle F_h\rangle/\epsilon$, where $\langle F_h\rangle$ is the mean amplitude of heavy atom structure factor and ϵ is the rms lack of closure error. [e] The thermal factor *B* of the Pb derivative was not refined but the occupancy was.

Table 2: Crystallographic Refinement Statistics

no. of non-hydrogen atoms	2400	(residues 1–312)		
no. of water molecules	45			
rms deviation, bond length (Å)	0.011			
rms deviation, bond angle (deg)	1.86			
rms deviation, improper angles (deg)	1.63			
resolution (Å)	8.0–2.83	2.83–2.26	2.26–1.98	1.98–1.80
no. of reflections ($F > 2\sigma$)	6664	6028	5491	4115
R-factor (shell)[a]	0.168	0.236	0.284	0.331
R-factor (cumulative)	0.168	0.188	0.205	0.217
reflections set aside from the atomic refinement				
no. of reflections ($F > 2\sigma$)	755	691	566	453
R_{free} (shell)	0.253	0.316	0.307	0.353
R_{free} (cumulative)	0.253	0.271	0.277	0.284

[a] *R*-factor = $\sum\|F_o| - |F_c\|/\sum|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively.

PROCHECK (Laskowski et al., 1993). The secondary structures are assigned by the criteria of Kabsch and Sander (1983), as implemented in PROCHECK. The final model atomic coordinates have been deposited with Brookhaven Protein Data Bank.

RESULTS AND DISCUSSION

Multiple Isomorphous Replacement. A single lead site was located from the Harker section ($^1/_2$,*v*,*w*) of the difference Patterson of the TMLA derivative. The heavy atom position was refined for both isomorphous and anomalous scattering data and results in a combined figure of merit of 0.67 at 3.0-Å resolution. An automated solvent flattening filtering improved the figure of merit to 0.85 for the protein phases; a solvent content of 40% was used, compared to 46% as estimated from the unit cell. These phases were then used to locate the mercury heavy atom sites by difference Fourier synthesis, which identified three major sites. A combined positional refinement of the lead and mercury heavy atoms yielded an overall combined figure of merit of 0.71 (Table 1), which was further improved by solvent flattening. The new phases were then used to re-refine the position of the heavy atoms, followed by solvent flattening. The final figure of merit is 0.88 for the protein phases within the resolution range 15–2.3 Å.

Refinement and Quality Assessment of the Current cex-cd Atomic Model. The refinement by simulated annealing of 2400 non-hydrogen atoms, *i.e.*, 312 residues of the native cex-cd, improved the *R*-factor from 38.4% to 24.2% at 2.3-Å resolution. The resolution was increased to 1.8 Å, and 45 water molecules were added to the model, followed by rounds of model building and refinement by X-PLOR. The current value of the *R*-factor is 21.7% and the R_{free} is 28.3%, for the resolution range 8–1.8 Å and $F > 2\sigma$ (Table 2). For the current model, the root-mean-square deviations from ideality are 0.011 Å for bond distances, 1.86° for bond angles, and 1.63° for "improper" angles. The residue real space correlation for the model has an average value per residue of 0.91 (σ = 0.02), with the lowest values between 0.80 and 0.84 for Glu76, Asn94, and Asn96. Among the non-glycine residues, 252 (91%) have their ϕ, ψ angles in the "most favored" regions of the Ramachandran plot, whereas the remaining 24 (9%) are in the "allowed" regions.

