Computational and experimental studies of the catalytic mechanism of Thermobifida fusca cellulase Cel6A (E2)

Gwenaëlle André-Leroux, P. Kanchanawong, R. Palma, H. Cho, X. Deng, D. Irwin, M.E. Himmel, D.B. Wilson, J.W. Brady

Computational and experimental studies of the catalytic mechanism of *Thermobifida fusca* cellulase Cel6A (E2)

G.André[1,2], P.Kanchanawong[1,3], R.Palma[1,4], H.Cho[5], X.Deng[5,6], D.Irwin[5], M.E.Himmel[7], D.B.Wilson[5] and J.W.Brady[1,8]

[1]Department of Food Science, Stocking Hall and [5]Department of Molecular Biology and Genetics, Biotechnology Building, Cornell University, Ithaca, NY 14853, USA, [6]Department of Chemical Engineering, Xiamen University, Xiamen 361005, China and [7]National Renewable Energy Laboratory, National Bioenergy Center, 1617 Cole Boulevard, Golden, CO 80401-3393, USA

[2]Permanent address: Institut National de la Recherché Agronomique, Centre de Recherché de Nantes, B.P. 527, 44026 Nantes Cedex 03, France

[3]Present address: Argule Inc., 19 Presidential Way, Woburn, MA 01810, USA

[4]Present address: GlaxoWellcome, Biomedsci Development Group, 3030 Cornwallis Road, RTP, NC 27709, USA

[8]Author to whom correspondence should be addressed
E-mail: jwb7@cornell.edu

Mutagenesis experiments suggest that Asp79 in cellulase Cel6A (E2) from *Thermobifida fusca* has a catalytic role, in spite of the fact that this residue is more than 13 Å from the scissile bond in models of the enzyme–substrate complex built upon the crystal structure of the protein. This suggests that there is a substantial conformational shift in the protein upon substrate binding. Molecular mechanics simulations were used to investigate possible alternate conformations of the protein bound to a tetrasaccharide substrate, primarily involving shifts of the loop containing Asp79, and to model the role of water in the active site complex for both the native conformation and alternative low-energy conformations. Several alternative conformations of reasonable energy have been identified, including one in which the overall energy of the enzyme–substrate complex in solution is lower than that of the conformation in the crystal structure. This conformation was found to be stable in molecular dynamics simulations with a cellotetraose substrate and water. In simulations of the substrate complexed with the native protein conformation, the sugar ring in the –1 binding site was observed to make a spontaneous transition from the $^4C_1$ conformation to a twist-boat conformer, consistent with generally accepted glycosidase mechanisms. Also, from these simulations Tyr73 and Arg78 were found to have important roles in the active site. Based on the results of these various MD simulations, a new catalytic mechanism is proposed. Using this mechanism, predictions about the effects of changes in Arg78 were made which were confirmed by site-directed mutagenesis.
Keywords: catalytic mechanism/cellulase Cel6A (E2)/molecular dynamics/mutagenesis/*Thermobifida fusca*

Introduction

Cellulases are a highly diverse group of enzymes whose function is crucial to the healthy functioning of the biosphere, since more than half of all the biomass on the planet consists of their substrate, cellulose (Himmel *et al.*, 1999). The actual mechanisms of these enzymes are incompletely understood, which is a practical problem in their industrial use to convert waste plant material into simple sugars for fermentation to ethanol or other chemicals. The principal industrial challenge is the low turnover rate, and attempts to produce more efficient enzymes by bioengineering are hampered by the lack of a detailed understanding of their catalytic mechanism on crystalline substrates.

Most cellulases have a multiple-domain structure, with a specific globular binding domain, which interacts directly with the cellulose substrate, a separate globular catalytic domain where the glycosidic linkages are hydrolyzed and a linker segment joining the two domains. *Thermobifida fusca* Cel6A (E2), which is the subject of this paper, has this overall structure, with a catalytic domain consisting of 286 residues, a C-terminal binding domain consisting of 103 residues and a glycosylated linker segment 22 residues long (Spezio *et al.*, 1993). [The nomenclature of cellulases has recently been systematized for greater clarity (Henrissat *et al.*, 1998) and in this newer system, E2 is designated Cel6A; however, for brevity, we will use E2 here.] The catalytic domains for many cellulases can be produced separately, as is the case for E2, and this peptide, designated E2cd, has full activity toward soluble substrates and residual activity on solid cellulose.

