Structure of the Endoglucanase I from *Fusarium oxysporum*: Native, Cellobiose, and 3,4-Epoxybutyl β-D-Cellobioside-Inhibited Forms, at 2.3 Å Resolution[†,‡]

Gerlind Sulzenbacher,[§] Martin Schülein,[||] and Gideon J. Davies*[,§]

Department of Chemistry, University of York, Heslington, York YO1 5DD, U.K., and Novo Nordisk A/S, Novo Allé, DK 2880 Bagsvaerd, Denmark

Received December 3, 1996; Revised Manuscript Received February 13, 1997[⊗]

ABSTRACT: The mechanisms involved in the enzymatic degradation of cellulose are of great ecological and commercial importance. The breakdown of cellulose by fungal species is performed by a consortium of free enzymes, known as cellobiohydrolases and endoglucanases, which are found in many of the 57 glycosyl hydrolase families. The structure of the endoglucanase I (EG I), found in glycosyl hydrolase family 7, from the thermophilic fungus *Fusarium oxysporum* has been solved at 2.3 Å resolution. In addition to the native enzyme, structures have also been determined with both the affinity label, 3,4-epoxybutyl β-D-cellobioside, and the reaction product cellobiose. The affinity label is covalently bound, as expected, to the catalytic nucleophile, Glu197, with clear evidence for binding of both the *R* and *S* stereoisomers. Cellobiose is found bound to the −2 and −1 subsites of the enzyme. In marked contrast to the structure of EG I with a nonhydrolyzable thiosaccharide analog, which spanned the −2, −1, and +1 subsites and which had a skew-boat conformation for the −1 subsite sugar [Sulzenbacher, G., et al. (1996) *Biochemistry 35*, 15280−15287], the cellobiose complex shows no pyranoside ring distortion in the −1 subsite, implying that strain is induced primarily by the additional +1 subsite interactions and that the product is found, as expected, in its unstrained conformation.

Cellulases catalyze the hydrolysis of the β-1,4-glycosidic linkages in cellulose, the most abundant biopolymer. In nature, turnover of cellulose is provided by a mélange of cellulolytic enzymes produced by a variety of microorganisms. Fungal systems tend to produce a consortium of free enzymes, while some bacterial species secrete a highly active multi-protein complex named the cellulosome. Most individual cellulases display a modular structure in which a catalytic core domain is linked to one or more nonhydrolytic domains (Gilkes et al., 1991). Cellulose-based products, mainly in the form of paper, constitute the great majority of municipal waste, and applications of cellulases in the conversion of waste biomass to foodstuffs have been proposed. Current commercial applications of cellulases are mainly limited to the detergent and textile industries, but present environmental concerns have refueled interest in the application of cellulases in biomass conversion. All current and potential industrial applications of these enzymes are frustrated by a lack of the detailed understanding of these enzymes and their interactions with substrates.

The catalytic core domains of cellulases are found in approximately 11 of the 57 families of glycosyl hydrolase (Henrissat, 1991; Henrissat & Bairoch, 1993, 1996). Structures are known for representatives of 21 of the 57 glycosyl hydrolase families (Davies & Henrissat, 1995). The *Fusarium oxysporum* endoglucanase I (EG I)[1] is found in glycosyl hydrolase family 7. Structurally, EG I has the jelly-roll topology as first observed in the plant legume lectins such as concanavalin A and now widely recognized as a ubiquitous fold (Edelman et al., 1972; Hardman & Ainsworth, 1972). Two native family 7 EG I structures have been described, those from *Humicola insolens* and *Trichoderma reesei* (Davies et al., 1997a). The *F. oxysporum* EG I bears most similarity to the *H. insolens* enzyme with which it shares approximately 57% sequence identity (Sheppard et al., 1994). Both the *H. insolens* and *F. oxysporum* enzymes are unusual in that they consist of a catalytic core domain only and have no cellulose binding domains. Family 7 of the glycosyl hydrolases also contains enzymes classified as cellobiohydrolases, which differ from the endoglucanases in that their active sites are enclosed within a tunnel. Indeed, it was a cellobiohydrolase, the CBH I from *T. reesei*, that was the first enzyme from this family to have its three-dimensional structure elucidated (Divne et al., 1994). EG I also shows similarity to the family 16 glycosyl hydrolases such as the *Bacillus* 1,3-1,4-glucanase (Keitel et al., 1993) with not only a structural relationship between these two families of enzymes but also a strict conservation of the catalytic residues and catalytic mechanism.

[†] This work was funded in part by the Biotechnology and Biological Sciences Research Council, Novo-Nordisk A/S, and the European Union (Contract BIO2-CT94-3018). G.J.D. is a Royal Society University Research Fellow.

[‡] Coordinates for the structure described in this paper have been deposited with the Brookhaven Protein Data Bank (accession references 1OVW, 2OVW, 3OVW, and 4OVW).

* Corresponding author. Telephone: -44-1904-432596. Fax: -44-1904-410519. E-mail: davies@yorvic.york.ac.uk.

[§] University of York.

[||] Novo Nordisk A/S.

[⊗] Abstract published in *Advance ACS Abstracts,* April 1, 1997.

[1] Abbreviations: CBH, cellobiohydrolase; CMC, carboxymethyl-substituted cellulose; EG, endoglucanase; MOPS, 3-(*N*-morpholino)-propanesulfonic acid; NAM, *N*-acetylmuramic acid; NAG, *N*-acetylglucosamine; NCS, noncrystallographic symmetry; PEG, polyethylene glycol; thio-DP5, methyl (*S*)-β-D-glucopyranosyl-(1→4)-(*S*)-4-thio-β-D-glucopyranosyl-(1→4)-(*S*)-4-thio-β-D-glucopyranosyl-(1→4)-(*S*)-4-thio-β-D-glucopyranosyl-(1→4)-4-thio-α-D-glucopyranoside.

S0006-2960(96)02963-7 CCC: $14.00

EG I catalyzes hydrolysis of the β-1,4-glycosidic linkages of cellulose with a net retention of the anomeric configuration at C-1 (Schou et al., 1993). This occurs *via* a double displacement reaction as described by Koshland (1953). Such a mechanism requires the presence of two functional groups: an acid/base which first functions as a Brønsted acid and thus protonates the glycosidic bond to promote aglycone departure and a nucleophile which assists in the departure of this leaving group concomitant with the formation of the glycosyl−enzyme intermediate. Following formation of this intermediate, the acid/base then functions as a Brønsted base and activates, by deprotonation, an incoming water molecule (or alternate nucleophile) for nucleophilic attack. Both these glycosylation and deglycosylation reactions proceed *via* transition states with substantial oxocarbenium ion character as indicated by secondary deuterium kinetic isotope effects ($k_H/k_D > 1$) (McCarter & Withers, 1994; Sinnott, 1990). In the case of the *F. oxysporum* EG I, the acid/base and nucleophile are Glu202 and Glu197, respectively. Recent work on the structure of the *F. oxysporum* EG I with a nonhydrolyzable thiooligosaccharide analog has established the role of substrate distortion, to yield an axial leaving group orientation, in catalysis by this enzyme (Sulzenbacher et al., 1996).

Kinetic studies have shown that EG I has four subsites for sugar binding labeled −2 to +2, with cleavage, by definition, taking place between the −1 and +1 subsites (Figure 1A) (Davies et al., 1997b). On small soluble cellodextrins, EG I preferentially liberates cellobiose as the leaving group (Schou et al., 1993), indicative of two aglycone subsites contributing to catalysis, as is also observed on all other cellulase systems. The family 7 enzymes, such as EG I, are also slightly unusual in that under certain circumstances they may liberate glucose as a leaving group (Biely et al., 1991; Schou et al., 1993). Kinetic studies are further complicated because at high saccharide concentrations EG I performs transglycosylation reactions to synthesise higher oligosaccharides. This transglycosylating ability has led to the employment of EG I in the chemoenzymatic synthesis of "designer" oligosaccharides, such as bifluorescently labeled cellodextrins (Armand et al., 1997).