Structural Features. The catalytic domain of Cex is a globular protein rich in secondary structure. There are three single-turn α-helices and ten longer ones. Alternating with the α-helices are ten β-strands, of which eight are parallel and form the elliptical central core of an (α/β)-barrel (Figure 1). The secondary structure elements are α0-β1-α1-β2-α2a-α2b-β3-α3a-α3b-β4-α4a-α4b-β5-α5-β6a-β6b-α6-β7-α7-β8-α8-β9-α9. The amino- and carboxy-terminal α-helices, α0 and α9, are adjacent and parallel to each other. Disulfide bridges were predicted to join Cys167 with Cys199 (CC1) and Cys261

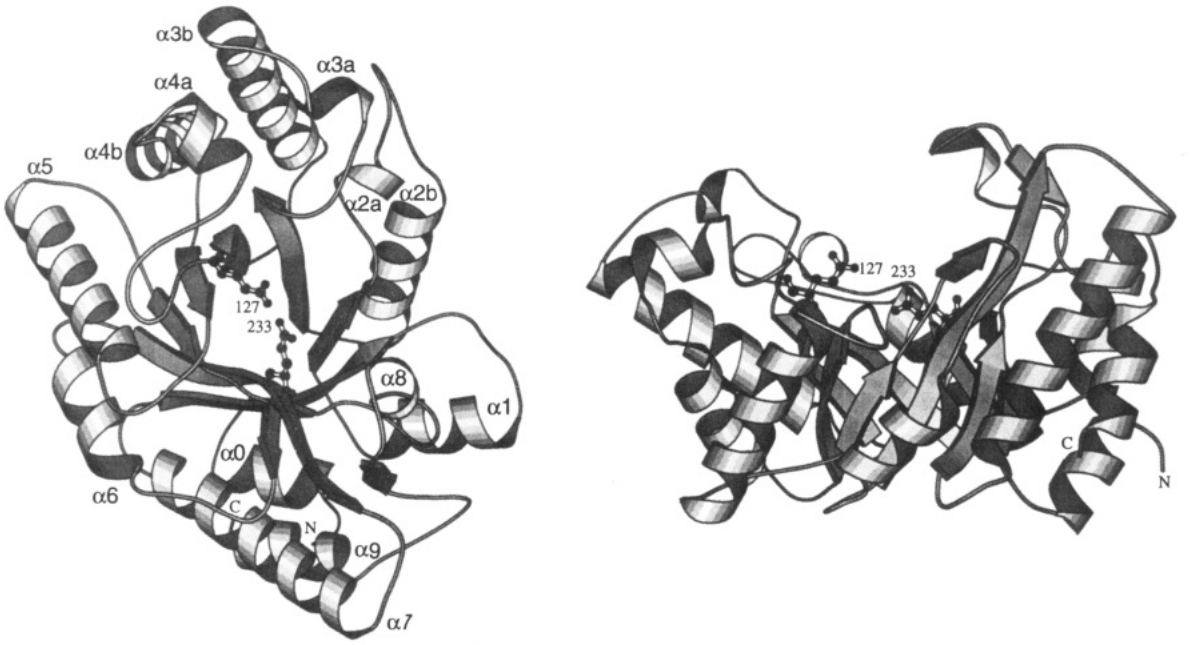

FIGURE 1: Folding of cex-cd, which forms an α/β-barrel, as drawn by the program MOLSCRIPT (Kraulis, 1991). Residue numbers refer to the acid/base catalyst Glu127 and the nucleophile Glu233, whereas N and C identify the amino and carboxy termini. The secondary structure locations are α0 (4–10), β1 (14–19), α1 (27–36), β2 (39–42), α2a (48–51), α2b (61–73), β3 (76–83), α3a (90–93), α3b (97–114), β4 (121–126), α4a (140–145), α4b (149–160), β5 (165–170), α5 (178–193), β6a (199–202), β6b (205–207), α6 (215–223), β7 (228–239), α7 (244–262), β8 (267–272), α8 (282–285), β9 (289–290), and α9 (304–311).

with Cys267 (CC2) (Gilkes et al., 1991a). Both disulfide bridges are now confirmed; CC1 joins the strands β5 and β6a, while CC2 constrains residues 264–266 to form a turn connecting α7 to β8.

