Crystal Structure of the Catalytic Domain of a Thermophilic Endocellulase[†,‡]

Michael Spezio, David B. Wilson, and P. Andrew Karplus*

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853

Received February 5, 1993; Revised Manuscript Received June 7, 1993[⊕]

ABSTRACT: One way to improve the economic feasibility of biomass conversion is to enhance the catalytic efficiency of cellulases through protein engineering. This requires that high-resolution structures of cellulases be available. Here we present the structure of $E2_{cd}$, the catalytic domain of the thermophilic endocellulase E2 from *Thermomonospora fusca*, as determined by X-ray crystallography. The structure was solved by multiple isomorphous replacement at 2.6-Å resolution and has been refined at 1.8-Å resolution to an *R*-value of 18.4% for all reflections between 10- and 1.8-Å resolution. The fold of $E2_{cd}$ is based on an unusual parallel β-barrel and is equivalent to the fold determined for the catalytic domain of cellobiohydrolase II, an exocellulase from *Trichoderma reesei* [Rouvinen et al. (1990) *Science 249*, 380–385]. The active site cleft of the enzyme, approximately 11 Å deep and running the entire length of the molecule, is seen to be completely free for ligand binding in the crystal. A 2.2-Å resolution analysis of crystals of $E2_{cd}$ complexed with cellobiose, an inhibitor, shows how cellobiose binds in the active site and interacts with several residues which line the cleft. Catalytic roles are suggested for three aspartic acid residues at the active site. A comparison of the $E2_{cd}$ and $CBHII_{cd}$ structures reveals a large difference in their active site accessibilities and supports the hypothesis that the main difference between endo- and exocellulases is the degree to which their active sites are accessible to substrate.

Cellulases make up a diverse class of enzymes which hydrolyze cellulose into small polymers of glucose. Individual cellulolytic microorganisms produce an array of different cellulases whose activities complement one another (Coughlan, 1992). The enzymes can be grouped into two broad classes: endocellulases, which cleave anywhere along cellulose chains; and exocellulases, which cleave only from the ends of cellulose chains. *In vitro* studies with purified enzymes demonstrate the need for these multiple components in order to achieve maximal cellulose degradation. Mixtures of different cellulases exhibit up to 8-fold synergism compared with individual enzymes (Irwin et al., 1993).

Interest in cellulases focuses mainly on two issues. The first concerns basic research into the following areas: how a cellulase isolates one strand of cellulose from an insoluble cellulose complex, the chemistry of the catalysis itself, and how cellulases interact to give synergistic activity. The second is more practical and results from the effort to move to renewable energy resources by making biomass conversion more economically feasible. Currently, cellulases account for 40% of the total cost of biomass conversion, so that an increase in their catalytic efficiency would result in significant savings for the industry (Lynd et al., 1991). Such an increase could be achieved through protein engineering. Since cellulase activity does not appear to be a limiting factor for microbial growth (Tribe, 1966), it is possible that cellulase catalysis has not been optimized through evolution and that protein engineering may be applied to increase the specific activity of cellulases. Both the basic and applied aspects of cellulase research require knowledge of the atomic structure of cellulases.

The thermophilic actinomycete *Thermomonospora fusca* produces at least six cellulases, E1 through E6, which show synergism (Wilson, 1988; Irwin et al., 1993). Endocellulase E2 (M_r = 42 000) exhibits the highest activity of any cellulase from *T. fusca* in cellulose fragmentation assays (Walker et al., 1992) and is optimally active at 55 °C. Several of its properties have been described (Calza et al., 1985), and it is very active on various cellulosic substrates. The gene for E2 has been cloned, sequenced, and expressed in *Escherichia coli* and *Streptomyces lividans* (Ghangas & Wilson, 1988; Lao et al., 1991). The natural protein is lightly glycosylated, and although the recombinant protein lacks the natural glycosylation, this appears to have no effect on activity (Ghangas & Wilson, 1988).

E2, like many cellulases, possesses two functional domains (Ghangas & Wilson, 1988). The amino-terminal domain carries out catalysis while the carboxy-terminal domain (cellulose binding domain) enables the enzyme to bind very tightly to insoluble cellulose. The catalytic domain ($E2_{cd}$,[1] M_r = 30 000) can be proteolytically cleaved from the rest of the protein and purified (Irwin et al., 1993). While the exact carboxy terminus of $E2_{cd}$ is not known, its amino terminus is identical to that of intact E2 (Irwin et al., 1993).

To date, structural determinations have been published for three unrelated cellulases. These are the NMR structure of the cellulose binding domain of cellobiohydrolase I (CBHI) from the fungus *Trichoderma reesei* (Kraulis et al., 1989), the crystal structure of the catalytic domain of cellobiohydrolase II (CBHII) from *T. reesei* (Rouvinen et al., 1990),

[†] Supported by Grant DE-FG02-84ER13233 from the Department of Energy and by a grant from the Cornell Biotechnology Program, which is sponsored by the New York State Science and Technology Foundation, a consortium of industries, the U.S. Army Research Office, and the National Science Foundation. M.S. was supported by a National Institutes of Health training grant and a Cornell Biotechnology Fellowship, funded by the U.S. Army Research Office.

[‡] Coordinates and structure factors have been deposited in the Protein Data Bank (Acquisition Number 1TML).

* To whom correspondence should be addressed.

[⊕] Abstract published in *Advance ACS Abstracts*, September 1, 1993.

[1] Abbreviations: CMC, (carboxymethyl)cellulose; AS, ammonium sulfate; PEG, poly(ethylene glycol); TLC, thin-layer chromatography; MIR, multiple isomorphous replacement; MIRAS, MIR-including anomalous scattering; SF-MIRAS, solvent-flattened MIRAS; pc, Patterson correlation; $E2_{cd}$, the catalytic domain of endocellulase E2; $CBHII_{cd}$, the catalytic domain of cellobiohydrolase II.

0006-2960/93/0432-9906$04.00/0

and the crystal structure of endoglucanase D (CelD) from the bacterium *Clostridium thermocellum* (Juy et al., 1992). Amino acid sequence comparisons (Lao et al., 1991) show that E2 is homologous to CBHII and these two enzymes belong to cellulase family B in the nomenclature of Henrissat et al. (1989). The catalytic domains of E2 and CBHII share 26% sequence identity.

Extensive characterization of the cellulolytic activity of E2 has recently been carried out (Irwin et al., 1993) and shows that E2, compared to CBHII, possesses 300-fold higher activity on (carboxymethyl)cellulose (CMC) and 20-fold higher activity on swollen cellulose (Irwin et al., 1993). The structure determination of $E2_{cd}$ provides an opportunity to obtain detailed structural explanations for these differences, to gain insight into structure–function relationships of this family of cellulases, and to initiate structure-based efforts to engineer improved catalytic efficiency into this enzyme.

EXPERIMENTAL PROCEDURES

Purification and Crystallization. Recombinant $E2_{cd}$ was purified from *S. lividans* as described (Irwin et al., 1993). The protein was kept at a concentration of 15–18 mg/mL in 2 mM 2-(*N*-morpholino)ethanesulfonic acid (Sigma) and 0.04% (w/v) NaN_3, pH 6.0, at −80 °C. All crystallization trials were performed using the hanging drop method (McPherson, 1989) set up with Linbro tissue culture plates (Flow Laboratories, Inc.). For the reproducible growth of diffraction-quality crystals, hanging drops containing 3 μL each of protein solution and reservoir [0.056 M disodium citrate, pH 4.0, 10 mM $CaCl_2$, and 12% ammonium sulfate (AS) as precipitant] and 1 μL of a suspension of microcrystals were equilibrated at room temperature against said reservoir. The suspension of microcrystals was made by adding a drop of crystals to 0.5 mL of crystal storage buffer (identical to reservoir but containing 25% AS). The crystals were crushed in a glass homogenizer, and a fresh aliquot of this homogenate was diluted 1:1000 with storage buffer for each set of crystallization experiments. Crystal storage buffer was always employed while the crystals were handled. Protein concentration was calculated from the absorbance of the protein solution at 280 nm (ϵ = 57 600 M^{-1} cm^{-1}; Irwin et al., 1993).

Data Collection. For diffraction experiments, single crystals were mounted in thin-walled glass capillaries (Charles Supper Co.). Their space group was determined by taking 15°-precession photographs of each major zone using a Supper precession camera and nickel-filtered Cu Kα radiation from a Philips XRG-3000 sealed-tube X-ray generator operating at 30 kV and 38 mA. All quantitative X-ray diffraction data were collected at room temperature by a single multiwire area detector from San Diego Multiwire Systems (SDMS; Hamlin, 1985) using a Huber three-circle goniostat and Cu Kα radiation obtained from a graphite monochrometer on a Rigaku RU-200 rotating anode operating with a 0.5 × 5 mm point focus at 50 kV and 150 mA. Data collection and reduction were done using standard software provided by SDMS (Howard et al., 1985). The crystal-to-detector distance was approximately 360 mm through air, and data were collected in 0.1–0.12° frames with an exposure time of 35–45 s/frame, depending on how well each crystal diffracted. Following the data collection strategy described by Xuong et al. (1985), a complete data set to 2.6-Å resolution could be collected within 24 h using four 70° sweeps in ω and stationing the detector at θ_c = −10°. Higher resolution data were collected at θ_c = −40°. For the second uranyl and mercurial data sets, anomalous diffraction data were collected by following each ω-sweep with an equivalent sweep carried out with $\varphi' = \varphi - 180°$ and $\chi' = -\chi$ (Xuong et al., 1985). Due to the small difference in the lengths of the *a*- and *c*-axes (*a* = 43.35 Å, *c* = 43.41 Å), the *h* and *l* indices were occasionally permuted during autoindexing. Such cases were recognized when the new data showed no correlation with the reference native data set, and the problem was easily corrected by invoking the *permind* function in the SDMS software during data reduction.