E2 is an endocellulase, capable of cleaving a cellulose chain at any linkage after the first one. It is a configuration-inverting enzyme, meaning that the configuration at the anomeric center of the reducing end produced by cleavage is α, rather than the equatorial β linkage found in the substrate. It is stable up to 65°C and its activity is relatively insensitive to pH over the range 4–10 (Wolfgang, 1998; Wolfgang and Wilson, 1999). The gene for E2 has been cloned into both *Escherichia coli* and *Streptomyces lividans* and it is expressed in sufficient quantities to allow structural studies (Ghangas and Wilson, 1988; Zhang *et al.*, 2000a).

The crystal structure of E2 has been determined by X-ray diffraction and is very useful in attempting to deduce the catalytic mechanism of this enzyme (Spezio *et al.*, 1993). However, there are problems in reconciling this structure with the known mutagenesis data. For example, mutation of Asp79 to Asn reduces activity more than 30-fold (see Table I), yet this residue is ~13 Å from the scissile bond. Mutagenesis studies of Asp117 have identified this acid as the essential catalytic acid in a general acid–base mechanism (Wolfgang, 1998; Wolfgang and Wilson, 1999), and other experiments have identified Asp265 as important in catalysis, although primarily for substrate binding. Asp117 and Asp265 are found on opposite sides of the binding cleft of the enzyme, bracketing the scissile bond, but both residues are more than 5.5 Å from the glycosidic oxygen atom of the scissile bond, too far away to be playing a role in catalysis by directly donating a proton

Table I. Activity of Cel6A mutants on different cellulose substrates

Cel 6A mutant	Relative specific activity (%)[a]				Dissociation binding constant K_d (M×10⁻⁸)	Corresponding MU(glc)₃[b] residue in *T.reesei* CBH II
	Swollen cellulose SC	Carboxymethyl-cellulose CMC	Filter-paper FP	Bacterial micro-crystalline cellulose BMCC		
Wild-type	100	100	100	100	2.1–5.3	
R78K	54	15	41	52	1.2	R174
R78A[c]	<1.3	<0.9	<18.7	<0.9	0.9	
D117A (Wolfgang, 1999)	0.02	0.03	9		5.7	D221
D117E (Wolfgang, 1999)	0.06	0.06	17		1.7	
D117N (Wolfgang, 1999)	0.02	0.02	5		Curve fit poor	
Y73F (Barr *et al.*, 1998)	5.7	8.4	74		2.1	Y169
Y73S (Barr *et al.*, 1998)	0.088	0.022	14		1.6	
Y73E (Wolfgang, 1998)	0.13	0.09	17		2.4	
D79A (Wolfgang, 1999)	0.7	1.3	42		3.5	D175
D79E (Wolfgang, 1999)	0.6	0.4	38		20	
D79N (Wolfgang, 1999)	2	3.9	55		4.7	
K259H (Zhang *et al.*, 2000b)	66	373	30		Too weak to determine	K395
K259A (Zhang *et al.*, 2000b)	0.5	262	17		Undetectable	
K259R (Zhang *et al.*, 2000b)	0.019	0.0033	0.82		Not determined	

[a]Specific activities for wild-type Cel6A in μmol cellobiose released/(min.μmol enzyme) were 421, 575, 1.52 and 2.76 for SC, CMC, FP and BMCC, respectively.
[b]MU(glc)₃ = 4-methylumbelliferyl-β-cellotrioside.
[c]Target digestion (10% for SC and BMCC, 4% for FP and 5% for CMC) could not be reached on any substrate.

to cleave the bond. This situation is similar to that found in the homologous exocellulase *Trichoderma reesei* CBH II (Rouvinen *et al.*, 1990) [also now called Cel6A (Henrissat *et al.*, 1998)], where it has been proposed that water molecules bridge between the catalytic acid residues and the scissile glycosidic oxygen atom. The results of Asp79 mutations in E2 and also those of other E2 mutations (Zhang and Wilson, 1997) suggest that in the enzyme–substrate complex there is a major rearrangement of the protein structure away from that of the crystal structure, involving the region containing Asp79.