There are two *cis*-peptides in the three-dimensional structure of cex-cd. The peptide bond Thr240–Pro241 is located in a turn between strand β7 and helix α7 and adopts a *cis* conformation; the *cis*-Pro241 is stabilized by Lys246 N$^{\zeta}$, which sits within hydrogen bond distance of the backbone oxygens of Arg239 and Pro241. The interpretation of this region is consistent with an omit map, in which residues 239–245 were omitted. Although a *cis*-peptide is energetically less favorable for residues other than proline, the peptide bond His80–Thr81 adopts the *cis* conformation; this electron density map interpretation was confirmed by an omit map in which residue 81 and its neighboring atoms within 6.5-Å distance were left out while the remainder of the model was subjected to simulated annealing, as proposed by Brünger (Hodel et al., 1992). Because the carbonyl oxygen of His80 would otherwise clash with the carbonyl of Asp123, it is strained and its angles, ψ and ω (peptide bond rotation angle), are rotated by about 180°, thus producing a *cis*-peptide. The His80 backbone O position is now stabilized by two water molecules in favorable hydrogen bond positions. The backbone of residues 80–82 and 123 form a "special β-bulge" of type "SP3" (Chan et al., 1993), a conformation which incorporates the unusual orientation of the backbone of *cis*-Thr81.

Folding. The catalytic domain of Cex folds into an α/β-barrel that contains an elliptical core of eight parallel β-strands (Figure 1). First observed in triose-phosphate isomerase (TIM) (Banner et al., 1975), the α/β-barrel is a common folding pattern for enzymes with a wide variety of different activities; these enzymes have been postulated to be related by divergent evolution (Farber & Petsko, 1990). The folding of the family 10 enzyme cex-cd is distinct from the reported crystal structures of enzymes of other bacterial and fungal families of β-glucanases and xylanases. The cex-cd folding differs from family 6 (previously family B) enzymes CbhII and $E2_{cd}$ (Rouvinen et al., 1990; Spezio et al., 1993), the most closely related microbial β-glycanase topologies reported to date, for the following two reasons: all the β-strands of the cex-cd barrel are linked from one to the next by at least one α-helix, whereas neither CbhII nor $E2_{cd}$ include an α-helix between βVI, βVII, and βVIII, and the cex-cd α/β-barrel core includes an eighth parallel β-strand, unlike those of CbhII and $E2_{cd}$. In addition, the cex-cd active site crevice appears to be more accessible in comparison to the tunnel-like CbhII cleft, even though both enzymes exhibit exo activity. Moreover, because of its eight-stranded α/β-barrel and ten α-helices, the cex-cd is unlikely to be related to the single-helix six-stranded β-barrel of EngV (Davies et al., 1993), which is a member of family 45 (previously family K). The structure of the family 10 cex-cd more closely resembles the plant (1–3)- and (1–3,1–4)-β-glycanases reported by Varghese et al. (1994) as well as α-amylase (Luo et al., 1994), all of which do adopt $(\alpha/\beta)_8$ folds. Therefore, the cex-cd structure introduces yet another folding pattern to the β-glycanases isolated from microorganisms.

Active Site. A significant crevice exists on the protein surface, at the carboxy-terminal end of the β-barrel. Since this crevice is suitably exposed to the solvent, and since active sites of β-barrel enzymes are invariably found at the C-terminal end of the β-barrel, this is the likely location of the active site of cex-cd. This conclusion is confirmed by the identification of the two key catalytic residues, the nucleophile, Glu 233, and the acid/base catalyst, Glu 127, within that crevice. The two residues are suitably disposed within the site, with their carboxyl groups facing together at a separation of 5.5Å . This separation is similar to the one in other "retaining" β-glycanases whose structures have been solved: 5.4 Å between Glu78 and Glu172 for the *Bacillus circulans* xylanase (Campbell et al., 1993) and 5.5 Å between Asp52 and Glu35 of hen egg white lysozyme (Imoto et al., 1972). Such a separation is presumably optimal for the efficient formation of a glycosyl-enzyme intermediate on Glu233 of cex-cd while at the same time allowing Glu127 to protonate the departing aglycon in a