Molecular Replacement. Molecular replacement (Blundell & Johnson, 1976b) attempts with X-PLOR (Brunger, 1990) used a polyalanine model of $CBHII_{cd}$ (Rouvinen et al., 1990; kindly provided by T. A. Jones) after the removal of segments which an amino acid sequence alignment indicated were not present in $E2_{cd}$. These segments consisted of residues 83–89, 119–127, 177–185, 227–233, 272–274, 294–296, 309–316, 342–346, 362–365, and 402–415. All calculations were done using data between 15- and 4-Å resolution, and Patterson correlation (pc) refinement was performed on each of the top 145 solutions of the rotation function, allowing each secondary structural element in $CBHII_{cd}$ to move independently as a rigid body.

Structure Solution. Phase determination proceeded by the method of multiple isomorphous replacement (MIR) (Blundell & Johnson, 1976a). Heavy atom derivatives were prepared by soaking native crystals in solutions of heavy atom salts in the dark at room temperature. In heavy atom soak experiments, storage solutions employed 0.075 M sodium acetate in place of citrate to avoid problems from the known chelating properties of citrate. Potential heavy atom positions were identified from Harker sections of difference Patterson maps, and cross-peaks expected for the potential sites identified from the Harker sections were checked against the respective maps. Heavy atom models were refined, and phases from these models were calculated by the method of Dickerson et al. (1968) as implemented in the program DAREFI (G. E. Schulz, personal communication). Final phases were calculated utilizing four heavy atom derivatives with their anomalous diffraction signal (Matthews, 1966).

For construction of an initial model of $E2_{cd}$, a 2.8-Å resolution MIR-including anomalous scattering (MIRAS) electron density map was displayed on acetate sheets (scale = 4 Å/cm), and a total of 46 C_α positions from five β-strands and two α-helices were located. The polyalanine model of $CBHII_{cd}$ was rotated into the electron density of $E2_{cd}$ based on an overlay (Kabsch, 1976, 1978) of 18 equivalent C_α atoms located in four β-strands. The amino acid sequence of the model was changed to that of $E2_{cd}$ by adding side chains in FRODO (Jones, 1985) and removing segments corresponding to insertions relative to $E2_{cd}$ and adding one segment corresponding to a deletion. After the resulting model was subjected to energy minimization in X-PLOR (Brunger et al., 1987), residues 1–287 of the model were manually fit into the $E2_{cd}$ MIRAS map using the graphics program FREIBAU (Karplus et al., 1990). A round of refinement at 2.8-Å resolution (Table III) was done in X-PLOR following described protocols (Brunger et al., 1990). A solvent-flattened MIRAS (SF-MIRAS) map was then calculated (Wang, 1985) with phases to 2.6-Å resolution, and the refined 2.8-Å model was refit to this improved, experimentally phased map. Solvent flattening was done in three rounds of 17 cycles each, using a higher solvent level for each progressive round. Since the crystals were estimated to contain 36% solvent (see Results), levels of 20%, 30%, and finally 33% solvent were used. The envelope calculated at 33% solvent flattened some surface side-chain density, and the 2.6-Å MIRAS map was thus used as a reference during fitting. Further refinement and fitting

to improve the model are described in Table III. At each stage, individual temperature factors were refined prior to the calculation of $(2F_o - F_c)\alpha_c$ maps. However, these values were not kept for the subsequent refinement step until data to 1.8-Å resolution were used. Water molecules were added conservatively at positions showing strong peaks ($\geq 4\sigma$) in $(F_o - F_c)\alpha_c$ maps and having reasonable hydrogen-bonding geometry. Manual refitting employed the graphics programs FREIBAU (Karplus et al., 1990) and CHAIN (Sack, 1988).

Ligands. All data on putative ligand-bound crystals were collected as described above. $E2_{cd}$ crystals soaked in 5 mM cellobiose, an inhibitor, failed to show any binding of the ligand, so a cocrystallization approach to complex formation was used. $E2_{cd}$ was crystallized in the presence of 5 mM cellobiose under normal crystallization conditions to yield an $E2_{cd}$–cellobiose complex. The crystal structure of cellobiose (Chu & Jeffrey, 1967) was used as an initial model for cellobiose in manual fitting to an $(F_{o,soak} - F_{o,nat})\alpha_c$ difference Fourier map employing the data from the resulting $E2_{cd}$–cellobiose cocrystals (Table I). A number of unsuccessful attempts were made to obtain a complex of $E2_{cd}$ and glucose, with or without cellobiose bound. A crystal which had been grown in the presence of 2 mM cellobiose was soaked at room temperature in storage buffer (25% AS) containing 5 mM cellobiose and 100 mM glucose. Native crystals of $E2_{cd}$ were also soaked in storage buffer containing 100 mM glucose in order that a complex of $E2_{cd}$ and glucose alone might be observed.

Enzymatic Activity of Crystals. To assay the activity of native $E2_{cd}$ crystals, the crystals were washed in storage buffer (2 × 2 mL) and placed in a 1-mm thin-walled glass capillary tube (in order that any significant dissolution of the crystal would be seen) in 50 μL of crystal storage buffer containing 3 mM cellotetraose (Sigma) and sodium acetate in place of disodium citrate. The sizes of the crystals ranged from 0.002 to 0.005 mm^3, corresponding to approximately 2–5 μg of protein. In a control experiment to assess the activity of noncrystalline enzyme in crystal storage buffer, 1 μg of soluble $E2_{cd}$ was incubated in 50 μL of the same buffer. Additional controls to assess the contribution of enzyme that may have dissolved from the crystal surface involved soaking individual washed crystals in storage buffer in capillaries overnight and using 10 μL of the "saturated" storage buffer from these soaks as enzyme. All reactions took place at room temperature and were stopped after 2, 10, or 30 min by the addition of 150 μL of Millipore water containing 10 mg of Norit-A (National Biochemical Corp.), which binds oligosaccharides. In the case of the crystals, the reaction buffer (approximately 40 μL) was removed from the capillary and added to Norit-A, leaving the crystal inside the capillary. The resulting mixtures were incubated at room temperature for 30 min, and the Norit-A was pelleted, washed twice with 400 μL of Millipore water to remove excess salt which interferes with the thin-layer chromatography, and repelleted, discarding the supernatant. Oligosaccharides were extracted from Norit-A with 25% (v/v) ethanol (2 × 200 μL extractions, incubating each for 10–30 min at room temperature). The resulting 400-μL samples were vacuum dried at room temperature overnight and resuspended in 15 μL of 25% (v/v) ethanol. Thin-layer chromatography (TLC) proceeded by loading the sample onto a lane of a 20 cm × 20 cm silica gel TLC plate (Whatman LK5D) and performing the TLC and staining the oligosaccharides as previously described (Chirico et al., 1985; Krebs et al., 1969).

Sequence Analysis and Alignment. Initially, the three endocellulases in family B were individually aligned with the $E2_{cd}$ sequence in order to assess their relative similarities to $E2_{cd}$. This procedure employed the BESTFIT program in the UWGCG program package (Devereux et al., 1992), using a gap penalty of 5 and a gap-length penalty of 0.3. An optimal family alignment was then built by adding the three endocellulase sequences, one at a time in order of decreasing similarity, to the sequence of $E2_{cd}$ (CelA, then CasA, then CenA). These alignments were done with the statistical alignment package from Kansas State University (Reeck et al., 1982), using the McLachlan similarity matrix (McLachlan, 1971). The programs were modified to disallow placement of gaps in user-defined secondary structural elements (P. A. Karplus, unpublished). The most statistically significant alignment was found by starting with a broad range of gap penalties and comparing to 100 randomizations at each penalty. Optimization of the alignment utilized 300 randomizations at each gap penalty within a narrower range.

For comparison of the structures of $E2_{cd}$ and $CBHII_{cd}$, the C_α coordinates of $CBHII_{cd}$ were obtained from the Protein Data Bank (Bernstein et al., 1977). To produce the structural alignment between $E2_{cd}$ and $CBHII_{cd}$, an initial overlay (Kabsch, 1976, 1978) of the two structures was conducted on the basis of 50 corresponding C_α coordinates, most of which were located within the β-strands. Using this alignment, corresponding residues in all major secondary structural elements were identified, and these were employed in a structural alignment by the method of Chothia and Lesk (1986).