In this paper, we describe the structure of the native form of EG I together with its complexes with the covalent inhibitor 3,4-epoxybutyl β-D-cellobioside (Figure 1B) and the natural reaction product β-D-cellobiose (Figure 1C). The epoxide reagent identifies the enzymatic nucleophile as Glu197, as inferred from other complexes (Sulzenbacher et al., 1996) and work on similar structures (Divne et al., 1994; Keitel et al., 1993) and as labeled unambiguously with 2-fluoro-2-deoxycellobiosides [described in the preceding paper by MacKenzie et al. (1997)]. In addition, comparisons between the EG I complex with cellobiose and that with the nonhydrolyzable thiosaccharide shed light on the interactions that contribute to substrate binding and catalysis by this family of glycosyl hydrolase.

MATERIALS AND METHODS

Crystallization, Data Collection, and Processing for Native EG I. The gene encoding the EG I from *F. oxysporum* has been cloned (Sheppard et al., 1994). EG I was secreted by an *Aspergillus oryzae* expression system, as described previously (Christensen et al., 1988). The protein was

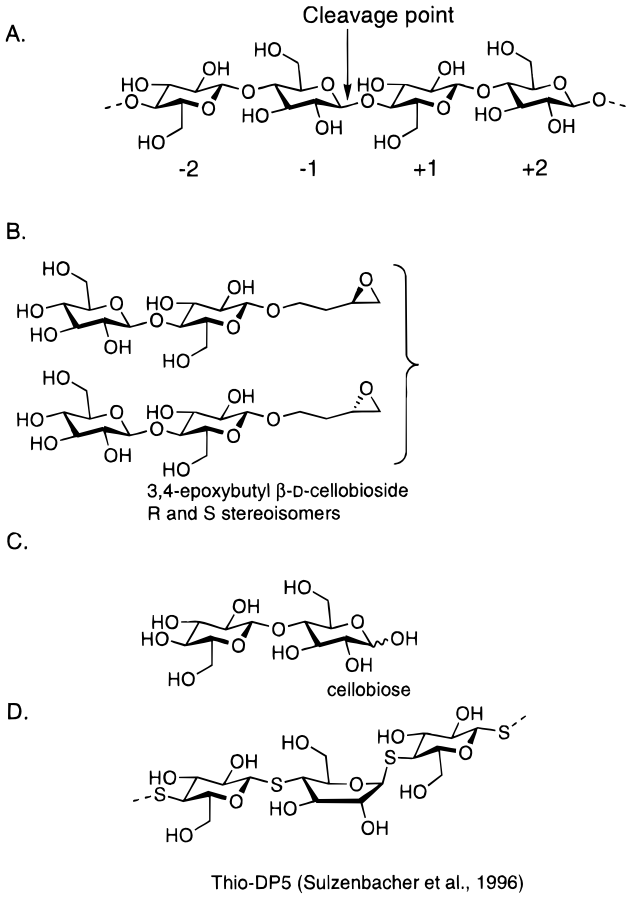

FIGURE 1: (A) Diagram showing the subsite specificity of *F. oxysporum* EG I. There are four significant subsites for saccharide binding. (B) Chemical structure of the inhibitor 3,4-epoxybutyl β-D-cellobioside, which is synthesised as a mixture of *R* and *S* stereoisomers as indicated. (C) Chemical structure of β-D-cellobiose. (D) Chemical structure of the crystallographically observed part of the nonhydrolyzable substrate analog, thio-DP5, as published previously (Sulzenbacher et al., 1996). Panels B−D are drawn to show the approximate subsite location of the compounds as reported in this study.

initially purified by ion exchange chromatography, deglycosylated by treatment with endoglycosidase F, and then finally purified by hydrophobic interaction chromatography on phenyl-Sepharose. Samples of native *F. oxysporum* EG I were concentrated and desalted on Amicon microconcentrators to a concentration of 10 mg mL^{-1}. Crystals were grown in hanging drops from solutions containing 0.2 M $MgCl_2$, 0.1 M MOPS at pH 6.5, and 10−20% (w/v) polyethylene glycol 8000. Crystals grow to a maximum size of 0.3 × 0.1 × 0.04 mm within 7 days at 16 °C. They are in monoclinic space group $P2_1$ with cell dimensions a = 57.4 Å, b = 81.9 Å, c = 91.0 Å, and β = 105.97° and with two molecules of EG I in the asymmetric unit. Data were collected from a single crystal mounted in a glass capillary using an RAXIS II imaging plate and Cu rotating anode. Data (116°) were collected in 1° frames from a sample mounted with the crystallographic c^* axis roughly parallel to the spindle axis of the camera, but missset slightly to reduce loss of data in the "blind region". Data were processed with the DENZO program (Z. Otwinokski, unpublished) and further reduced with programs from the CCP4 suite of programs (Collaborative Computational Project Number 4, 1994).

Table 1: Data Quality and Completeness, Given in Resolution Bins, for the Native *F. oxysporum* Endoglucanase I and the Complexes with the Inhibitor 3,4-Epoxybutyl β-D-Cellobioside and with the Product β-D-cellobiose

	native		epoxide inhibited		cellobiose	
resolution (Å)	R_{merge}[a]	completeness (%)	R_{merge}	completeness %)	R_{merge}	completeness (%)
15.0−4.9	0.048	98	0.041	99	0.049	89
3.9	0.055	98	0.050	98	0.052	93
3.4	0.063	97	0.061	97	0.062	95
3.1	0.076	97	0.072	97	0.075	95
2.9	0.097	96	0.088	96	0.110	96
2.7	0.122	96	0.107	96	0.141	96
2.6	0.134	95	0.124	95	0.181	97
2.5	0.162	95	0.147	88	0.213	95
2.4	0.184	73	0.175	63	0.245	76
2.3	0.185	51	0.201	47	0.274	54
totals	0.072	91	0.071	88	0.077	89

[a] $R_{merge} = \sum_{hkl}\sum_i|I_{hkli} - \langle I_{hkl}\rangle|/\sum_{hkl}\sum_i\langle I_{hkl}\rangle$

Complex Structures. *F. oxysporum* EG I was inhibited as follows. An aqueous solution of EG I was concentrated to 9.5 mg/mL (equivalent to 0.19 mmol), in 20 mM phosphate buffer at pH 7.5. A total of 2 mL was then incubated with a fresh solution of 8.25 mmol of 3,4-epoxybutyl β-D-cellobioside for 3 h at 40 °C. The enzyme was then deglycosylated and repurified, as described above for the native enzyme. Crystals of the inhibited form were grown from protein solutions buffered with 100 mM Tris-HCl at pH 7.7 using 16−22% monomethyl ether PEG 2000 as the precipitant. The crystals are in space group $P2_1$ with cell dimensions $a = 65.9$ Å, $b = 82.6$ Å, $c = 73.2$ Å, and $\beta = 94.2°$, and there are two molecules of EG I in the asymmetric unit. Crystals were harvested into a cryoprotectant mother liquor containing 25% (w/v) monomethyl ether PEG 2000, 100 mM Tris-HCl at pH 7.7, and 10% glycerol. Crystals were stable in this cryoprotectant for no more than 5 s, and so a single crystal was immediately mounted in a rayon fiber loop and placed in a boiling nitrogen stream at 100 K. Data were collected using an RAXIS II imaging plate system with a Cu rotating anode and utilizing long, focusing, mirror optics (Yale/Molecular Structure Corp.).