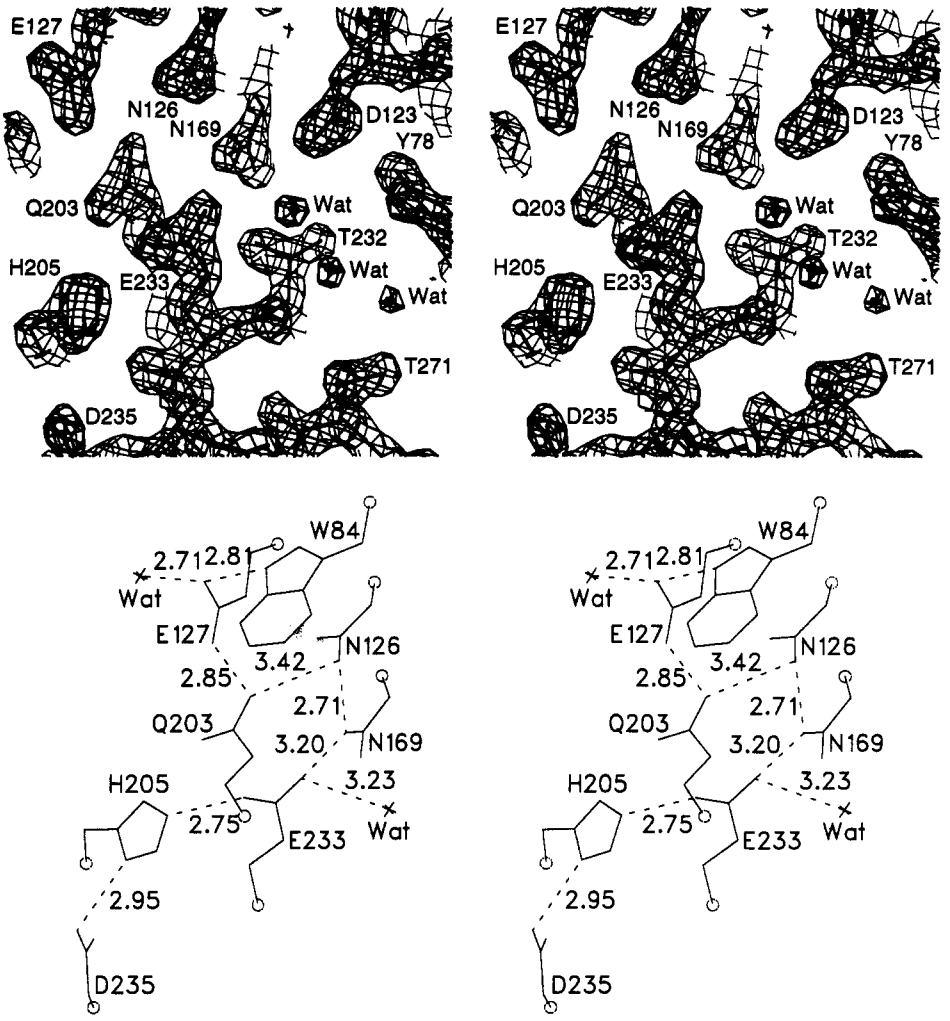

FIGURE 2: (a, top) Stereoview displayed with the program O of a representative section of the electron density map $2mF_o - DF_c$ contoured at 2.5σ. The view focuses on the active site of cex-cd. (b, bottom) Stereo diagram of the active site of cex-cd, as drawn by the program SETOR (Evans, 1993). For clarity, only selected side chains are shown along with their respective Cα (circles). Probable hydrogen bonds (dashed lines) and their lengths are indicated.

concerted manner. A greater separation between active site carboxyl groups is observed for the inverting β-glycosidase E2, an endocellulase from *T. fusca* (Spezio et al., 1993), in which 9.5 Å separates the acid catalyst Asp117 and Glu265; this distance is consistent with the different mechanism in which a water molecule, as well as the substrate, must be located between the two carboxyl groups.