RESULTS AND DISCUSSION

Crystal Growth and Characterization. The effort to crystallize $E2_{cd}$ began by screening a series of AS and poly(ethylene glycol) (PEG) conditions at a variety of pH values (McPherson, 1989). All initial screening was done with 0.2 M McIlvaine's citrate–phosphate buffer system, which mixes citric acid and disodium phosphate to generate pH values between 4.0 and 8.0 (McIlvaine, 1921). During initial screening, crystalline aggregates (Figure 1a) were obtained at pH 4.0 and 20% AS at room temperature. Although these crystals were not useful for structure analysis, X-ray diffraction experiments with one such aggregate mounted in a glass capillary showed diffraction extending beyond 3.0-Å resolution, providing incentive for further trials to optimize conditions. Early attempts yielded only aggregates or, at lower concentrations of AS, no crystals at all. Microseeding with crushed crystalline aggregates as seeds did, however, yield improved crystals. Three rounds of microseeding, each time using improved crystals as seeds, resulted in reproducible large single crystals appearing within 2 days, reaching maximal growth within 3 weeks. The crystals (Figure 1b) grow up to 0.15 mm × 0.40 mm × 1.8 mm. These results suggest that, at conditions which allow nucleation to occur (near 20% AS), the growth of $E2_{cd}$ crystals is too rapid to yield good crystals, while at conditions optimal for controlled growth (near 12% AS), nucleation does not occur and requires the addition of seeds.

The space group was determined to be $P2_1$ by analysis of 15°-precession photographs of each major plane. Preliminary unit cell constants taken from the precession photographs were $a = 44$ Å, $b = 66$ Å, $c = 44$ Å, and $\beta = 108°$. Refined unit cell parameters from two crystals used for native data collection are $a = 43.35$ Å, $b = 65.94$ Å, $c = 43.41$ Å, and $\beta = 107.69°$. This unit cell only allows for a single molecule per asymmetric unit, corresponding to a solvent content of 36%, which is in the normal range expected for protein crystals (Matthews, 1968). The densely packed $E2_{cd}$ crystals are unusually well ordered, as they diffract to beyond 1.4-Å resolution.

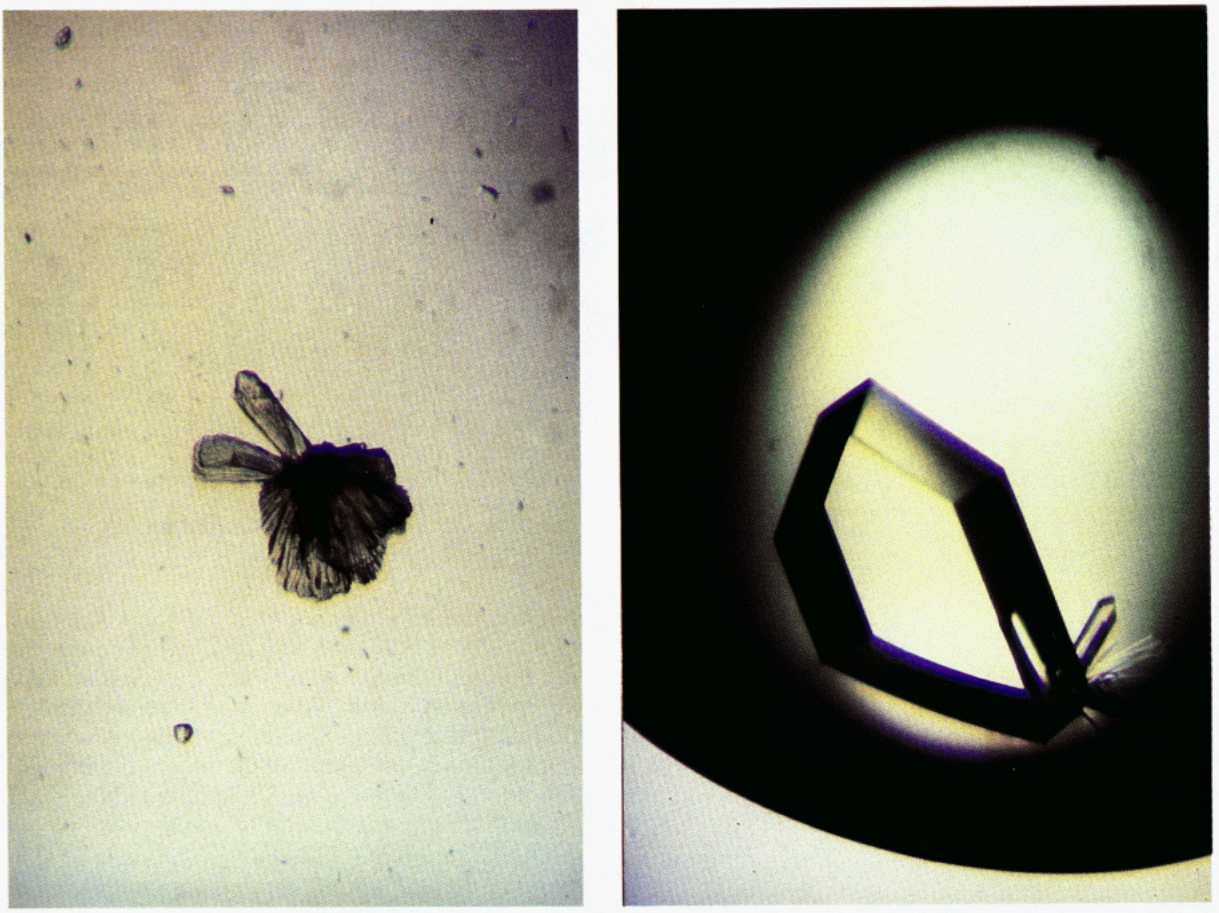

FIGURE 1: Crystals of $E2_{cd}$. (a, left) Initial crystals of $E2_{cd}$ grown from 20% ammonium sulfate. Largest dimension approximately 0.3 mm. (b, right) Crystals of $E2_{cd}$ grown from microseeds at 12% ammonium sulfate. Largest dimension approximately 1.2 mm.

Table I: Data Collection Statistics for $E2_{cd}$

crystal (no. of crystals)	resolution (Å)	soak time (days)	soak concn (mM)	measured reflections	unique reflections (% complete, % complete anomalous)	R^a_{sym} (%)	$R_{nat}{}^b$ (%)	ΔB^c (Å^2)
native (2)	1.8			79 576	21 081 (97)	4.59		
$UO_2(NO_3)_2{\cdot}6H_2O$ (1)	2.7	3	20	15 682	6 293 (94, 35)	8.47	17	0.85
$UO_2(NO_3)_2{\cdot}6H_2O$ (1)	2.6	4	30	31 658	6 602 (94, 85)	8.05	17	0.48
$HgCl_2$ (1)	2.6	8	30	31 350	6 768 (94, 85)	5.94	20	0.08
Na_2AuCl_4 (1)	2.6	4	0.4	17 225	6 809 (94, 37)	10.28	36	0.36
K_2PtCl_4 (1)	2.6	4	0.5	17 224	6 895 (95, 34)	12.14	25	5.50
cellobiose cocrystal (1)	2.2			20 844	10 654 (92)	2.82	8	2.3

[a] $\sum|I - \langle I\rangle|/\sum I$. [b] $\sum|F_{PH} - KF_P|/\sum F_P$. [c] Overall temperature factor of derivative data set relative to native.

Structure Determination. Molecular replacement with the modified CBHII model failed to yield a clear solution to the rotation function. When subsequent pc refinement of top results from the rotation search also failed to provide a definitive answer, translation searches were done using the top pc-refinement solutions. This too gave no solution, and molecular replacement was abandoned in favor of a search for heavy atom derivatives.

A total of 17 heavy atom compounds were screened, and 4 of these proved to be useful isomorphous derivatives. Data collection statistics for the 4 derivatives are reported in Table I. Phasing was initiated when the Patterson map of the uranyl derivative was interpreted and the resulting two-site model was refined. The single isomorphous replacement phases thus obtained were used to find sites in other derivatives by standard difference Fourier techniques. These sites were checked for consistency against the respective difference Patterson maps of the derivatives and were refined. To determine the absolute hand of the structure, two trial refinements, one for each hand, were carried out using only the uranyl derivative and its anomalous signal. Difference Fourier analysis of the other derivatives uniformly showed higher signal-to-noise ratios for one of the hands, and this was chosen. Because preliminary phase calculations showed that the uranyl derivative clearly gave the best signal at higher resolution, a soak at higher uranyl nitrate concentration was used to collect fairly complete anomalous diffraction data. A second data set to maximize anomalous diffraction data was also collected for a higher mercuric chloride soak. Final phasing was carried out with the second uranyl and mercurial data sets and data sets from two additional derivatives (Table I) to yield phases with an overall figure of merit of 0.732 to 2.6-Å resolution (Figure 2).

Reference to the refined $E2_{cd}$ model shows that most of the heavy atom sites are primarily at glutamic acid, methionine, or histidine residues (Table II), as is commonly observed for these metals. The three exceptions are weakly occupied sites of the mercury, gold, and platinum derivatives (Table II). However, difference maps calculated using final refined phases confirm the validity of these sites.