Crystals of the cellobiose complex were grown similarly, but using 20−26% (w/v) PEG 8000 as the precipitant with 0.15−0.3 M $MgCl_2$ and 0.1 M MOPS at pH 6.4. Crystals were harvested into a stabilizing solution containing 24% (w/v) PEG 8000, 0.2 M $MgCl_2$, and 0.1 M MOPS at pH 6.5, and with the addition of 20 mM cellobiose. Crystals were then transferred to a cryoprotectant mother liquor [24% (w/v) PEG 8000, 0.2 M $MgCl_2$, 0.1 M MOPS at pH 6.5, 20 mM cellobiose, and 22% (w/v) glycerol] and data collected from a single crystal at 120 K essentially as described above. The crystals are in space group $P2_1$ with unit cell dimensions $a = 67.8$ Å, $b = 78.3$ Å, $c = 142.5$ Å, and $\beta = 97.0°$, and there are four molecules of EG I in the asymmetric unit.

Structure Solution and Refinement. Molecular replacement calculations for the native enzyme structure were performed with the program AMoRe (Navaza, 1994) using the A molecule (protein atoms only) of the refined *H. insolens* EG I as the search model (Davies et al., 1997a). An outer radius of Patterson integration of 20 Å was used together with all data between 20 and 4 Å resolution. The translation search was performed with data between 10 and 4 Å resolution. Electron density maps were calculated and averaged according to the noncrystallographic symmetry using the program RAVE (Kleywegt & Jones, 1994). The initial averaged electron density map was of high quality and permitted correction of all the errors associated with the differences between the *F. oxysporum* and *H. insolens* (search model) sequences. This, and subsequent, manual rebuilding of the model was performed using the O program (Jones et al., 1991). Five percent of the observations were set aside for cross validation analysis (Brünger, 1992) and used to monitor various refinement strategies such as geometric, temperature-factor, and NCS restraint values and the insertion of solvent water. Refinement was initially performed with a single molecule of EG I strictly constrained by NCS. A standard XPLOR slow cooling protocol was used with a starting temperature of 3000 K (Brünger et al., 1987). Further cycles of manual rebuilding were interspersed with standard least-squares refinement, with tight NCS restraints, using the maximum-likelihood program REFMAC (Murshudov et al., 1997). As all observed data were employed in the refinement, a low-resolution bulk solvent correction was applied, as implemented in the REFMAC program. The behavior of R_{free} was closely monitored and suggested that tight NCS restraints for both positional and temperature factors were appropriate. Water molecules were initially built into difference density maps also averaged according to the NCS. The epoxybutyl cellobioside and cellobiose complexes were solved by molecular replacement using the program AMoRe in a manner equivalent to that described above for the native structure solution except that the refined A molecule of the native *F. oxysporum* EG I was employed as the initial search model. Refinement of these complexes was performed in a manner similar to that described for the native enzyme with the stereochemical dictionaries for the ligands based upon the coordinates of β-D-cellotetraose (Raymond et al., 1995). Coordinates for the structures described in this work have been deposited with the Brookhaven Protein Data Bank with the generic reference OVW (Bernstein et al., 1977).

RESULTS AND DISCUSSION

A summary of the data quality and completeness, described in this study, is given in Table 1. The native EG I data consist of 78 153 observations of 32 773 unique reflections with 886 observations (constituting 1.1% of the observations) rejected during the data reduction procedure. A large nonorigin Patterson peak at $^1/_2$, 0, $^1/_2$ (with a peak height corresponding to 42% of the origin at 4.5 Å resolution) indicated that the two molecules of EG I in the asymmetric unit were related by predominantly translational symmetry (Figure 2). The averaged electron density maps were of high

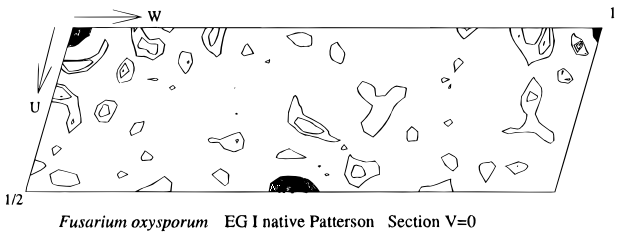

FIGURE 2: Native Patterson, section $V = 0$, for the native crystal form of EG I calculated at 4.5 Å resolution. The nonorigin peak at $U = {}^1/_2$, $V = 0$, $W = {}^1/_2$ (peak height 42% of the origin) is indicative of two similarly oriented molecules of EG I in the asymmetric unit related by a translation vector.

Table 2: Refinement and Structure Quality Statistics for the EG I from *F. oxysporum*

	native	epoxide	cellobiose
no. of molecules per asymmetric unit	2	2	4
no. of protein atoms[a]	6070	6060	12092
no. of ligand/*N*-linked glycosylation atoms	0/56	46/28	92/112
no. of solvent waters	423	700	1104
resolution used in refinement (Å)	15−2.3	15−2.3	15−2.3
R_{cryst}	0.17	0.18	0.20
R_{free}	0.22	0.27	0.28
1−2 bonds (Å)	0.008	0.010	0.013
1−3 angles (Å)	0.031	0.034	0.037
chiral volumes (Å^3)	0.107	0.120	0.126
average main chain *B* (Å^2)	21	17	27
average side chain *B* (Å^2)	25	19	30
mean *B* ligand (Å^2)	N/A	21	33
main chain Δ*B*, bonded atoms (Å^2)	1.5	1.5	1.8
NCS positional deviation, main chain (Å)	0.036	0.046	0.020

[a] Apparent discrepancies in the number of atoms per protein molecule result from (for epoxide) modelling of part of the side chain of Glu197 as "inhibitor" and (for cellobiose) differing C-terminal disorder in the different crystal forms.

quality and permitted continuous tracing of amino acids 1−400. Protein atoms were only modeled for the first 400 amino acids of the sequence; the 11 C-terminal residues were presumed to be either disordered or absent in the crystal. The first 18 amino acids at the N terminus were not present, confirming that these residues form a leader peptide as implied from the sequence (Sheppard et al., 1994).

The final model structure has a crystallographic *R* factor of 0.17 with an R_{free} of 0.22. All the non-glycine residues have conformational angles (φ, ψ) in permitted regions of the Ramachandran plot (Ramachandran et al., 1963) with 0.4% of these in the "generously allowed regions" as defined by PROCHECK (Laskowski et al., 1993). A plot of *R* factor against resolution or the σ_A method gives an upper estimate for the mean coordinate error of approximately 0.18 Å (Read, 1986). The average main chain and side chain *B* values are approximately 21 and 24 Å^2, respectively, for both molecules within the asymmetric unit. The behavior of R_{free} for different refinement strategies suggested that very tight positional and temperature factor NCS restraints, for both native and complex structures, were appropriate. This is reflected in the very small rms difference in the position of main and side chain atoms between the independent molecules in the asymmetric unit (Table 2). In contrast to the native crystal form, all EG I complexes crystallize in highly sensitive and fragile crystal forms which do not diffract sufficiently well for X-ray structural determination at room temperature. For this reason, all complex structure determinations utilized cryocrystallographic techniques. Refinement of the complex structures was performed in a manner equivalent to that for the native enzyme. Final refinement statistics for all three structures are given in Table 2.