The X-ray crystal structure of CBH II and some mutagenesis studies indicate that CBH II Asp221 (the residue corresponding to E2 Asp117) functions as the catalytic acid and CBH II Asp175 (E2 Asp79) acts to raise the pK_a of the catalytic acid (Rouvinen *et al.*, 1990; Koivula *et al.*, 1996). However, there was no evidence for an essential catalytic base. Considerable work has been done on another family 6 endocellulase, *Cellulomonas fimi* CenA (Damude *et al.*, 1995). Although no crystal structure for this enzyme has been determined, site-directed mutagenesis data provided evidence for a catalytic acid, CenA Asp252 (E2 Asp117), and a catalytic base, CenA Asp392 (E2 Asp265). However, the Asp392–Asn mutation was not produced and this was the mutation in E2 which showed that this residue was not essential for activity.

Unfortunately, it has not yet been possible to co-crystallize E2 with an oligosaccharide substrate or with a non-cleavable substrate analog. It is also not possible to diffuse small, soluble oligosaccharide substrates into the crystal. However, in a previous paper we reported molecular mechanics (MM) conformational energy calculations which placed a tetrasaccharide substrate into the binding cleft and proposed a possible structure of the complex of the protein with this substrate (Taylor *et al.*, 1995). In principle, it should be possible to use similar calculations to evaluate whether the protein conformational energetics would permit significant shifts in the position of the loop containing Asp79 and also to simulate the interactions of the substrate with the individual residues of the protein

and the solvent water. In this paper, we report such MM conformational energy calculations and also molecular dynamics (MD) simulations of E2 complexed with a tetrasaccharide substrate. The results of these simulations are used to suggest a new mechanism for E2. This proposed mechanism suggests that Arg78 participates in catalysis. Mutations in this residue were subsequently prepared and characterized. Their activities, reported here, were found to be consistent with our new mechanism.

Methods

Computational

This paper reports two sets of simulations: standard MD simulations of the dynamics of the previously obtained enzyme–substrate complex in aqueous solution and a search for alternative conformations of the loop containing Asp79 using constrained MD simulations and energy minimization calculations. In the latter calculations, alternative conformations were examined to look for any which had mechanical energies comparable to or lower than the energy of the crystal structure. In particular, because mutagenesis data demonstrated the critical role of Asp79 in the catalytic activity, alternative conformations for the loop containing residue 79 were sought which placed this residue closer to the scissile bond of the substrate. The MD simulations were used to refine the energy-minimized structures, exploring the configuration space around these structures to determine if there were other nearby conformations separated by energy barriers and also to give a picture of the motions of the enzyme–substrate complex that might contribute to the catalytic mechanism.