The four derivative data sets were nearly 100% complete to 2.8 Å and less than 80% complete between 2.8 and 2.6 Å. Therefore, model building was initiated with a 2.8-Å MIRAS

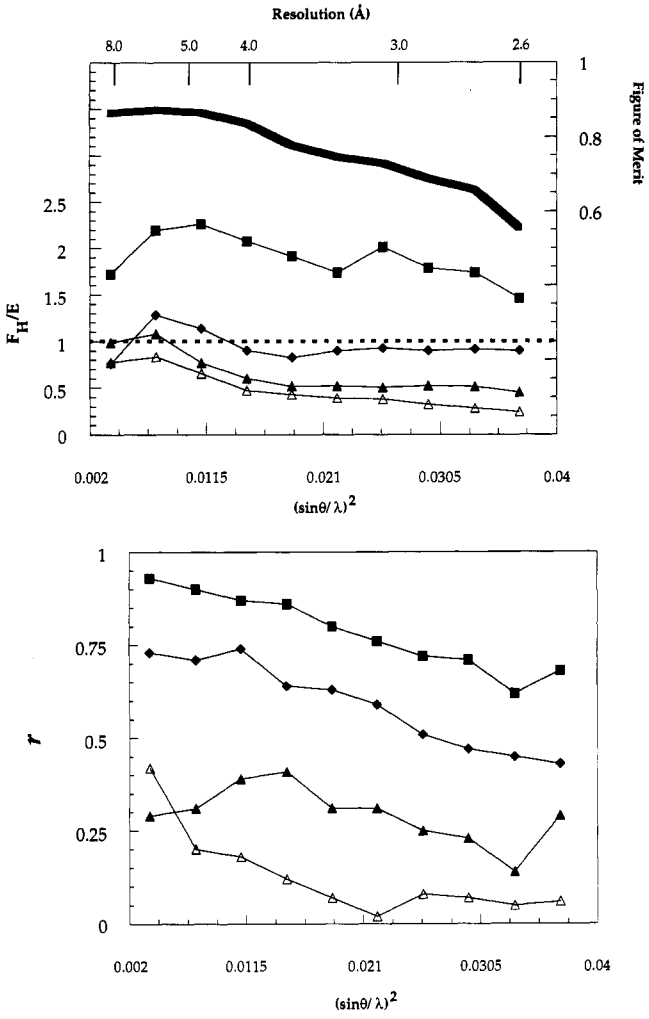

FIGURE 2: (a, top) Phasing power and figure of merit. Phasing power for the four derivatives used and the figure of merit of the resulting phases are plotted as a function of resolution. Symbols: (—) figure of merit; (■) uranyl nitrate; (◆) mercuric chloride; (▲) sodium tetrachloroaurate; (△) potassium tetrachloroplatinate. Phasing power is defined as the rms calculated heavy atom scattering divided by the rms lack of closure error for the phasing triangles. (b, bottom) Correlation between the observed and calculated anomalous signal. The correlation coefficient between the observed and calculated anomalous signal is plotted as a function of resolution for all four derivatives used in MIRAS phasing. The symbols for derivatives are identical to those in (a). Both the uranyl and mercurial derivatives have high correlation coefficients, showing that both gave a good anomalous signal. Correlation coefficient $r = \sum[(\Delta F_o - \overline{\Delta F_o})\Delta F_c - \overline{\Delta F_c})]/[\sum(\Delta F_o - \overline{\Delta F_o})^2\sum(\Delta F_c - \overline{\Delta F_c})^2]^{1/2}$.

map. This was facilitated by using $CBHII_{cd}$ as a template in the β-barrel region (see Experimental Procedures). One round of simulated annealing refinement using data between 7.0- and 2.8-Å resolution in X-PLOR brought the R-value down from 50.2% to 24.1%, but the subsequent $(2F_o - F_c)\alpha_c$ map showed heavy model bias and was not suitable for use in fitting. Further attempts to better fit this refined model to the MIRAS density were hampered by ambiguous connectivity and the lack of carbonyl and side-chain density in some regions of the map. This problem was alleviated by using all of the derivative data, extending the MIRAS phasing to 2.6 Å, and carrying out solvent flattening (Wang, 1985). This map showed carbonyl bumps and solvent density in many places (Figure 3a), and completely rebuilding the $E2_{cd}$ model into this improved density resulted in a coordinate set which showed much better agreement with the 2.6-Å MIRAS phases (Table III) and which refined easily. The course of refinement is given in Table III. The current model has an R-value of

Table II: Individual Heavy Atom Sites

heavy atom	site	X	Y	Z	occupancy[a]	B[a]	amino acid neighbors[b]
U	1	16.07	65.94	−1.98	0.70	28.5	E281, $O_{\epsilon 1}$, $O_{\epsilon 2}$; N1*, $O_{\delta 1}$
	2	−2.81	38.99	−32.93	0.47	81.5	E135, $O_{\epsilon 1}$, $O_{\epsilon 2}$; E132, $O_{\epsilon 2}$
Hg	1	33.78	52.48	−13.17	0.36	24.1	E15, $O_{\epsilon 2}$; W162*, O
	2	15.88	44.71	−35.04	0.34	38.1	H45, $N_{\epsilon 2}$
	3	5.40	27.91	−11.19	0.40	35.8	H182, $N_{\delta 1}$
	4	−3.04	31.58	−1.09	0.15	26.0	M251, S_δ
	5	−4.10	48.52	−13.81	0.26	28.0	H163, $N_{\epsilon 2}$; S120, O_γ
	6	4.61	38.98	−31.34	0.05	5.6	R95, $N_{\eta 1}$, $N_{\eta 2}$
Au	1	−4.52	46.27	−14.22	0.76	68.4	M168, S_δ; S120, O_γ
	2	9.83	42.08	−24.01	1.00	31.2	Y94, O_η; E115, $O_{\epsilon 1}$
	3	25.14	54.45	−16.96	0.12	14.6	
Pt	1	−3.37	31.04	−1.05	0.63	44.7	M251, S_δ; R237*, $N_{\eta 1}$
	2	0.64	49.64	−29.72	0.14	50.1	G81, N; C80, O

[a] $F = \sum AZ\exp[-BS^2]\exp[2\pi i(hx + ky + lz)]$, where A is the fractional occupancy, Z is the atomic number, and $S = 2\sin\theta/\lambda$. Fractional occupancies are reported on an absolute scale arrived at after refinement was complete. [b] Neighbors are those polar atoms within 3.5 Å of the heavy atom site. The symbol * denotes a residue in a symmetry-related molecule.

18.4% against all reflections from 10.0 to 1.8 Å and includes all atoms for residues 1–286 of $E2_{cd}$, 73 bound water molecules, and 1 bound sulfate. Ala^{286} is the last residue for which electron density is seen. The molecular weight calculated for $E2_{cd}$ using Ala^{286} as its C-terminus is 30 410, which agrees well with the 30 000 value estimated by gel electrophoresis (Ghangas & Wilson, 1988). If there are residues extending beyond Ala^{286}, they appear to be disordered.

In order to determine whether any of the ordered waters are in fact calcium ions, we inspected the environment of each water having a temperature factor lower than that of the best ordered protein atom within a 4-Å radius. This analysis revealed no evidence for the presence of calcium ions in the structure. Crystals like the one shown in Figure 1b have been obtained in the absence of calcium ions, indicating that these ions are not required for crystallization.

The overall root-mean-square deviations from ideality for the current model are 0.014 Å in bond lengths, 2.8° in bond angles, 1.1° in fixed improper dihedrals, and 1.4 Å^2 in individual temperature factors between bonded atoms. A Ramachandran plot (data not shown) indicates that almost all residues in the model fall within acceptable regions. The three exceptions are Asp^2, His^{83}, and Trp^{162}. The first two residues are in flexible regions of the protein, as revealed by their higher temperature factors relative to the rest of the protein. Trp^{162}, with (ϕ,ψ) angles of (−146, −105), is well ordered, judging by both its strong electron density (Figure 3b) and its average temperature factor (15 Å^2). Residues having (ϕ,ψ) angles similar to those of Trp^{162} have been observed in other well-refined protein structures (Karplus & Schulz, 1987).

Molecular Fold. $E2_{cd}$ is globular and slightly ellipsoidal (Figure 4a), with approximate dimensions of 53 × 38 × 36 Å^3. It consists of one domain which is an atypical α/β barrel structure, consisting of eight β-strands and eight major and two minor helices (Figure 4b). As expected, the topology is virtually identical to that described for $CBHII_{cd}$ (Rouvinen et al., 1990). Also similar to $CBHII_{cd}$, but unusual for α/β barrel proteins in general (Branden, 1991; Farber & Petsko, 1990), the barrel closure involves only one hydrogen bond, with Gly^{39} in strand I donating an amide proton to the carbonyl oxygen of Trp^{257} at the carboxy-terminal end of strand VII. Strand VIII interacts in an antiparallel fashion with the amino-terminal portion of stand VII (Figure 4b), forming a β-bridge involving two hydrogen bonds between Thr^{244} and Phe^{255}. This eighth strand also contributes the side chain of Ser^{243}, whose