A hydrogen-bonding network involving the catalytic residues is found at the active site and is likely important in determining their ionization behavior during catalysis (Figure 2). The nucleophile, Glu233, probably exists in a deprotonated form in the free enzyme to facilitate attack at the anomeric center. Glu233 is within hydrogen bond distance of both His205 and Asn169. His205 in turn interacts with Asp235, another conserved residue. Interestingly, in only 1 of the 20 or more enzymes from this family, ORF4 from *Caldocellum saccharolyticum* (Luethi et al., 1990), is this aspartic acid residue not conserved; it is replaced by a threonine. However, a compensating replacement of the (otherwise) conserved His205 by an asparagine maintains charge equivalence. As this trio of residues (Glu233–His205–Asp235) is highly conserved within the family, we postulate that it plays an important role in maintaining the ionization state of the nucleophile.

By contrast, the acid catalyst Glu127 must be in its protonated state for the first step in catalysis. This ionization state of Glu127 may be facilitated by interactions with Asn126, Gln203, Trp84, and a water molecule. Formation of the glycosyl-enzyme intermediate on Glu233 will affect the ionization behavior of Glu127 by decreasing the local charge density and possibly by altering the hydrogen-bonding network. As a consequence, Glu127 should be in an appropriate state for its role as a base in the second step of the mechanism.

A fourth conserved carboxylic amino acid located near the active site is Asp123. This residue is hydrogen-bonded to the active site conserved residue His80, which itself may be involved in substrate recognition. MacLeod (1994) noted that mutants in which Asp123 is replaced by Ala have considerably reduced catalytic activity, suggesting either some direct role of Asp123 in catalysis or some important role in modulating other interactions. A direct role in catalysis is unlikely, because the Asp123 side chain is not exposed in the active site. From the structure, it appears that replacement of this residue by Ala compromises the ability of His80 to interact with the substrate. Interestingly, much of the lost activity could be restored to the mutant by addition of exogenous azide. The structure would suggest that this restoration is due to azide binding at the site of the missing side chain and satisfying local charge requirements, thereby enabling His80 to better fulfill its role. Similar results have been obtained with mutants of trypsin, which could be reactivated by addition of exogenous acetate (Perona et al., 1994).

Inspection of sequence alignments of the proteins within

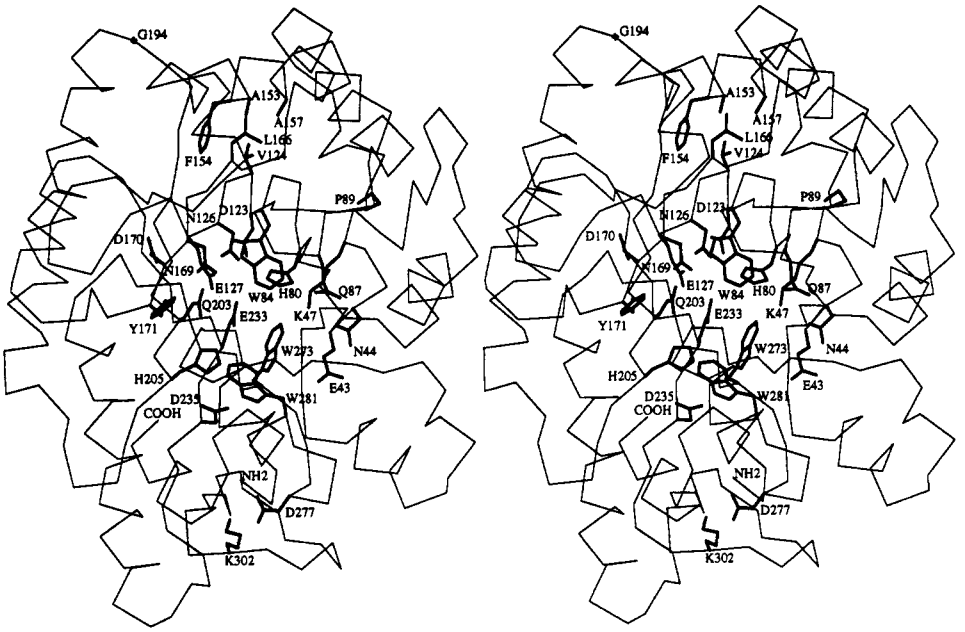

FIGURE 3: Projection in stereo, from the program O, of the Cα of cex-cd with side chains of residues (bold) conserved in family 10. Conserved residues are either identical in all, or all but one, known family 10 sequences.