Native Enzyme Structure. EG I has the jellyroll topology, first observed in the structures of the plant legume lectins, a concave sheet of seven antiparallel β-strands, above a convex sheet of eight such strands. The concave sheet forms the sides and base of a substrate binding groove which is approximately 50 Å long (Figure 3). The structure is stabilized by the presence of nine disulfide bridges between residues 18 and 24, 48 and 70, 60 and 66, 138 and 365, 172 and 195, 176 and 194, 215 and 234, 223 and 228 and 239 and 315. The N-terminal glutamine residue was observed to be modified to form the cyclic pyroglutamate moiety observed in the structures of other enzymes from this family. The substrate binding groove contains a number of acidic residues, located deeply at the bottom of this canyon, which play key roles in catalysis. Glu202, the enzymatic acid/base sits approximately 5.5 Å away from the nucleophile Glu197, consistent with the spatial separation of these residues in a retaining enzyme mechanism (McCarter & Withers, 1994; Wang et al., 1994; White et al., 1994). The nucleophile, Glu197, hydrogen bonds to an adjacent aspartate residue, Asp199. This direct interaction demands that, prior to a catalytic cycle, one of these two residues is present in a protonated form, and the likely reaction mechanism dictates that this is to be Asp199 with the negatively charged Glu197 poised for nucleophilic attack on the anomeric C-1 atom of the substrate (Figure 4A). Further details on the nature of the active site and the interactions therein will be presented following description of the two complex structures.

Substrate Binding Cleft and Processivity in Cellulases. Family 7 of the glycosyl hydrolase classification contains enzymes traditionally classified as cellobiohydrolases (EC 3.2.1.91) as well as endoglucanases (EC 3.2.1.4). Although this distinction reflects the intrinsic ability of cellobiohydrolases to hydrolyze intact crystalline cellulose, a substrate which is unresponsive to attack by endoglucanases, it is current practice to distinguish between endoglucanases and cellobiohydrolases by the ability of "endoglucanases" to hydrolyze carboxymethylcellulose (CMC), a substrate somewhat more resistant to catalytic attack by the "cellobiohydrolases". In truth, this artifactual distinction on an artificial substrate merely reflects the inability of some enzymes' subsites to tolerate the carboxymethyl substituents present in CMC (which is clearly a greater problem for an enzyme with a much larger number of subsites). Furthermore, this distinction has led to the persistence of the proposal that cellobiohydrolases are "exoenzymes", a definition which would demand that they display a kinetic requirement for the chain ends of cellulose, which is manifestly not the case (Claeyssens et al., 1989; Armand et al., 1997).

We, and others, have proposed that these artificial classifications and distinctions could profitably be replaced with the simple concept of processivity, as is widely accepted in other families of glycosyl hydrolase and which is simply explained by the active site topologies (Davies & Henrissat, 1995; Henrissat, 1994). Enzymes with an open active site

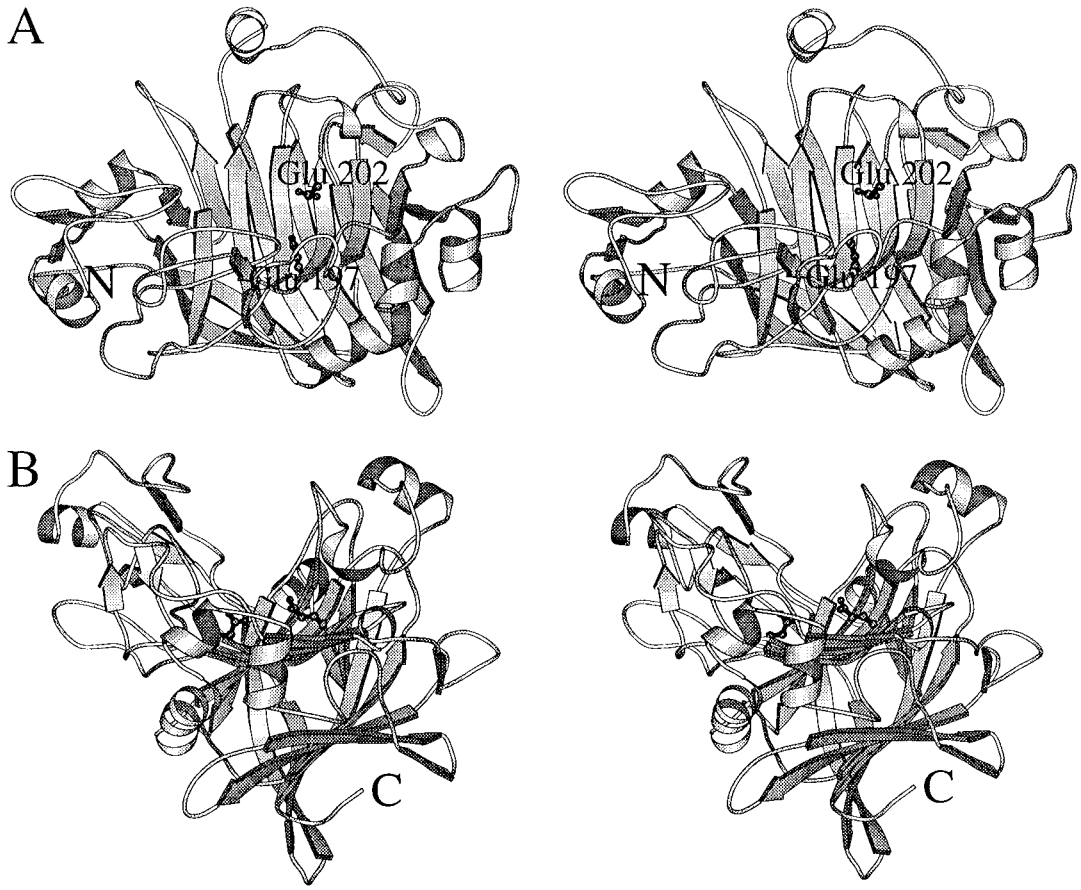

FIGURE 3: Stereo (wall-eyed) MOLSCRIPT (Kraulis, 1991) diagrams of the EG I from *F. oxysporum.* (A) Side-on view showing the open active site canyon formed by the β-strands. The catalytic acid/base Glu202 and the catalytic nucleophile, Glu197, are shown in "ball-and-stick" representations and the N terminus is indicated. (B) View down the active site canyon. Glu197 is on the right and Glu202 on the left. The C terminus is indicated.

cleft and a small number of sugar binding subsites are unlikely to retain the polysaccharide chain following a catalytic event. Consequently, on intact crystalline substrates, the displaced polysaccharide chain could then readhere to the cellulose crystal and prevent further hydrolysis. This is reflected in the inability of these enzymes to hydrolyze crystalline substrates. Conversely, those enzymes possessing an extended active site, with many sugar binding subsites enclosed by a tunnel, would be much less likely to release the polysaccharide chain after a catalytic event, as was initially proposed by Jones and colleagues (Rouvinen et al., 1990). Thus, the tunnel and extended substrate binding interactions create the necessary conditions for a processive degradation of the crystalline substrate and account for the ability of CBH I to hydrolyze crystalline substrates. Schematic representations of the processivity concept, as applied to cellulases, have been described (Davies & Henrissat, 1995; Tomme et al., 1996).

All three family 7 EG I structures have their active sites enclosed in a long open canyon formed by the concave curvature of the "upper" β-sheet, as described above. Although the substrate binding groove has potential interaction sites for many glucose units, these enzymes display only four kinetically significant subsites for saccharide (Biely et al., 1991; Schou et al., 1993; Schülein, 1997). The cellobiohydrolases from this family have an overall fold similar to and catalytic machinery identical to that of EG I (Divne et al., 1994). Extended loop regions, adjacent to the active site of CBH I, enclose much of the substrate binding groove in a form of tunnel. In addition, CBH I displays a kinetic preference for much longer oligosaccharides (at least seven subsites, but the insolubility of higher cellodextrins prevents determination of the exact number), and recent preliminary crystallographic reports suggest 10 or 11 subsites (Divne et al., 1996). Together, these two structural differences are reflected in the ability of CBH I to hydrolyze intact crystalline cellulose, a substrate which is comparatively inert to attack by EG I.