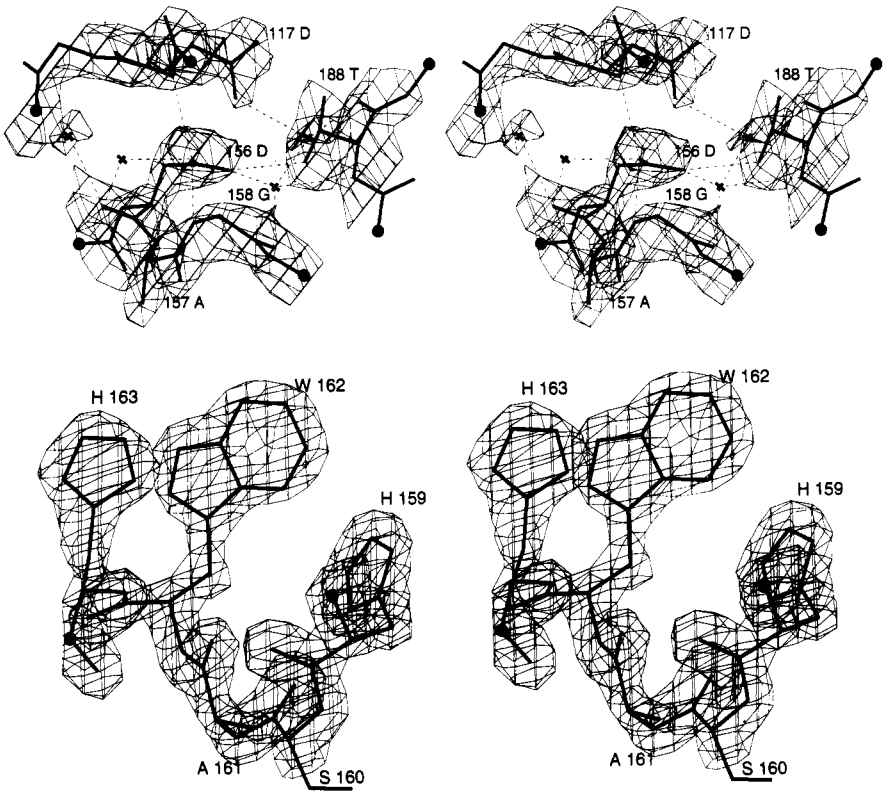

FIGURE 3: (a, top) Quality of the experimental electron density and the environment of Asp[156]. A stereoview of the model refined at 1.8-Å resolution is shown with the SF-MIRAS electron density map calculated from ∞ to 2.6-Å resolution. The contour level is equal to the rms value of the electron density map. Dashed lines indicate potential hydrogen bonds (distances shown are ≤3.2 Å), and crosses represent ordered water molecules. The density for two of the waters is not visible at this contour level but is visible when the contour level is lowered. The putative catalytic general acid, Asp[117], and its buried neighbor, Asp[156], can both be clearly seen. The $O_{\delta 1}$ oxygen of Asp[156] forms hydrogen bonds with a bound water molecule and with the amide nitrogens of Asp[117] and Gly[158]. The unusual geometry of the hydrogen bonds with these two nitrogens should be noted. (b, bottom) Quality of the electron density around Trp[162]. A stereoview of the refined model is shown with a $(2F_o - F_c)\alpha_c$ map calculated from 10.0- to 1.8-Å resolution. The region is well ordered, and the unusual main-chain and side-chain (X_2 = 4.2°) torsion angles of Trp[162] are visible.

Table III: Course of Refinement

model[a]	no. of atoms[b] protein	solvent	no. of unique reflections	resolution range (Å)	$R_{initial}$[c]	R_{final}[c]	Δphase[d] (deg)
I	2148	0	5409	7.0–2.8	0.502	0.241	62
II	2148	0	6844	7.0–2.6	0.349	0.211	53
	2144[e]	0	11427	7.0–2.2	0.257	0.240	51
III	2144	49	11427	7.0–2.2	0.262	0.219	51
	2144	49	15209	7.0–2.0	0.239	0.233	51
IV	2144	72	20704	7.0–1.8	0.265	0.183	49
	2144	72	20951	10.0–1.8	0.210	0.187	48
V	2144	78	20951	10.0–1.8	0.190	0.184	48

[a] Model I, manual fit to 2.8-Å MIRAS map; model II, manual refit to 2.6-Å SF-MIRAS map; model III, manual refit of 2.2-Å-refined model to $(2F_o - F_c)\alpha_c$ map; model IV, manual refit of 2.0-Å-refined model to $(2F_o - F_c)\alpha_c$ map; model V, manual refit of 1.8-Å-refined model to $(2F_o - F_c)\alpha_c$ map. [b] Hydrogen atoms are not included. [c] $\sum|F_o - kF_c|/F_o$. [d] The phase difference between the phases of the refined model and the appropriate experimental phases. [e] Residue Gly[287] was removed from the model at this stage due to the lack of strong density at this position.

O_γ oxygen donates its proton in a hydrogen bond to the carbonyl oxygen of Pro[37] and accepts the main-chain amide proton of Trp[257]. These interactions presumably aid in the closure of the barrel.

Loops at the C-terminal edge of the β-sheet consisting of residues 78–89, 124–127, 227–236, and 264–273 form a cleft approximately 11 Å deep at its deepest point and 10 Å wide, running the length of the molecule (Figure 4a). This cleft is not uniformly deep over its entire length but becomes more open and shallow near both of its ends. Both disulfide bonds in the protein (Cys[80]–Cys[125] and Cys[232]–Cys[267]) are observed in the exposed loops forming the sides of the cleft, one on each side. The electron density calculated at 1.8-Å resolution (data not shown) shows that these loops are reasonably well ordered and that they are involved in crystal contacts. The residues directly involved in these contacts (defined as having atoms within 4.5 Å of atoms from a neighboring molecule) are 78–82, 84, 230–232, 265–267, and 270. Also, residues 77–85 are involved in binding a sulfate molecule. It is clear from an analysis of crystal packing that the cleft itself has no symmetry-related protein atoms within it and is thus available for ligand binding.

Active Site. The active site was expected, by analogy to CBHII, to lie in the cleft of $E2_{cd}$, and difference map analysis at 2.2 Å of isomorphous cocrystals of $E2_{cd}$ with the known inhibitor cellobiose (Ghangas & Wilson, 1988) shows the inhibitor binding in the cleft (Figure 5). Cellobiose is estimated to bind at 50% occupancy, with the nonreducing glucosyl ring bound more tightly than the reducing glucosyl ring (see legend, Figure 5).

In order to determine whether the observed crystal structure represents an active conformation of $E2_{cd}$, native $E2_{cd}$ crystals were assayed for activity using cellotetraose, a soluble substrate consisting of four glucosyl units (see Experimental Procedures). Crystals of $E2_{cd}$ had no greater activity than the supernatant from crystals which had been soaked in storage buffer overnight and were much less active than soluble enzyme under crystal storage conditions (data not shown). This may indicate that the conformation seen in the crystals is not active. However, since soaks of cellobiose did not show binding to the active site, while cocrystallizations did show binding, the lack of activity shown by the crystals may be simply due to the inability of cellotetraose to gain access to the active site of crystalline $E2_{cd}$.