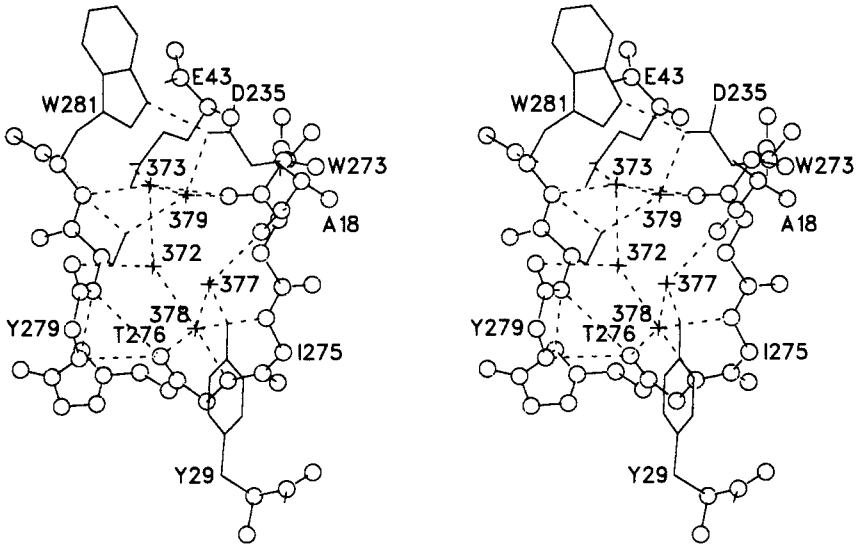

FIGURE 4: Stereo diagram, made with SETOR, of a cluster of five water molecules (crosses and unlettered numbers), which interacts with the loop of residues 273–281. For clarity, the backbone is shown with circles, and the side chains of only the residues that are within 3.5 Å of the cluster of water molecules are shown.

this family identifies a number of insertions that are present in some members. As might be expected, all these insertions are found to be on the outer surface of the protein, mostly located in loop regions, and are thus easily accommodated. Other critical residues for binding of the polysaccharide chain substrate can be inferred on the basis of their location in the structure and their conservation on sequence alignment. Figure 3 shows the location of the conserved residues from the sequence alignment of the 20 known members of family 10 (Gilkes et al., 1991b). With the exception of a five-residue hydrophobic cluster, most of these residues are located at or near the active site. This observation suggests a highly similar catalytic mechanism and substrate recognition for these enzymes.

One particularly useful piece of information in trying to identify the polysaccharide binding site and to assess the orientation of the chain within that site is the recent work on mutants of xylanase A from *Streptomyces lividans* (Moreau et al., 1994), an enzyme from the same sequence-related family. On the basis of kinetic studies with a mutant in which the conserved residue Asn173 was replaced by Asp, the authors concluded that this residue was involved in important hydrogen-bonding interactions with a xylose residue three units away from the cleavage site, in the direction of the reducing terminus. The structure of cex-cd is in complete agreement with this proposal, since this residue (Asn172 in Cex) is located some 11.3 Å away from the nucleophilic residue Glu233. Such a distance is consistent with the placement of at least two complete sugar residues between these two amino acid residues, with Asn172 interacting with the third. This observation also suggests that when the scissile bond is located between Glu127 and Glu233, the reducing end of the polysaccharide is oriented toward Asn172, and thus the nonreducing disaccharide moiety extends toward Gln87, another completely conserved residue.