Structure of the 3,4-Epoxybutyl Cellobiose Complex. Epoxyalkyl glycosides have frequently been used to identify the active site nucleophile of glycosyl hydrolases (Withers & Aebersold, 1995). Protonation of the epoxide, in a manner analogous to the enzyme mechanism itself, results in the generation of a highly reactive species which may be subject to attack by the enzymatic nucleophile (Figure 4B). If the true nucleophile is to be identified, then a certain amount of fortune may be required since these reagents are notoriously promiscuous and frequently make explicit liaisons with residues other than the mechanistic nucleophile [for example, see Havukainen et al. (1996)]. In our case, the length of the butyl linker region corresponds to the spacing of a single sugar residue and the 3,4-epoxybutyl cellobioside successfully labels the residue previously implicated as the catalytic nucleophile, Glu197 (Figure 5). This absolute assignment of this residue as the catalytic nucleophile is shown by specific reversible labeling using the mechanism-based inhibitor, 2,4-dinitrophenyl-2-deoxy-2-fluorocellobioside, in the preceding paper by MacKenzie et al. (1997).

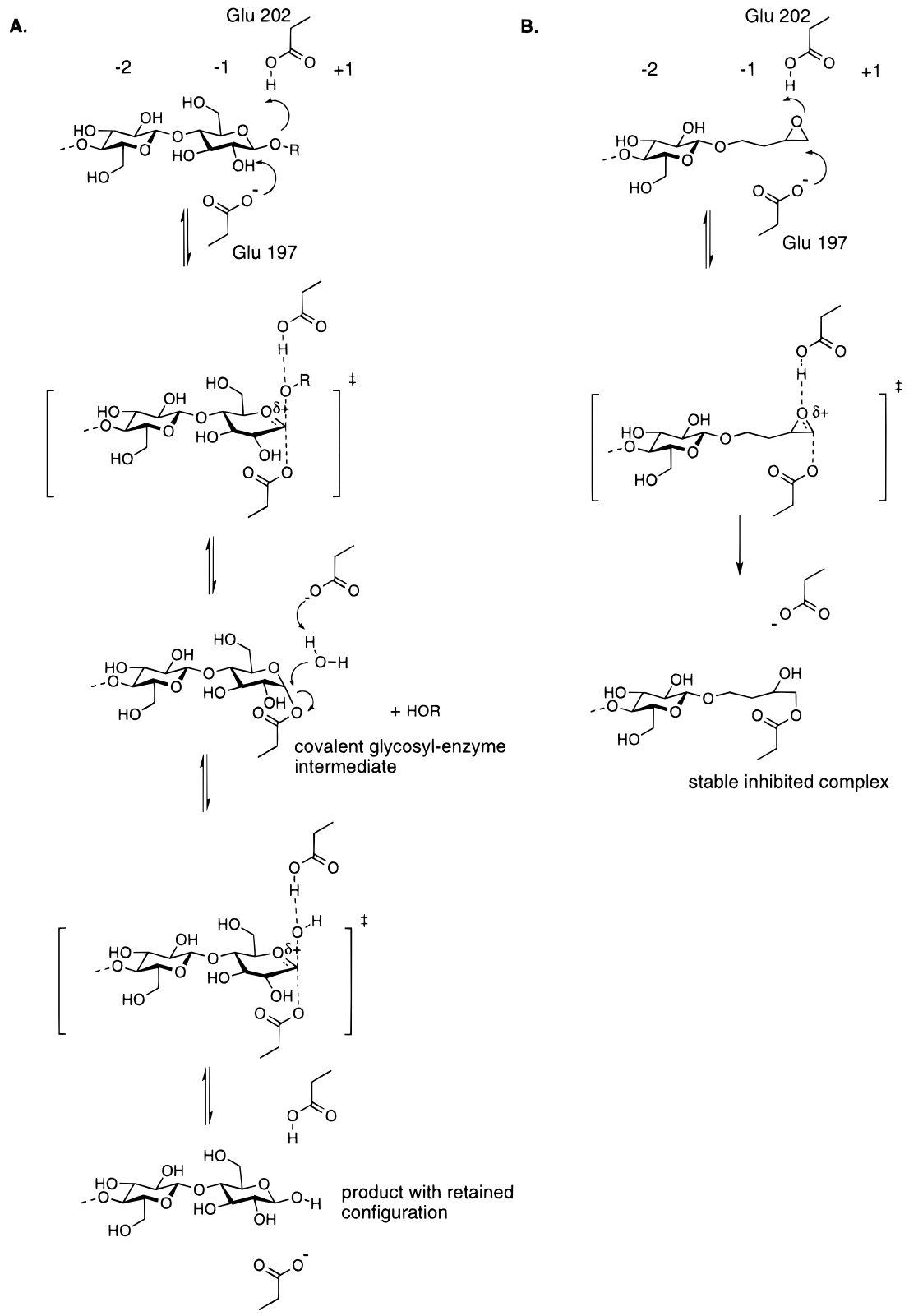

FIGURE 4: (A) Reaction scheme for glucoside catalysis by EG I. Catalysis occurs in a double-displacement reaction mechanism *via* a covalent enzyme intermediate which is formed and subsequently hydrolyzed *via* oxocarbenium ion transition states. Glu202 functions as the Brønsted acid/base and Glu197 as the nucleophile. (B) Likely reaction scheme for labeling by the epoxybutyl cellobioside.

Standard synthesis of 3,4-epoxybutyl cellobioside results in a mixture of both *R* and *S* stereoisomers (Legler & Bause, 1973; Rodriguez & Stick, 1990). Similar epoxide labeling of the catalytically equivalent *Bacillus* 1,3-1,4-glucanase structure identified the analogous nucleophile but did not reveal a density of sufficient quality to discriminate between the *R* or *S* stereoisomers (Keitel et al., 1993). In this case, refinement strongly suggests that both isomers are present in the crystal. Large difference density peaks, corresponding to the oxygen atom of the alternate isomer, result if the inhibitor is modeled as a single isomer only. The most stable refinement of the inhibitor occurred when it was modeled

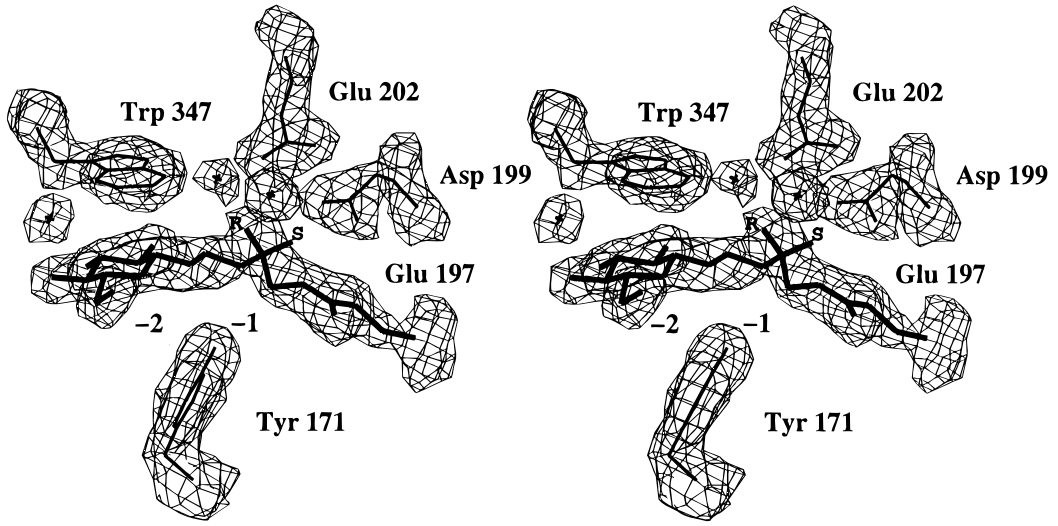

FIGURE 5: Stereo $2F_o-F_c$ electron density (maximum likelihood/σ_A weighted), contoured at 0.33 eÅ^{-3}, for the 3,4-epoxybutyl cellobioside inhibitor covalently bound to the catalytic nucleophile, Glu197. Only a single β-D-glucopyranose ring of the inhibitor is present; the other is presumed to have been cleaved off during incubation. Density is present for both the *R* and *S* stereoisomers of the inhibitor, and these have been modeled with occupancies of 0.7 and 0.3, respectively.