All of the calculations reported here used the CHARMM molecular mechanics program and the recently developed CHARMM22 parameter set for protein atoms (MacKerell *et al.*, 1998). The sugar atoms were modeled using parameters specifically developed for carbohydrates (Palma *et al.*, 2000). Water molecules were represented using the TIP3P force field
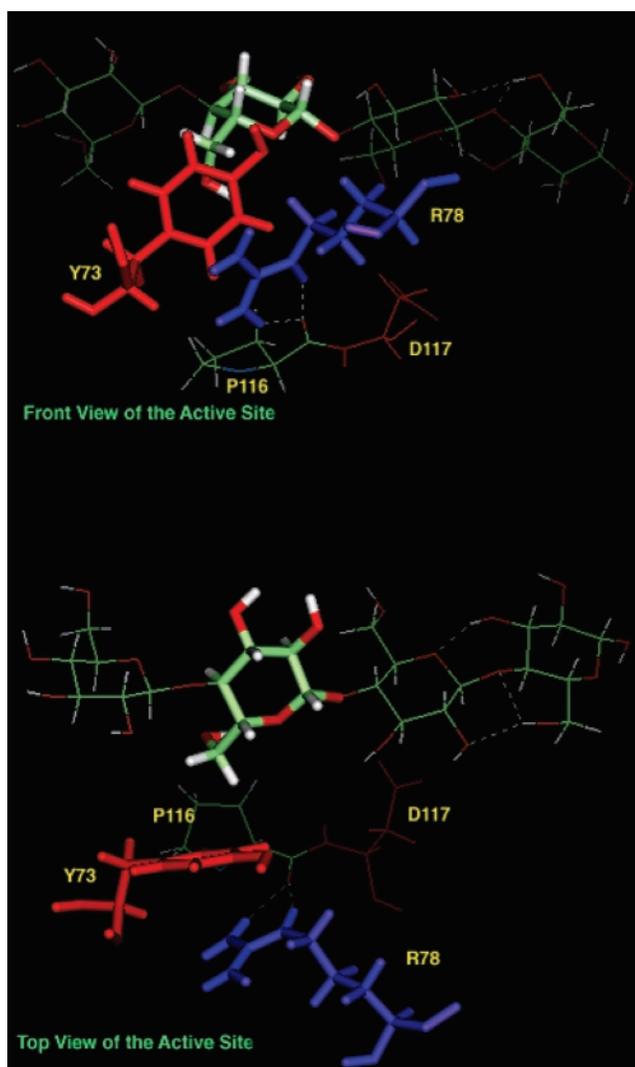
Fig. 1. Two views of the relative positions of the sugar ring in the −1 binding site, shown in thicker lines than the other three sugar rings, and Tyr73, Arg78, Pro116 and Asp117 in the minimized complex of the substrate with the crystallographic protein structure. Asp117 is the catalytic acid group, which, as can be seen, is far from the scissile bond.

Fig. 2. Illustration of the positions of the two loops P1 and P2 in E2, as defined in the present simulations, and an indication of the constraint distance between their two centers of mass used to draw them together.

(Jorgensen *et al.*, 1983). In all calculations, the long-range interactions were smoothed to zero using switching functions between 10 and 12 Å, since Ewald summation could not be used in these non-periodic systems. In all calculations, the lengths of chemical bonds to hydrogen atoms were kept fixed using the constraint algorithm SHAKE (van Gunsteren and Berendsen, 1977).

All of the standard equilibrium MD simulations (wild-type and mutants) started from a previously calculated structure for E2cd, complexed with a tetrasaccharide substrate (Taylor *et al.*, 1995). This structure was produced by building the tetrasaccharide into the binding site using the crystal coordinates for E2cd bound to cellobiose as the starting point. In this model, the second ring of the substrate, in the so-called −1 position immediately preceding the scissile bond, is tilted sharply with respect to the other rings of the chain. The tilting of this ring away from the generally flat ribbon-like cellulose conformation was also observed in CBH II (Rouvinen *et al.*, 1990) and may have functional significance. The tilting apparently results from steric repulsions with the non-polar aliphatic
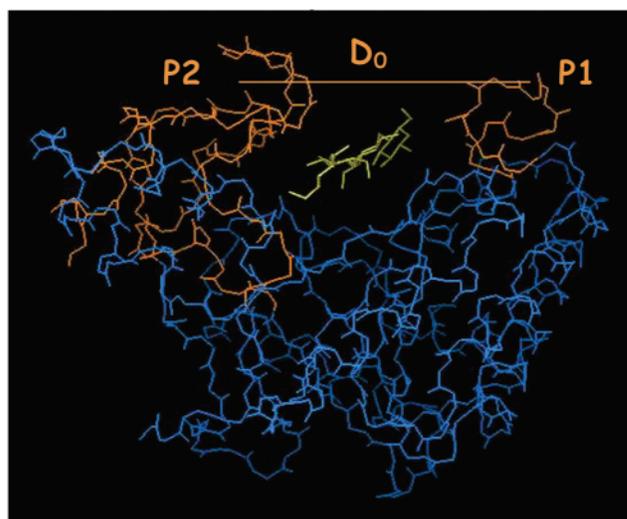
portion of the side chain of Lys259. Figure 1 illustrates two views of the tilt of the −1 sugar ring, showing the close interaction of the primary alcohol group of this sugar residue with the aromatic ring of Tyr73.