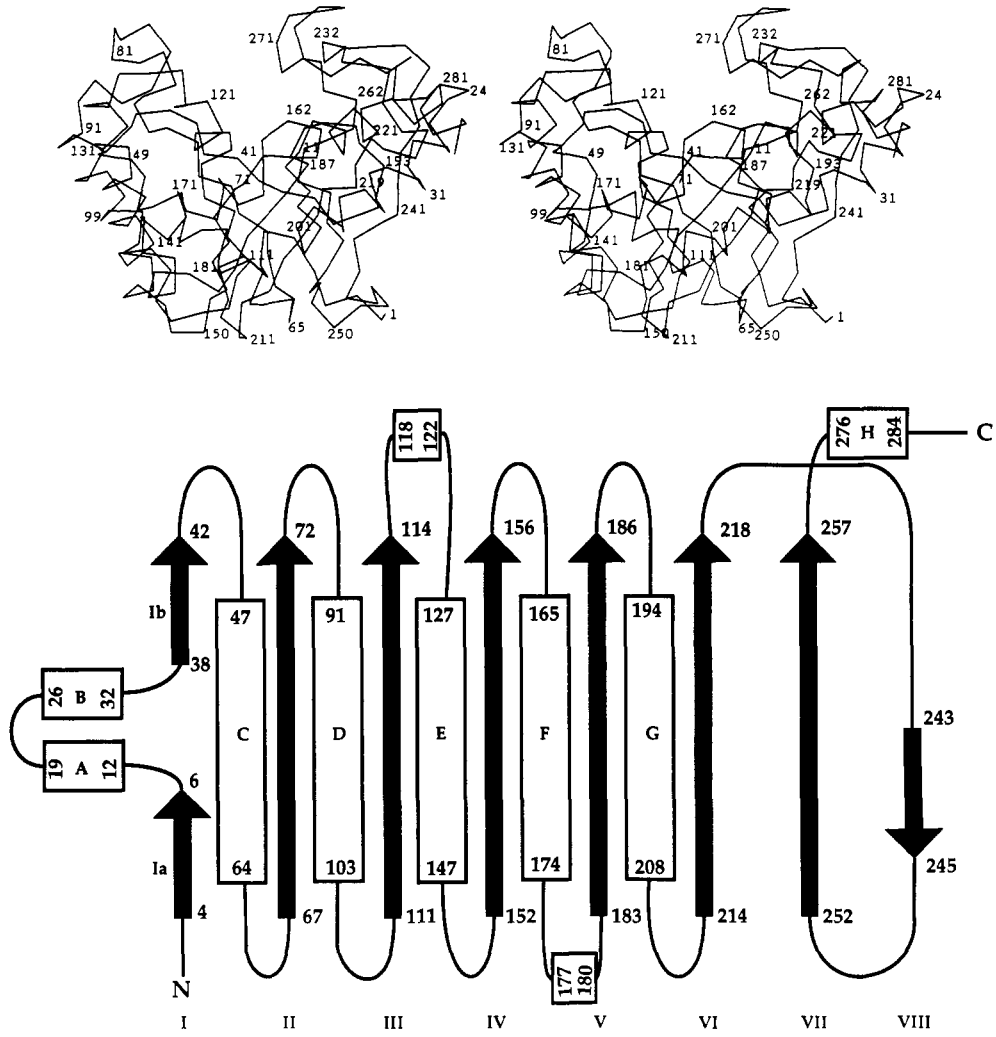

FIGURE 4: Overall fold of $E2_{cd}$. (a, top) Stereoview of the α-carbon backbone of $E2_{cd}$. The active site is located within the cleft at the C-terminal end of the β-barrel. The fold is an atypical α/β barrel. Residues 41 and 162 are tryptophans at each end of the cleft and are most likely involved in binding of the substrate (see Figure 6). (b, bottom) Topology diagram of $E2_{cd}$. Secondary structural elements were assigned on the basis of the criteria described (Kabsch & Sander, 1983), with slight lengthening of the β-stands where β-bridges were identified. Solid arrows denote β-strands and open boxes represent α-helices, and the initial and final residue numbers for each of these structural elements are shown. Strand I is made up of two sections, both interacting with strand II. The barrel closure is formed by a hydrogen bond between Gly^{39} and Trp^{257}. Note that strand VIII interacts in an antiparallel fashion with the N-terminal portion of strand VII, helping to stabilize the atypical barrel.

Residues lining the active site cleft include three tryptophans and an array of polar amino acids (Figure 6). Such a distribution of amino acids is a common feature of carbohydrate binding sites in proteins (Vyas, 1991). The binding cleft is long enough to accommodate at least four glucosyl units, and the side chains of Trp^{41} and Trp^{162} are well exposed at the base of the cleft (Figure 6). In the observed binding mode of cellobiose, the hydrophobic surface of the nonreducing glucosyl unit is packed against the side chain of Trp^{41} (Figure 5), suggesting that Trp^{41} and Trp^{162} may provide broad nonpolar surfaces onto which incoming glucosyl rings of cellulose dock during substrate binding. It should be noted that these two tryptophans are present in all known homologs of E2 (Figure 7). Such an important role for Trp^{162} would help to explain why it is conserved despite its unusual main-chain and side-chain torsion angles (Figure 3b).

Residues Lys^{259} and Glu^{263} participate in hydrogen bonds with hydroxyl groups on the nonreducing end of cellobiose (Figure 5), yet an E263G mutant of E2 shows no detectable loss in activity toward the soluble substrate CMC (Kroupis, 1991). Several residues also interact with the reducing end of the inhibitor. The side chain of Ser^{189} appears to contact the reducing end through a water molecule which is bound upon binding of the ligand (Figure 5). Additionally, the side chains of Asp^{265} and Lys^{259} interact with individual hydroxyl groups on the reducing end.

E2 is known to invert the configuration of the anomeric carbon at the point of cleavage of a cellulose chain (Gebler et al., 1992). One possible mechanism by which this may occur is for an active site base to abstract a proton from water, giving an hydroxide ion which then would cleave the cellulose chain by nucleophilic attack at the anomeric carbon. An active site acid could then protonate the bridging oxygen to stabilize the leaving group. Rouvinen et al. (1990) observed that, in $CBHII_{cd}$, four aspartic acid residues are positioned near the proposed site of cleavage of the cellulose chain: Asp^{175}, Asp^{221}, Asp^{263}, and Asp^{401}. By considering each residue's environment and conservation pattern, as well as some results from mutagenesis experiments, they suggested that either Asp^{263} or Asp^{401} serves as the active site base and that Asp^{221} serves as the active site acid, with Asp^{175} acting to raise the pK_a of Asp^{221}. These four aspartates are all conserved in E2 (Figure 7) and correspond to Asp^{79} (Asp^{175} in CBHII), Asp^{117} (Asp^{221} in CBHII), Asp^{156} (Asp^{263} in CBHII), and Asp^{265} (Asp^{401} in CBHII). The positions of these residues in $E2_{cd}$ and updated amino acid sequence data allow us to further refine the proposed roles of these residues.

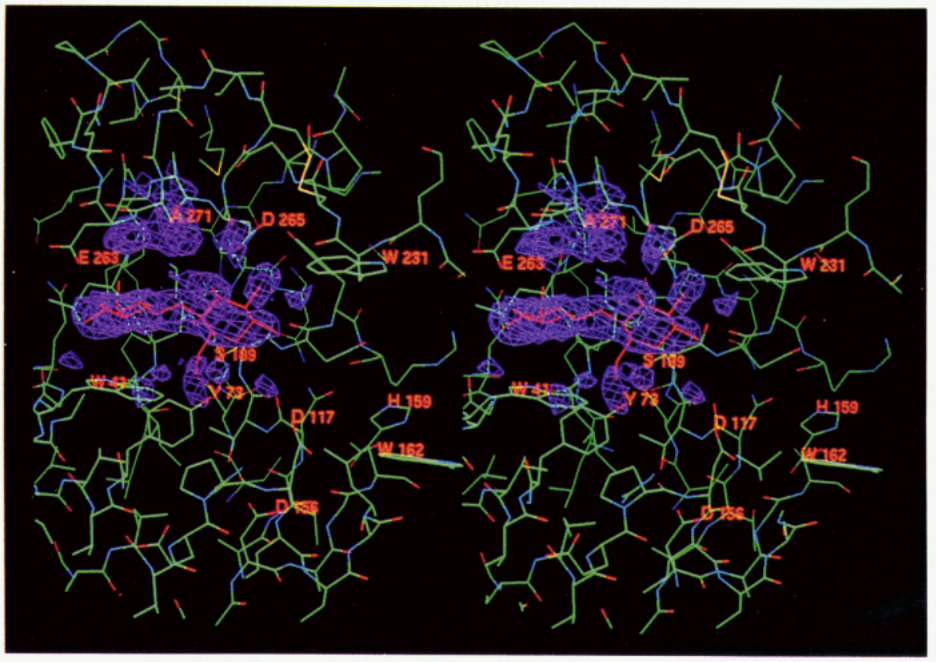

FIGURE 5: Cellobiose (red) bound to the active site. Stereoview looking down on the cleft, showing cellobiose in red fit to an $(F_{o,liganded} - F_{o,nat})\alpha_c$ map calculated from ∞ to 2.2-Å resolution. The long dimension of the cleft is horizontal, and the main-chain atoms are colored by atom type as follows: carbon, green; nitrogen, blue; oxygen, red; sulfur, yellow. Several key residues which interact with the cellobiose or potentially play a role in binding and catalysis (see text) are labeled. The density for the nonreducing end of cellobiose (left end) is stronger than that of the reducing end, suggesting that the nonreducing glucosyl unit is more ordered than the reducing glucosyl unit. This may be due to the packing interactions between the nonreducing end and the side chain of Trp^{41}. Several positive difference peaks are seen near Ala^{271} (in front of Glu^{263}) and Asp^{265}, suggesting a slight conformational change in the protein upon ligand binding. We estimate that cellobiose binds at 50% occupancy because a $(4F_{o,liganded} - 3F_{o,nat})\alpha_c$ map yields cellobiose density on the same level as protein density (Cheetham et al., 1992).

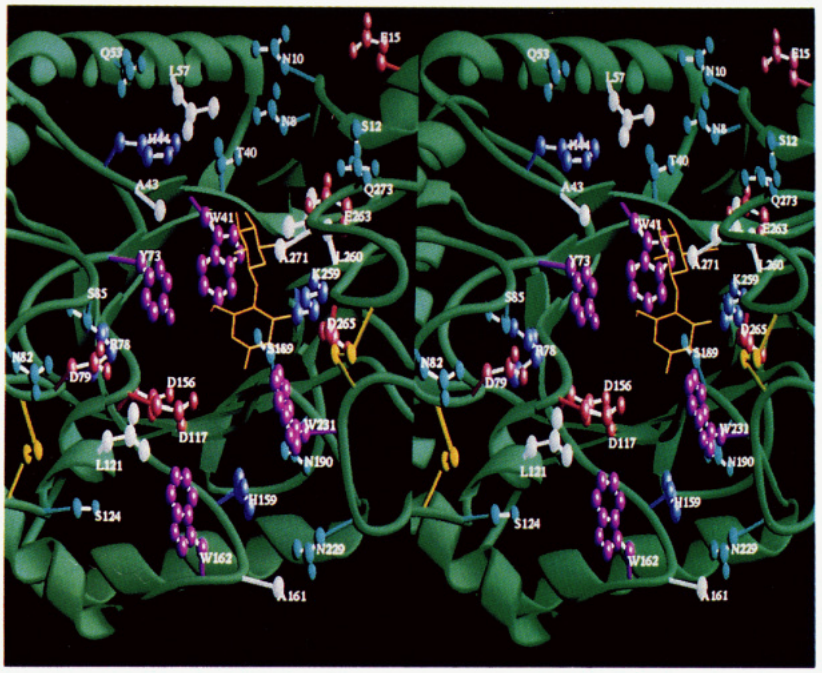

FIGURE 6: Active site residues in $E2_{cd}$. Stereoview looking directly down upon the active site of $E2_{cd}$. The long dimension of the cleft runs vertically. The main-chain ribbon is in green, acidic residues are in red, basic residues are in dark blue, aromatic residues are in purple, polar residues are in aqua, and nonpolar residues are in white. The two disulfide bonds are shown in yellow. Residues depicted were selected by requiring their presence on the "walls" of the cleft and (with the exception of Asp^{156} and the disulfide bonds) by requiring a surface accessibility of at least 10 $Å^2$ by the method of Kabsch and Sander (1983). The side chains of Trp^{41} and Trp^{162} are facing up and are accessible to an incoming cellulose chain. Asp^{117} and Asp^{265} are across the cleft from one another, while Asp^{79} is at the top of the left side of the cleft. Cellobiose (bright yellow) is shown as it binds in the cleft. Display created using RIBBONS (Carson, 1987).