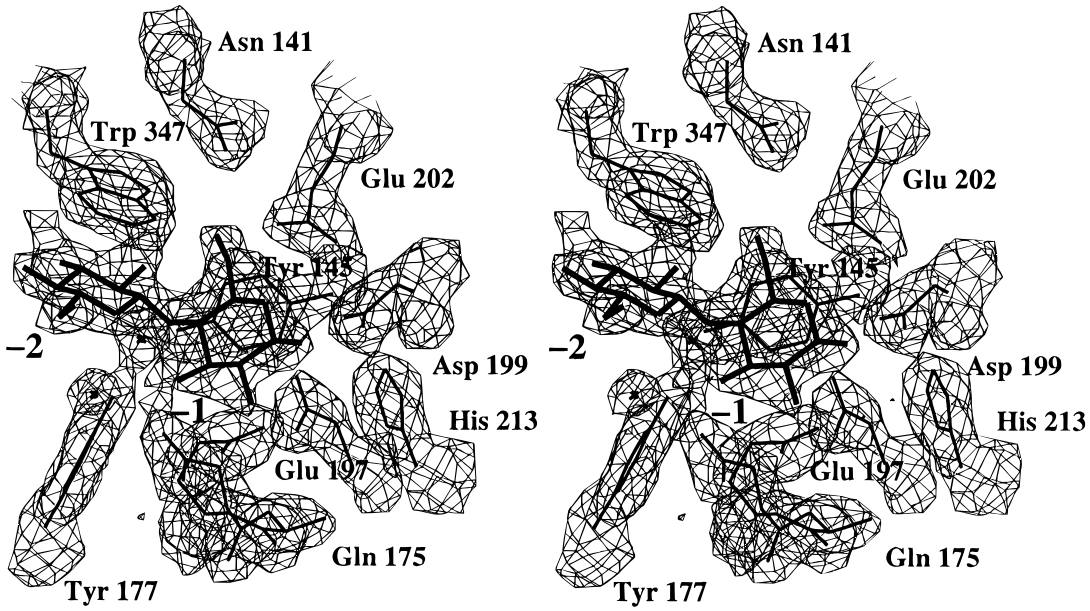

FIGURE 6: Stereo $2F_o-F_c$ electron density (maximum likelihood/σ_A weighted), contoured at 0.30 eÅ^{-3}, for cellobiose bound to −2 and −1 subsites of EG I.

as a single species, but with both *R* and *S* oxygens present at the chiral center, (Figure 5). The electron density is much stronger for the *R* stereoisomer. This isomer has been modeled with a crystallographic occupancy of 0.7 and the *S* stereoisomer with a correspondingly reduced occupancy of 0.3. These respective occupancies result in both atoms having an equal temperature factor of approximately 21 Å^2. The density is clear for the butyl linker region (spanning the −1 subsite) and for one glucopyranosyl unit in the −2 subsite, which is in the standard $^4C_1$ conformation. No density is observed, however, for the second of the two glucose residues of the cellobiose moiety of the inhibitor. This could reflect the presence of only four significant binding sites (−2 to +2) as indicated by kinetic studies, but it is more likely that, during the long time course of inactivation, the terminal glucose residue has been hydrolyzed by the enzymatic action of EG I. Certainly, the electron density terminates extremely abruptly with the C-4 hydroxyl of the −2 subsite saccharide, more indicative of only one sugar ring rather than a glycosidic linkage to a second, more disordered, ring.

Cellobiose Complex. Cellobiose is found clearly bound to the −2 and −1 subsites (Figure 6). Both sugar rings are in the standard $^4C_1$ conformation. This is markedly different from the structure of EG I with a nonhydrolyzable thiosaccharide analog in the −2, −1, and +1 sites, in which the −1 subsite sugar was distorted toward a skewed-boat conformation with an axial leaving group orientation (Sulzenbacher et al., 1996). The consequences of this will be discussed below. There is a relative paucity of interactions between the −2 subsite sugar and the enzyme, with only the C-2 hydroxyl making a direct hydrogen bond to the protein (Figure 7A). The C-6 hydroxyl group points into the solvent region and is disordered. The C-4 and C-3 hydroxyls all interact with the protein *via* a network of solvent water molecules. The main feature of the −2 subsite

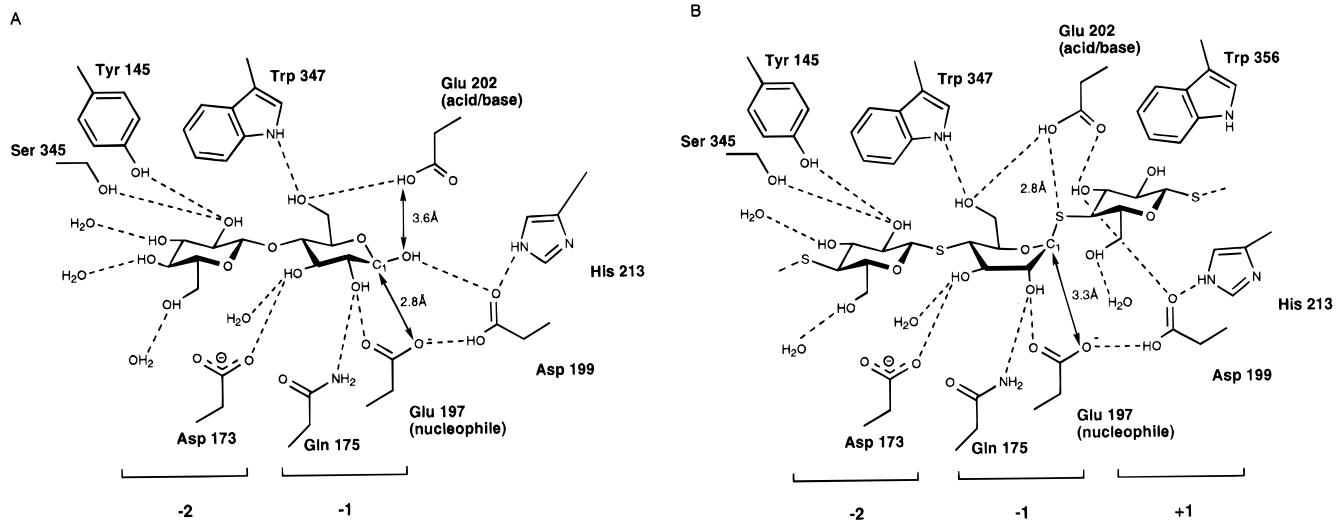

FIGURE 7: (A) Schematic diagrams of the interactions of cellobiose with EG I from *F. oxysporum*. (B) Schematic diagrams of the interactions of thio-DP5 with EG I from *F. oxysporum* (Sulzenbacher et al., 1996). Relevant distances are shown.

interactions is the aromatic stacking with Trp347. Tryptophan stacking interactions with the hydrophobic faces of the pyranose ring are a common feature of sugar−protein interactions (Vyas, 1991).