MD simulations of the hydration of the enzyme–substrate complex

Unfortunately, owing to the large size of E2cd, it was not practical to solvate it by placing it in a very large box of water and applying periodic boundary conditions. In the present MD simulations, two alternative approximations were used to include solvation, depending on the objectives of the calculation. In simulations which sought to model large-scale motions such as potential conformational shifts of the loop containing Asp79, the entire protein was covered with a layer 5.2 Å thick of water molecules which were not constrained and thus were allowed to relax as any conformational shifts occurred. This method has been used in many previous simulations, including the recent folding calculation of Duan and Kollman (Duan and Kollman, 1998), in which an unfolded polypeptide underwent substantial conformational changes, presumably driven in part by solvation contributions. For simulations that focused on events taking place in the immediate vicinity of the catalytic site, a different procedure was used. In these simulations, a previously equilibrated sphere of water with a radius of 18 Å was placed with its origin at the glycosidic oxygen atom of the scissile bond in the enzyme–substrate structure obtained in the previous study (Taylor *et al.*, 1995). Those water molecules that overlapped with any of the protein or substrate atoms were then removed. Unlike the situation in the first type of simulations, the water molecules were then constrained to remain within this sphere using an inwardly directed boundary constraining force. This constraint prevents any water molecules from escaping from the solvation sphere, but also tends to inhibit large-scale conformational motions of the enclosed protein chain. Water molecules in the outer 2 Å of this sphere were governed by a Langevin equation of motion using the Stochastic Boundary MD procedure implemented in CHARMM (Brünger *et al.*, 1984; Brooks *et al.*, 1985), while those atoms in the interior of the sphere were governed by conventional Newtonian equations.

In simulations of the hydration of the crystal structure of the wild-type protein or point mutants complexed with substrate, the starting structure was equilibrated for 20 ps followed by 40 ps of dynamics used for data collection. During the equilibration period the entire system was coupled to a thermal bath through the use of a Langevin equation of motion with a friction coefficient of 5 kcal/mol.Å^2. Since this structure had been previously extensively equilibrated, only this short 20 ps sequence of additional equilibration was needed. During the data collection period the Newtonian equations of motion were integrated as a microcanonical ensemble for all atoms which were within a sphere of 16 Å centered on the glycosidic oxygen of the scissile bond. The atoms in the outer 2 Å buffer region of the solvation sphere were governed by Langevin equations of motion (Brünger *et al.*, 1984; Brooks *et al.*, 1985), as were proteins atoms more than 18 Å from the center of the sphere. The two sulfate groups found in the crystal structure were included in these simulations, although they are apparently unimportant to the activity of the enzyme and are present in the experimental structure only as a consequence of its precipitation from ammonium sulfate.

Search for alternative low-energy conformations

The proposal that E2 undergoes a conformational change is further supported by the observation that in the homologous protein CBH II, the corresponding residue is close to the active site (Rouvinen *et al.*, 1990). To look for alternative E2 conformations, the segment from Gly77 to Ala88 was placed in the conformation of the CBH II protein by manually changing each (ϕ, ψ) angle pair in this segment to the values found in the analogous residues in CBH II. However, the loop in CBH II contains several residues missing from the E2 sequence, which makes it difficult to map the E2 loop conformation on to that in CBH II. This loop in E2 also is anchored to the chain segment behind it through a disulfide bond from Cys80 to Cys125 (since the sulfate groups from the crystal structure of E2 are also in this region and could potentially artificially inhibit loop motion, they were discarded in these simulations). Simply attempting to shift the chain mechanically toward the active site pulls against this bond and thus raises the energy significantly. Other attempts to move the segment containing Asp79 toward the active site by rotating about only the two terminal residues of the chosen segment, Gly77 and Ala88 (which follow and precede prolines, respectively), similarly failed, since such rigid loop motion also pulls directly against the disulfide bond.