The active site cleft of cex-cd also contains several tryptophans (84, 273, 281); these three residues extend their side chains into the vicinity of Glu127 and Glu233. The tryptophans are likely to be involved in the stacking of either

cellulose or xylan in the active site, but Trp273 and Trp281 may also have a structural role. These two tryptophans delimit the extent of the loop which joins strand β8 to helix α8. The backbone of this loop is tightly stabilized on the surface of the protein by a well-ordered, internal cluster of five water molecules (Figure 4). These are within tight hydrogen-bonding distance to several backbone and side-chain atoms within the loop. Furthermore, the middle of this loop is pinned to the rest of the structure by a tight salt bridge between Asp277 and Lys302, two residues that are completely conserved in all members of the family. This interaction further points to an important role for this loop, though the nature of this role has not yet been defined.

ACKNOWLEDGMENT

We are grateful to Alasdair MacLeod, Dedreia Tull, Emily Kwan, and Peter Tomme for useful discussions, Kathy Johns for help with crystal handling, and Drs. R. Antony J. Warren and Doug Kilburn for their support and encouragement.

NOTE ADDED IN PROOF

Since acceptance of this paper, a 2.6-Å resolution structure of another family 10 glycosyl hydrolase has been published (Derewenda et al., 1994).

REFERENCES

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., & Waley, S. G. (1975) *Nature (London) 255*, 609–614.

Bedarkar, S., Gilkes, N. R., Kilburn, D. G., Kwan, E., Rose, D. R., Miller, R. C., Jr., Warren, R. A. J., & Withers, S. G. (1992) *J. Mol. Biol. 228*, 693–695.

Brünger, A. T. (1992) *Nature (London) 355*, 472–475.

Brünger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458–460.

Campbell, R., Rose, D., Wakarchuk, W., To, R., Sung, W., & Yagachi, M. (1993) in *Proceedings of the second TRICEL symposium on Trichoderma reesei cellulases and other hydrolases* (Suominen, P., & Reinikainen, T., Eds.) pp 63–72, Helsinski, Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland.

Chan, A. W. E., Hutchinson, E. G., Harris, D., & Thornton, J. M. (1993) *Protein Sci. 2*, 1574–1590.

Davies, G. J., Dodson, G. G., Hubbard, R. E., Tolley, S. P., Dauter, Z., Wilson, K. S., Hjort, C., Mikkelsen, J. M., Rasmussen, G., & Schulein, M. (1993) *Nature (London) 365*, 362–364.

Derewenda, U., Swenson, L., Green, R., Wei, Y., Morosoli, R., Shareck, F., Kluepfel, D., & Derewenda, Z. S. (1994) *J. Biol. Chem. 269*, 20811–20814.

Evans, S. (1993) *J. Mol. Graphics 11*, 134–138.

Farber, G. K., & Petsko, G. A. (1990) *Trends Biochem. Sci. 15*, 228–234.

Furey, W., & Swaminathan, S. (1990) *Proceedings of the American Crystallography Association Meeting*, New Orleans, LA.

Gilkes, N. R., Langsford, M. L., Kilburn, D. G., Miller, R. C., Jr., & Warren, R. A. J. (1984) *J. Biol. Chem. 259*, 10455–10459.

Gilkes, N. R., Claeyssens, M., Aebersold, R., Henrissat, B., Meinke, A., Morisson, H. D., Kilburn, D. G., Warren, R. A. J., & Miller, R. C., Jr. (1991a) *Eur. J. Biochem. 202*, 3657–3677.

Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr., & Warren, R. A. J. (1991b) *Microbiol. Rev. 55*, 303–315.

Hazlewood, G. P., & Gilbert, H. J. (1993) in *Hemicellulose and hemicellulases* (Coughlin, M. P., & Hazlewood, G. P., Eds.) pp 102–126, Portland Press, London.

Henrissat, B., & Bairoch, A. (1993) *Biochem. J. 293*, 781–788.

Hodel, A., Kim, S.-H., & Brünger, A. T. (1992) *Acta Crystallogr. A48*, 851–858.

Howard, A. J., Neilsen, C., & Xuong, N. H. (1985) *Methods Enzymol. 114*, 452–472.

Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C., & Rupley, J. A. (1972) in *The Enzymes* (Boyer, P. D., Ed.) pp 666–668, Academic Press, New York.

Jones, T. A., & Thirup, S. (1986) *EMBO J. 5*, 819–822.

Jones, T. A., Bergdoll, M., & Kjelddgaard, M. (1990) in *Crystallographic and modeling methods in molecular design* (Bugg, C. E., & Ealick, S. E., Eds.) pp 63–72, Springer-Verlag, New York.

Jones, T. A., Zou, J.-Y., Cowan, S. W., & Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110–119.

Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssens, M., Béguin, P., & Aubert, J.-P. (1992) *Nature (London) 357*, 89–91.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.

Keitel, T., Simon, O., Borriss, R., & Heinemann, U. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5287–5291.

Kleywegt, G. J., & Jones, T. A. (1994) *MAPMAN—the manual (Version 3.3)*, Uppsala University, Uppsala.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283–291.

Luethi, E., Love, D. R., McAnulty, J., Wallace, C., Caughey, P. A., Saul, D., & Bergquist, P. L. (1990) *Appl. Environ. Microbiol. 56*, 1017–1024.

Luo, Y., Withers, S. G., & Brayer, G. D. (1994) in Proceedings of the American Crystallography Association Meeting, Atlanta, GA.

MacLeod, A. M. (1994) Ph.D. Thesis, University of British Columbia, Vancouver, Canada.

MacLeod, A. M., Lindhorst, T., Withers, S. G., & Warren, R. A. J. (1994) *Biochemistry 33*, 6371–6376.

Meinke, A., Gilkes, N. R., Kwan, E., Kilburn, D. G., Warren, R. A. J., & Miller, R. C., Jr. (1994) *Mol. Microbiol. 12*, 413–422.

Moreau, A., Shareck, F., Kluepfel, D., & Morosoli, R. (1994) *Eur. J. Biochem. 219*, 261–266.

O'Neill, G., Goh, S. H., Warren, R. A. J., Kilburn, D. G., & Miller, R. C., Jr. (1986) *Gene 44*, 325–330.

Perona, J. J., Hedstrom, L., Wagner, R. L., Rutter, W. J., Craik, C. S., & Fletterick, R. J. (1994) *Biochemistry 33*, 3252–3259.

Pickersgill, R. W., Jenkins, J. A., Scott, M., & Connerton, I. (1993) *J. Mol. Biol. 229*, 246–248.

Read, R. J. (1986) *Acta Crystallogr. A42*, 140–149.

Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., & Jones, T. A. (1990) *Science 249*, 380–386.

Shen, H., Tomme, P., Meinke, A., Gilkes, N. R., Kilburn, D. G., Warren, R. A. J., & Miller, R. C., Jr. (1994) *Biochem. Biophys. Res. Commun. 199*, 1223–1228.

Souchon, H., Spinelli, S., Béguin, P., & Alzari, P. M. (1994) *J. Mol. Biol. 235*, 1348–1350.

Spezio, M., Wilson, D. B., & Karplus, P. A. (1993) *Biochemistry 32*, 9906–9916.

Tull, D., & Withers, S. G. (1994) *Biochemistry 33*, 6363–6370.

Tull, D., Withers, S. G., Gilkes, N. R., Kilburn, D. G., Warren, R. A. J., & Aebersold, R. (1991) *J. Biol. Chem. 266*, 15621–15625.

Varghese, J. N., Garrett, T. P. J., Colman, P. M., Chen, L., & Høj, P. B. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2785–2789.

Wang, B. C. (1985) *Methods Enzymol. 115*, 90–112.

Withers, S. G., Dombroski, D., Berven, L. A., Kilburn, D. G., Miller, R. C. Jr., Warren, R. A. J., & Gilkes, N. R. (1986) *Biochem. Biophys. Res. Commun. 139*, 487–494.

Wong, K. K. Y., & Saddler, J. N. (1992) *Crit. Rev. Biotechnol. 12*, 413–435.