The −1 subsite makes more direct interactions with the protein. The base of the subsite is constructed from the side chain of Tyr145 (Figure 7A). The C-3 hydroxyl forms H bonds both to solvent water and to Asp173. Two hydroxyl interactions merit special consideration. The C-6 hydroxyl is present in the *gauche−gauche* conformation which permits H bonding to both the N-ϵ hydrogen of Trp347 and to the carboxylate of the catalytic acid/base Glu202. The C-2 hydroxyl forms H bonds to both Gln175 and the nucleophile Glu197. The C-2 hydroxyl interactions, to both a side chain amide hydrogen and the carboxylate of the nucleophile, are very reminiscent of those found in the complexes of the, structurally unrelated, family 5 and 10 cellulases which have recently been determined (Sakon et al., 1996; White et al., 1996). The C-1 hydroxyl makes an H bond to Asp199, but it is over 3.5 Å away from the catalytic acid/base Glu202, with which it must interact during the catalytic cycle (see below). The OE2 atom of the nucleophile Glu197, sits 2.8 Å below the anomeric C-1 with its *syn* lone pairs directed toward the C-1 atom. The hydroxyl at the anomeric C-1 is present only in the β configuration and not as a mixture of α and β configurations as might be expected by mutarotation. An α configuration at C-1 would result in a severe steric clash of the hydroxyl with the carboxylate group of Glu197.

Comparison with the Thio-DP5 Complex and Implications for Catalysis by EG I. We have recently described the structure of the *F. oxysporum* EG I in a complex with a nonhydrolyzable thiooligosaccharide (thio-DP5), bound to the −2, −1, and +1 subsites, with an intact thioglycosidic linkage spanning the active site (Sulzenbacher et al., 1996). The salient feature of that structure was the fact that the −1 subsite pyranoside ring was distorted, resulting in a quasi-axial glycosidic bond and leaving group orientation. Such a distortion, between a $^{1}S_{3}$ skew-boat and a $^{1,4}B$ boat conformation on the pyranose ring interconversion pathway, was also recently described for the complex of chitobiase with its unhydrolyzed substrate chitobiose (Tews et al., 1996). The primary difference between the equivalent sugars of the thio-DP5 structure and the cellobiose complex described here is that the −1 subsite sugar has a ground state $^{4}C_{1}$ conformation in the cellobiose complex in marked contrast to the skew-boat of the thio-DP5 complex.

The interactions of EG I with both thio-DP5 and cellobiose are shown in the schematic diagram in Figure 7. Nearly all of the −2 and −1 subsite interactions are conserved between the two complexes, including the positions of all the solvent molecules. The distortion of the thio-DP5 complex results in a quasi-axial orientation for the glycosidic bond and leaving group which provides many mechanistic advantages for glycosyl hydrolysis. This distortion and the accompanying axial glycosidic bond orientation permit protonation of the glycosidic oxygen by the enzymatic acid/base (in this case Glu202) which is not possible when the sugar is in the ground state $^{4}C_{1}$ chair conformation with an equatorial glycosidic bond (as observed in the cellobiose product complex). Such an interaction is essential for the hydrolysis of substrates whose leaving groups require substantial protonic assistance, such as the natural substrate. Probably the most important result of pyranoside ring distortion is that the anomeric C-1 is presented in a much more favorable orientation for "in-line" nucleophilic attack by the enzymatic nucleophile Glu197. In the $^{4}C_{1}$ chair conformation, nucleophilic attack at C-1 is sterically hindered, primarily by the hydrogen on the anomeric carbon, but also by the H-3 and H-5 protons. Such steric hindrance is well-documented in organic chemistry, primarily by work on conformationally challenged systems such as 2-adamantyl compounds [for review, see Bentley and Schleyer (1977)]. The observed distortion from the $^{4}C_{1}$ chair toward a skew-boat conformation "opens up" the anomeric carbon and thus permits nucleophilic attack at C-1, concomitant with leaving group departure. Distortion toward a boat conformation is also consistent with a stereoelectronic role in glycoside catalysis since it results in an orientation for one of the ring oxygen lone pairs antiperiplanar to the glycosidic bond (Deslongchamps, 1983; Kirby, 1984).

The lack of a ground state saccharide distortion for the −1 subsite sugar in the product complex (*i.e.*, in the absence of a covalent link to the +1 subsite sugar) is not surprising; indeed, it is what one would expect, since stabilization of a strained product complex would demand more interactions with the enzyme and thus presumably a more unfavorable

situation for product release. It is, however, in marked contrast to similar studies on both hen egg white and bacteriophage T4 lysozymes, all of which have suggested distorted sugar rings in ground state product complexes (Hadfield et al., 1994; Kuroki et al., 1993; Strynadka & James, 1991). In EG I, no barrier to a $^4C_1$ conformation for the −1 subsite sugar exists in the absence of a covalent link to a sugar in +1. In the preceding paper, MacKenzie et al. (1997) demonstrate that regeneration of the glycosyl enzyme intermediate is 100 times faster with a saccharide nucleophile than by spontaneous hydrolysis. This presumably reflects utilization of the aglycone binding energy during catalysis. It is tempting to speculate that, since substrate distortion appears to be driven by these aglycone (+1/+2) binding site interactions, these provide the structural basis for the kinetic preference of saccharide nucleophiles over solvent water.

In addition to protonation of the glycosidic oxygen and nucleophilic attack at C-1, the noncovalent interactions of the sugar hydroxyl functions are known to be crucial in enzymatic glucoside hydrolysis (Namchuk & Withers, 1995). In many systems, two of these interactions, the C-6 and C-2 hydroxyls in the −1 subsite sugar, are extremely important. We note, with interest, that it is these hydroxyls whose positions appear to change most when we compare the $^4C_1$ conformation in the cellobiose complex with the skew-boat of the thio-DP5 complex. At this resolution, however, any apparent change in distance is only of a magnitude similar to the likely experimental error and so must be treated with caution. In solution, the presence of a hydroxyl function at C-2 is known to impair nonenzymatic catalysis, presumably by an inductive destabilization of the transition states. On enzyme, however, the interactions of the C-2 hydroxyl contribute in a positive way to catalysis, leading to the suggestion that the interactions of this group must in some way stabilize the conformationally distorted transition state. This is consistent with the role of a charged hydrogen bond with one of the oxygens of the nucleophile, in the enzymatic transition state (Namchuk & Withers, 1995; White et al., 1996). Although the thio-DP5 complex structure would appear to make "better" interactions between the C-2 hydroxyl and the enzyme, thio-DP5 is not a transition state analog and in the absence of an appropriate analogue interpretation of these interactions remains extremely speculative.

Although many favorable interactions result from the substrate distortion toward a skew-boat conformation, the distance from the nucleophile oxygen to the anomeric C-1 increases from approximately 2.8 to 3.3 Å. Since these two atoms form a covalent bond in the stable glycosyl−enzyme intermediate, any increase in this distance could be considered deleterious. Binding of the substrate in the distorted boat conformation is, however, clearly beneficial for the reasons outlined above: protonation of the glycosidic bond, removal of steric hindrance to nucleophilic attack, and potentially favorable stereoelectronics; but also the skew-boat conformation may be a necessary prerequisite for the passage through the half-planar oxocarbenium ion transition state toward the α configuration of the glycosyl−enzyme intermediate.