Because of these difficulties, an alternative general scheme was developed which relied on constrained MD simulations to attempt to bring Asp79 closer to the active site while allowing the balance of the protein to relax to accommodate the new position of this residue. In this approach, a harmonic constraining force was applied which was a function of the distance being minimized. This constraining potential could have been made a direct function of the distance between the Asp79 oxygen atoms and the glycosidic oxygen atom of the scissile linkage. However, it was felt that this force would pull too strongly on the target residue and distort the conformation of the rest of the chain as it was pulled along. It might alternatively, and more probably, simply pull the lighter substrate chain out of the binding cleft. Accordingly, the distance chosen was the distance from the geometric center of the Gly77–Ala88 loop, on one side of the binding cleft, to the geometric center of the Asp217–Ala286 segment on the other

side of the cleft. Figure 2 illustrates these two loops in the minimized complex of the substrate with the crystallographic conformation. The constrained distance between the center of mass of the Asp79 loop, labeled P1 in the figure, and the loop on the opposite side of the cleft is also illustrated. A series of MD simulations were performed in which a harmonic constraining potential as a function of this distance was applied, with the minimum-energy distance d_0 for the constraining force decreasing by 0.5 Å with each successive simulation to gradually shift the loop toward the active site. During these dynamics simulations additional harmonic constraints were applied to the backbone torsion angles ϕ and ψ for the residues constituting the secondary structural elements of the protein to inhibit large, global conformational shifts. For each of the simulations in the series, after a period of 10 ps of Brownian dynamics simulation, the new conformation was energy minimized (without any constraints) for 500 steps using a conjugate gradient minimization algorithm. The value of d_0 was then decreased and another simulation begun starting from the final frame of the previous simulation, not the corresponding minimized structure. The lowest energy structures at each value of d_0 along this pathway were then used as an approximation to the reaction path for the process of shifting Asp79 towards the active site.

Experimental

Strains, plasmids and protein purification

E.coli BL21gold(DE3) (Stratagene, La Jolla, CA) was used for plasmid DNA isolation and production of the mutant proteins. The parent plasmid, pD905 (Zhang *et al.*, 2000a), which contains the pet26b(+) (Novagen) vector ligated to the wild-type Cel6A coding sequence, was used to construct the mutant plasmids. The forward primers containing a unique BsrG1 restriction site (bold) and the desired codon change (underlined) were as follows:
R78A-TCCTGGTCG**TGTACA**ACGCCCCGGGC<u>GCC</u>GACTGC
and
R78K-TCCTGGTCG**TGTACA**ACGCCCCGGGC<u>AAG</u>GACTGCG
PCR was used to create C-terminal fragments of the Cel6A coding sequence containing the desired mutations. These fragments were cut by *Bsr*G1 and *Sfi*1 and gel purified. The desired fragments were ligated into *Bsr*G1, *Sfi*1 restricted gel-purified pD905 and transformed into BL21gold(DE3) cells. The Cel6A coding regions from the *Bsr*G1 to the *Sfi*1 restriction site in the mutant plasmids were sequenced by the Cornell Biotechnology Facility to ensure that they contained only the desired mutations. The E2 Arg78 Ala and Arg78 Lys proteins were produced and purified as reported previously (Irwin *et al.*, 1993; Zhang *et al.*, 2000a/b).