In $E2_{cd}$, Asp^{117}, Asp^{156}, and Asp^{265} are positioned similarly to their counterparts in $CBHII_{cd}$, while Asp^{79} is not. The carboxyl groups of Asp^{117} and Asp^{265} are closest to the proposed cleavage site. They are appropriately positioned across the active site cleft from one other, with their carboxyl oxygens within 5.7 Å of the anomeric carbon on the reducing end of cellobiose, allowing enough space for a water molecule in between (Figure 5). As cellobiose is not a transition-state inhibitor, we do not believe its anomeric carbon is exactly positioned as it would be during catalysis, though we expect that it is near that position. The environment of Asp^{265} appears to favor the ionized form of this residue, because its side chain is actually swung slightly away from the active site (Figures 5 and 6) to form a strong salt bridge with Arg^{221} and weaker

FIGURE 7: Alignment of the sequences of cellulase family B. $E2_{cd}$ shares 67% (one gap), 43% (four gaps), and 41% (nine gaps) sequence identity with CelA from *Microbispora bispora* (Yablonsky et al., 1988), CasA from *Streptomyces* sp. (Gilkes et al., 1991; Nakai et al., 1988), and CenA from *Cellulomonas fimi* (Greenberg et al., 1987), respectively. The sequence of $CBHII_{cd}$ was added on the basis of the structural alignment. Solid boxes denote β-strands, and open boxes represent regions of α-helix [see legend of Figure 4b for assignments in $E2_{cd}$; assignments for $CBHII_{cd}$ were taken from Rouvinen et al. (1990), excluding the assignment of strand VIII]. The gray boxes between the sequences of $CBHII_{cd}$ and $E2_{cd}$ identify regions which are structurally well conserved. These regions are defined as segments in which three or more consecutive C_α positions are $\leq$3.0 Å away from the corresponding C_α positions in the overlaid structure. Such a criterion points out those regions which might be most reliably modeled during homology model building. The circles denote residues in $E2_{cd}$ which are highlighted in Figure 6 (forming the cleft), and closed symbols (circles and squares) identify residues conserved among these five sequences.

interactions with Lys^{259} and Tyr^{191}. Asp^{117}, on the other hand, is more likely to be protonated, as its $O_{\delta 1}$ oxygen is positioned 3.9 Å away from that of Asp^{156}. It should be noted that the $O_{\delta 1}$ oxygen of Asp^{156} is involved in hydrogen bonds with the amide protons of Asp^{117} and Gly^{158}, as well as with a bound water (Figure 3a). These interactions will partially shield Asp^{117} from the charge of Asp^{156}. Because Asp^{156} is completely buried (accessibility = 0 $Å^2$), it is unlikely that it is able to function directly as a catalytic residue.

In $E2_{cd}$, Asp^{79} is well displaced from the catalytic center. It is at the top of one side of the cleft (Figure 6), with its carboxyl group approximately 11.1 Å away from that of Asp^{117} and 11.3 Å away from the anomeric carbon on the reducing end of cellobiose. There is a 4-Å shift in C_α position between Asp^{79} and its equivalent in $CBHII_{cd}$ (Asp^{175}), which may be related to the difference between endo- and exocellulases (see below). Because the loop containing Asp^{79} may move during substrate binding, the observed position in the crystal structure does not rule out its involvement in substrate binding or catalysis. However, the observed residual activity of 20% for the D175A mutant of CBHII (Rouvinen et al., 1990) suggests that it does not play a crucial role. Site-directed mutagenesis of Asp^{79} is planned so that the role of this residue in E2 can be better determined.

The above observations argue for tentatively assigning Asp^{265} as the general base in the reaction and Asp^{117} as the proton donor to the leaving group, with Arg^{221} and Asp^{156}, respectively, playing supporting roles in influencing their pK_a's. A possible argument against the importance of Asp^{265} is that, although the level of sequence similarity (43% identity) between E2 and CasA makes it reasonable to assume that they will operate by the same mechanism, automated sequence alignments of E2 with cellulase CasA do not show conservation of Asp^{265}. CasA does possess two "Asp-Gly" segments near Asp^{265}, but in order to align them, one must place either a single gap (as we have done in Figure 7) or a two-residue insertion in a region well conserved among the other known sequences. To justify this, we note that there are clear changes in CasA in two additional regions conserved among other family members: the loop containing Cys^{232} ($E2_{cd}$ numbering) is deleted, and the sequence of CasA aligned with the C-terminal helix contains two internal prolines (Figure 7), suggesting that the helix is much shorter or not present at all in CasA. Interestingly, the C-terminal helix and the loop near Cys^{232} fold to provide much of the structural context of Asp^{265} in $E2_{cd}$ (Figure 4a). Thus, with their removal, it seems reasonable that a gap or insertion in the region of Asp^{265} could be allowed. In addition, the replacement of Gly^{262} (ϕ,ψ = 106, –25 in $E2_{cd}$) with the β-branched amino acid valine in CasA guarantees that the local conformation of the chain in CasA is different. These arguments, together with the complete conservation of Arg^{221} and Tyr^{191}, which we suggest may interact to position and activate Asp^{265}, make it plausible that Asp^{265} is indeed conserved in CasA.

Endocellulases and Exocellulases. Sequence analysis of the cellulases in family B indicates that they cluster according to their overall endo- or exocellulase function. Since the crystal structure of a representative of each of these classes within family B has now been solved, previous hypotheses based on these sequence alignments may be evaluated.

The two cellulases overlay with 240 core residues, giving an rms deviation of 1.9 Å in α-carbon positions. Their central β-strand regions are structurally highly conserved, with 41 corresponding C_α positions in the eight β-strands giving an rms deviation of 0.67 Å. In a structural alignment of $E2_{cd}$ and $CBHII_{cd}$ (Figure 7), the β-strand regions correspond well, both in placement and in length. More variation is seen in helical regions, which often differ significantly in length. The

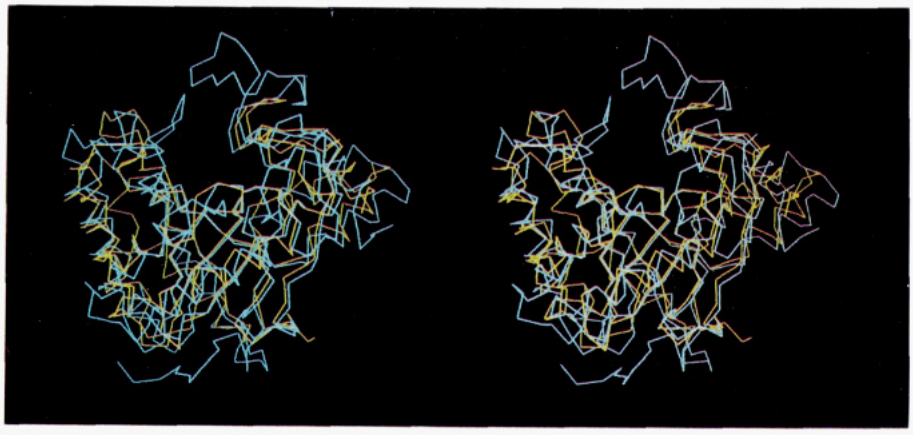

FIGURE 8: Overlay of $E2_{cd}$ (yellow) and $CBHII_{cd}$ (cyan). Stereoview of an overlay of the two structures based on the α-carbon positions of 240 corresponding residues (see Experimental Procedures). Note the cleft in $E2_{cd}$ and the tunnel in $CBHII_{cd}$. The 20-residue loop covering the right side of the tunnel in $CBHII_{cd}$ is deleted in $E2_{cd}$. However, on the top left side of the tunnel in $CBHII_{cd}$ is a loop which is present but adopts another conformation in $E2_{cd}$. These two modifications result in a more open active site in $E2_{cd}$, and the picture illustrates clearly that the major structural difference between the endocellulase (E2) and the exocellulase (CBHII) is the difference in active site enclosure. This evidence supports the hypothesis that the major determinant of endo- or exocellulytic activity is active site accessibility. [Coordinates of $CBHII_{cd}$ were obtained from the Protein Data Bank (Bernstein et al., 1977) using Acquisition Number 3CBH.]

rms difference in C_α position for the 95 corresponding residues in helices is 2.3 Å. This is a typical degree of structural divergence for proteins sharing only 26% sequence identity (Chothia & Lesk, 1986) and may explain why attempts at molecular replacement failed.