Summary. This paper, together with our previous structural study (Sulzenbacher et al., 1996) and the accompanying paper by Mackenzie et al. (1997), helps to dissect the interactions and roles of the protein−carbohydrate interactions in catalysis by EG I. Substrate binding is accompanied by pyranose ring distortion in the −1 subsite. This distortion is driven in part by the shape of the substrate binding surface at this position and by utilization of the aglycone binding energy. The resultant pyranoside conformation has an axial leaving group orientation which permits protonation of the interglycosidic oxygen, removes steric hindrance to nucleophilic attack at C-1, and is also favored by stereoelectronic factors. These results combine the essential features of the double-displacement reaction mechanism proposed by Koshland (1953) with pyranoside ring distortion as originally described by Phillips and co-workers (Ford et al., 1974). It is likely that these features of the EG I catalytic mechanism will be applicable to many of the vast spectrum of β-retaining glycosyl hydrolases found in nature.

ACKNOWLEDGMENT

The authors thank Bernard Henrissat, Steve Withers, Lloyd Mackenzie, and Alwyn Jones, FRS, for frequent helpful discussions.

REFERENCES

Armand, S., Drouillard, S., Schülein, M., Henrissat, B., & Driguez, H. (1997) *J. Biol. Chem.* (in press).

Bentley, T. W., & Schleyer, P. v. R. (1977) *Adv. Phys. Org. Chem. 14*, 1−67.

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

Biely, P., Vrsanska, M., & Claeyssens, M. (1991) *Eur. J. Biochem. 200*, 157−163.

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

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

Christensen, T., Wöldike, H., Boel, E., Mortensen, S. B., Hjortshøj, K., Thim, L., & Hansen, M. T. (1988) *Bio/Technology 6*, 1419−1422.

Claeyssens, M., Van Tilbeurgh, H., Tomme, P., & Wood, T. M. (1989) *Biochem. J. 261*, 819−825.

Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D 50*, 760−763.

Davies, G., & Henrissat, B. (1995) *Structure 3*, 853−859.

Davies, G. J., Tolley, S. P., Sulzenbacher, G., Divne, C., Jones, T. A., Wöldike, H., & Schülein, M. (1997a) *J. Mol. Biol.* (in press).

Davies, G. J., Wilson, K. S., & Henrissat, B. (1997b) *Biochem. J. 321*, 557−559.

Deslongchamps, P. (1983) *Stereoelectronic Effects in Organic Chemistry*, Pergamon Press, Oxford.

Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T., & Jones, A. (1994) *Science 265*, 524−528.

Divne, C., Ståhlberg, J., & Jones, T. A. (1996) in *IUCr: XVII Congress and General Assembly* (Griffin, J. E., Ed.) p C-193, International Union of Crystallography, Seattle.

Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J., & Wang, J. L. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 2580−2584.

Ford, L. O., Johnson, L. N., Machin, P. A., Phillips, D. C., & Tijan, T. (1974) *J. Mol. Biol. 88*, 349−371.

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

Hadfield, A. T., Harvey, D. J., Archer, D. B., MacKenzie, D. A., Jeenes, D. J., Radford, S. E., Lowe, G., Dobson, C. M., & Johnson, L. N. (1994) *J. Mol. Biol. 243*, 856−872.

Hardman, K. D., & Ainsworth, C. F. (1972) *Biochemistry 11*, 4910−4919.

Havukainen, R., Törrönen, A., Laitinen, T., & Rouvinen, J. (1996) *Biochemistry 35*, 9617−9624.

Henrissat, B. (1991) *Biochem. J. 280*, 309−316.

Henrissat, B. (1994) in *4th European workshop on crystallography of biological macromolecules* (Mattevi, A., & Wilson, K., Eds.) p O-33, EC Human Capital and Mobility Programme, Como, Italy.
Henrissat, B., & Bairoch, A. (1993) *Biochem. J. 293*, 781−788.
Henrissat, B., & Bairoch, A. (1996) *Biochem. J. 316*, 695−696.
Jones, T. A., Zou, J.-Y., Cowan, S. W., & Kjeldgaard, M. (1991) *Acta Crystallogr. A 47*, 110−119.
Keitel, T., Simon, O., Borriss, R., & Heinemann, U. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5287−5291.
Kirby, A. J. (1984) *Acc. Chem. Res. 17*, 305−311.
Kleywegt, G. J., & Jones, T. A. (1994) in *From first map to final model* (Bailey, S., Hubbard, R., & Waller, D., Eds.) pp 59−66, EPSRC, Daresbury, U.K.
Koshland, D. E. (1953) *Biol. Rev. 28*, 416−436.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Kuroki, R., Weaver, L. H., & Matthews, B. W. (1993) *Science 262*, 2030−2033.
Laskowski, R. A., McArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 282−291.
Legler, G., & Bause, E. (1973) *Carbohydr. Res. 28*, 45−52.
MacKenzie, L., Davies, G. J., Schülein, M., & Withers, S. G. (1997) *Biochemistry 36*, 5893−5901.
McCarter, J. D., & Withers, S. G. (1994) *Curr. Opin. Struct. Biol. 4*, 885−892.
Murshudov, G. N., Vagin, A. A., & Dodson, E. J. (1997) *Acta Crystallogr. D* (in press).
Namchuk, M. N., & Withers, S. G. (1995) *Biochemistry 34*, 16194−16202.
Navaza, J. (1994) *Acta Crystallogr. A 50*, 157−163.
Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963) *J. Mol. Biol. 7*, 95−99.
Raymond, S., Heyraud, A., Qui, D. T., Kvick, A., & Chanzy, H. (1995) *Macromolecules 28*, 2096−2100.
Read, R. J. (1986) *Acta Crystallogr. A 42*, 140−149.
Rodriguez, E. B., & Stick, R. V. (1990) *Aust. J. Chem. 43*, 665−679.
Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., & Jones, T. A. (1990) *Science 249*, 380−386.
Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R., & Karplus, P. A. (1996) *Biochemistry 35*, 10648−10660.
Schou, C., Rasmussen, G., Kaltoft, M.-B., Henrissat, B., & Schülein, M. (1993) *Eur. J. Biochem. 217*, 947−953.
Schülein, M. (1997) *J. Biotechnol.* (in press).
Sheppard, P. O., Grant, F. J., Oort, P. J., Cindy, A. S., Foster, D. C., Hagen, F. S., Upshall, A., McKnight, G. L., & O'Hara, P. J. (1994) *Gene 150*, 163−167.
Sinnott, M. L. (1990) *Chem. Rev. 90*, 1171−1202.
Strynadka, N. C. J., & James, M. N. G. (1991) *J. Mol. Biol. 220*, 401−424.
Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M., & Davies, G. J. (1996) *Biochemistry 35*, 15280−15287.
Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S., & Vorgias, C. E. (1996) *Nat. Struct. Biol. 3*, 638−648.
Tomme, P., Kwan, E., Gilkes, N. R., Kilburn, D. G., & Warren, R. A. J. (1996) *J. Bacteriol. 178*, 4216−4223.
Vyas, N. K. (1991) *Curr. Opin. Struct. Biol. 1*, 732−740.
Wang, Q., Graham, R. W., Trimbur, D., Warren, R. A. J., & Withers, S. G. (1994) *J. Am. Chem. Soc. 116*, 11594−11595.
White, A., Withers, S. G., Gilkes, N. R., & Rose, D. R. (1994) *Biochemistry 33*, 12546−12552.
White, A., Tull, D., Johns, K., Withers, S. G., & Rose, D. R. (1996) *Nat. Struct. Biol. 3*, 149−154.
Withers, S. G., & Aebersold, R. (1995) *Protein Science 4*, 361−372.

BI962963+