Activity assays

The activities of wild-type E2, the Arg78 Lys and the Arg78 Ala mutant enzymes were determined on carboxymethylcellulose (CMC), phosphoric acid-swollen cellulose (SC) and bacterial microcrystalline cellulose (BMCC) at substrate concentrations of 10, 2.5 and 2.5 mg/ml for 30, 30 and 1000 min incubations at 50°C, respectively, in 0.05 M sodium acetate pH 5.5; filter-paper (FP) assays were run with 8.5 mg/ml for 1000 min. The reducing sugars produced were determined with dinitrosalicylic acid reagent (DNS) (Ghose, 1987) and calculations were done as reported previously (Irwin *et al.*, 1993) using the amount of enzyme required to reach the target digestion (10% for SC and BMCC, 4% for FP and 5% for CMC). All assay points
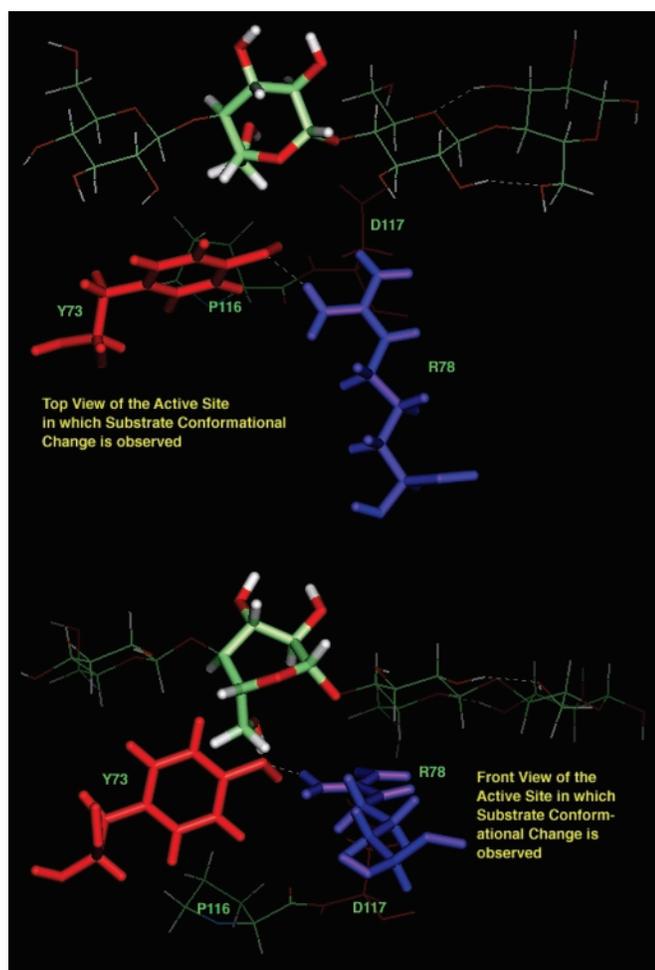
Fig. 3. Two views of the sugar ring in the −1 binding site after it has spontaneously 'flipped' during the course of a dynamics simulation into the $^2S_0$ conformation. Also shown are Tyr73 and Asp117.

Fig. 4. Another view of the substrate in the binding cleft after 50 ps of simulation with the unprotonated Asp117 residue, indicating the relationship of the −1 sugar ring to various active site residues. Tyr73 is shown in yellow and Asp117 is shown in red below the oligosaccharides while Asp265 is shown in red above the substrate. Lys259 is shown in green behind the chain; steric crowding by this residue is responsible for tilting the plane of the sugar in the −1 site out of the plane of the chain. Arg78 is shown in green and several water molecules are also included. Hydrogen bonds are indicated by dashed white lines.
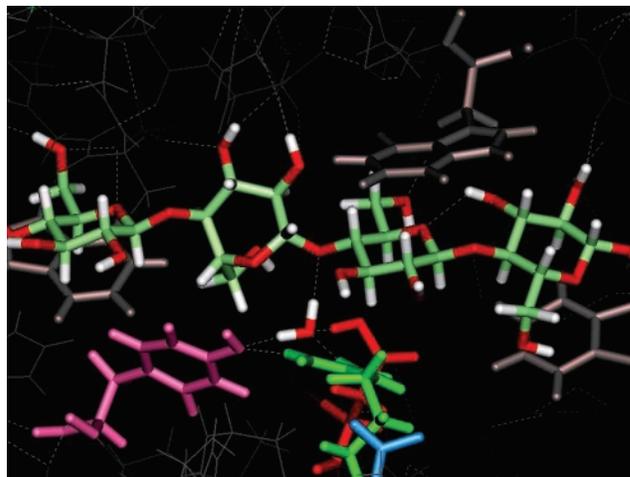
Fig. 5. Another view of the tetrasaccharide bound into the active site, in this case for the simulation in which Asp117, shown in red, is protonated, indicating how one water molecule bridges between this acid group and the glycosidic oxygen atom of the scissile bond. Also shown are Tyr73 in pink and Arg78 in green, with their hydrogen bonds to the same water molecule which serve to localize this molecule strongly in its bridging position. Asp79 is also shown in blue, far from the catalytic site, and Trp 41, 162 and 231, which stack against the hydrophobic faces of the sugar rings, are shown in gray.