Rouvinen et al. (1990) showed that two loops in $CBHII_{cd}$ enclose its active site in a tunnel, and sequence alignments of CBHII and three endocellulases of family B indicated that these loops were lacking in the endocellulases (Rouvinen et al., 1990). They thus proposed that the fundamental difference between exocellulases and endocellulases is the relative openness of their respective active sites, with the two active site enclosing loops in CBHII deleted in the endocellulases.

Comparison of $E2_{cd}$ and $CBHII_{cd}$ does indeed reveal a large difference in active site enclosure, with the active site of $E2_{cd}$ much more open than that of $CBHII_{cd}$ (Figure 8). However, this difference is not due solely to deletions relative to the exocellulase. The shortened loop near residue Cys^{267} in $E2_{cd}$ does correspond to a deletion of 20 residues which cover one side of the active site tunnel in $CBHII_{cd}$ (Figures 7 and 8), but the second loop enclosing $CBHII_{cd}$'s active site is not deleted. It corresponds to residues 76–89 in $E2_{cd}$, and the five-residue deletion occurs after this loop to shorten helix D (Figure 8). The active site cleft in $E2_{cd}$ is made more accessible because this loop is pulled back away from the top of the cleft in $E2_{cd}$ and is involved in binding a sulfate molecule in the crystal. It may be that in solution this loop has the flexibility to close down and contribute to substrate binding, adopting a conformation similar to that seen in $CBHII_{cd}$.

Significant differences between $E2_{cd}$ and CBHII also occur with respect to ligand binding. Cellobiose does not bind at an equivalent position in the two proteins, and while CBHII binds glucose with high affinity, difference Fourier analysis of data from crystals of $E2_{cd}$ soaked in the presence of 0.1 M glucose gives no evidence of binding (data not shown). Previous studies have identified four glucosyl unit binding sites (A, B, C, D) in CBHII (Tilbeurgh et al., 1989). Sites A, C, and D were subsequently seen complexed with ligand in crystals of $CBHII_{cd}$: glucose was observed in site A and cellobiose in sites C and D (Rouvinen et al., 1990). However, in $E2_{cd}$, cellobiose clearly binds to the equivalent of sites A and B. Although the reasons for these differences in ligand binding between $E2_{cd}$ and CBHII are not clear, they may result from differences in the active site environments between the endo- and exocellulase. It is hoped that a detailed comparison of the two enzymes will provide further insight into this question.

The structure of $E2_{cd}$ will aid in the understanding of the other endocellulases in family B. Given the differences in sequence between the endocellulases and CBHII, $E2_{cd}$ should be a better structural model for the endocellulases in the family, while $CBHII_{cd}$ will undoubtedly be a more suitable prototype for exocellulases. This pair of structures, as well as their structural alignment (Figures 7 and 8), will thus allow more accurate homology modeling and better guide future mutagenesis experiments.

ACKNOWLEDGMENT

We thank Kay Diederichs for very helpful discussions and comments, John Brady for assistance in obtaining coordinates for cellobiose, and T. Alwyn Jones for generously providing the polyalanine coordinates of $CBHII_{cd}$.

REFERENCES

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.

Blundell, T. L., & Johnson, L. N. (1976a) in *Protein Crystallography*, pp 151–239, Academic Press Inc., London.

Blundell, T. L., & Johnson, L. N. (1976b) in *Protein Crystallography*, pp 443–464, Academic Press Inc., London.

Branden, C. (1991) *Curr. Opin. Struct. Biol. 1*, 978.

Brunger, A. T. (1990) *Acta Crystallogr. A46*, 46–57.

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

Brunger, A. T., Krukowski, A., & Erikson, J. W. (1990) *Acta Crystallogr. A46*, 585–593.

Calza, R. E., Irwin, D. C., & Wilson, D. B. (1985) *Biochemistry 24*, 7797–7804.

Carson, M. (1987) *J. Mol. Graphics 5*, 103–106.

Cheetham, J. C., Artymiuk, P. J., & Phillips, D. C. (1992) *J. Mol. Biol. 224*, 613–628.

Chirico, W. J., & Brown, R. D. (1985) *Anal. Biochem. 150*, 264–272.

Chothia, C., & Lesk, A. M. (1986) *EMBO J. 5*, 823–826.

Chu, S. S., & Jeffrey, G. A. (1967) *Acta Crystallogr. B24*, 830–838.

Coughlan, M. P. (1992) *Bioresour. Technol. 39*, 107–116.

Devereux, W., Haeberli, P., & Smithies, O. (1992) *Nucleic Acids Res. 12*, 387–395.

Dickerson, R. E., Weinzierl, J. E., & Palmer, R. A. (1968) *Acta Crystallogr. B24*, 997.
Farber, G. K., & Petsko, G. A. (1990) *Trends Biochem. Sci. 15*, 228.
Gebler, J., Gilkes, N. R., Claeyssens, M., Wilson, D. B., Beguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller, R. C., Jr., Warren, R. A., & Withers, S. G. (1992) *J. Biol. Chem. 267*, 12559–12561.
Ghangas, G. S., & Wilson, D. B. (1988) *Appl. Enviro. Microbiol. 54*, 2521–2526.
Gilkes, N. R., Claeyssens, M., Aebersold, R., Henrissat, B., Meinke, A., Morrison, H. D., Kilburn, D. G., Warren, R. A., & Miller, R. C., Jr., (1991) *Eur. J. Biochem. 202*, 367–377.
Greenberg, N. M., Warren, R. A., Kilburn, D. G., & Miller, R. C. (1987) *J. Bacteriol. 169*, 646–653.
Hamlin, R. (1985) *Methods Enzymol. 114*, 416–452.
Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., & Mornon, J.-P. (1989) *Gene 81*, 83–95.
Howard, A. J., Nielsen, C., & Xuong, N. H. (1985) *Methods Enzymol. 114*, 452–472.
Irwin, D., Spezio, M. L., Walker, L., & Wilson, D. B. (1993) *Biotechnol. Bioeng.* (in press).
Jones, T. A. (1985) *Methods Enzymol. 115*, 157–171.
Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssens, M., Beguin, P., & Aubert, J.-P. (1992) *Nature 357*, 89–91.
Kabsch, W. (1976) *Acta Crystallogr. A32*, 922.
Kabsch, W. (1978) *Acta Crystallogr. A34*, 826.
Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.
Karplus, P. A., & Schulz, G. E. (1987) *J. Mol. Biol. 195*, 701–729.
Karplus, P. A., Daniels, M. J., & Herriott, J. R. (1990) *Science 251*, 60–66.
Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., & Gronenborn, A. M. (1989) *Biochemistry 28*, 7241–7257.
Krebs, K. G., Heusser, D., & Wimmer, H. (1969) in *Thin-Layer Chromatography* (Stahl, E., Ed.) pp 854–909, Springer-Verlag, Berlin.
Kroupis, C. (1991) Master's Thesis, Cornell University.
Lao, G., Ghangas, G. S., Jung, E. D., & Wilson, D. B. (1991) *J. Bacteriol. 173*, 3397–3407.
Lynd, L. R., Cushman, J. H., Nichols, R. J., & Wyman, C. E. (1991) *Science 251*, 1318–1323.
Matthews, B. W. (1966) *Acta Crystallogr. 20*, 82–86.
Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491–497.
McIlvaine, T. C. (1921) *J. Biol. Chem. 49*, 183–186.
McLachlan, A. D. (1971) *J. Mol. Biol. 61*, 409–424.
McPherson, A. (1989) in *Preparation and analysis of protein crystals*, pp 82–159, Robert Ekrieger Publishing Co., Malabar, FL.
Nakai, R., Horinouchi, S., & Beppu, T. (1988) *Gene 65*, 229–238.
Reeck, G. R., Jackson, P. J., & Teller, D. C. (1982) *Nature 300*, 76–78.
Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., & Jones, T. A. (1990) *Science 249*, 380–385.
Sack, J. S. (1988) *J. Mol. Graphics 6*, 224–225.
Tilbeurgh, H. v., Lootiens, F. G., Engelborgs, Y., & Claeyssens, M. (1989) *Eur. J. Biochem. 184*, 553–559.
Tribe, H. T. (1966) *Trans. Br. Mycol. Soc. 49*, 457–466.
Vyas, N. D. (1991) *Curr. Opin. Struct. Biol. 1*, 732–740.
Walker, L. P., Wilson, D. B., Irwin, D. C., McQuire, C., & Price, M. (1992) *Biotechnol. Bioeng. 40*, 1019–1026.
Wang, B.-C. (1985) *Methods Enzymol. 115*, 90–117.
Wilson, D. B. (1988) *Methods Enzymol. 160*, 314–323.
Xuong, N. H., Nielsen, C., Hamlin, R., & Anderson, D. (1985) *J. Appl. Crystallogr. 18*, 342–350.
Yablonsky, M. D., Elliston, K. O., & Eveleigh, D. E. (1988) in *Enzyme Systems for Lignocellulose Degradation* (Coughlan, M. P., Ed.) Elsevier Applied Science, New York.