were done in triplicate. The activities of E2 and the Ser to Ala mutant enzymes were determined on CMC and hydroxy-ethylcellulose (HEC) overlay assays. Colony overlay assays were done with 0.1% CMC (Hercules) and 0.5% HEC (Sigma) in 0.7% agarose for 2 and 8 h, respectively, and then stained with 1% Congo Red (Wood *et al.*, 1988). Dissociation constants (K_d) for the binding of 4-methylumbelliferyl-β-cellotrioside [MU(glu)$_3$] were determined by direct fluorescence titration using an Aminco SLM 8000C spectrofluorimeter as described previously (Barr *et al.*, 1998).

Results and discussion

Water structure in the active site

In our previous modeling studies, a tetrasaccharide substrate was docked into the E2 binding cleft using energy minimization to optimize the conformations of the enzyme, substrate and their interactions (Taylor *et al.*, 1995). While the starting cellotetraose structure was a planar fragment of cellulose (Aabloo *et al.*, 1994), placing this substrate into the putative binding site caused the sugar ring in the −1 position to tilt significantly away from the planar arrangement (see Figure 1), while the other three rings remained relatively planar. On inspection, this distortion was found to be largely caused by steric repulsions arising from non-bonded clashes with Lys259, and also hydrogen bonding interactions with Tyr73. This tilting
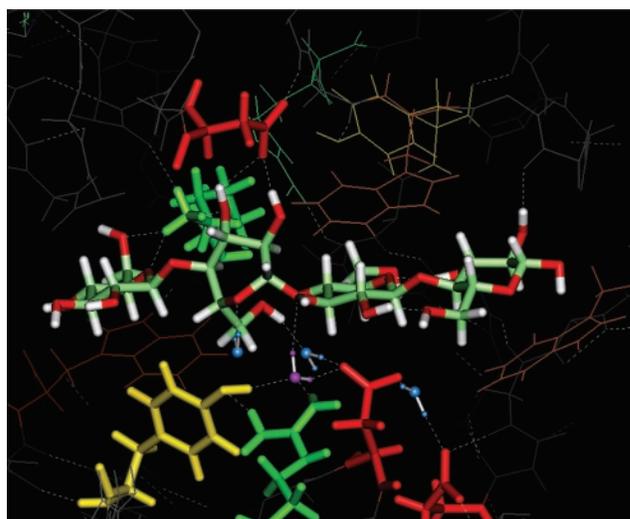
resulted from forcing the glycosidic linkage torsion angles ϕ and ψ away from the cellulose values, but without distorting the conformation of the ring significantly. When the Lys259 side chain was manually moved aside using an interactive graphics program, followed by energy minimization, to allow it to make a hydrogen bond to Asp265 rather than to the substrate, the sugar ring in the −1 position no longer tilted out of the average plane of the other three rings. Tilting of the sugar ring in the −1 position was also observed in recent crystal structures of cellotetraose in the active site of CBH II, but in that case the sugar ring also distorted from the $^4C_1$ chair
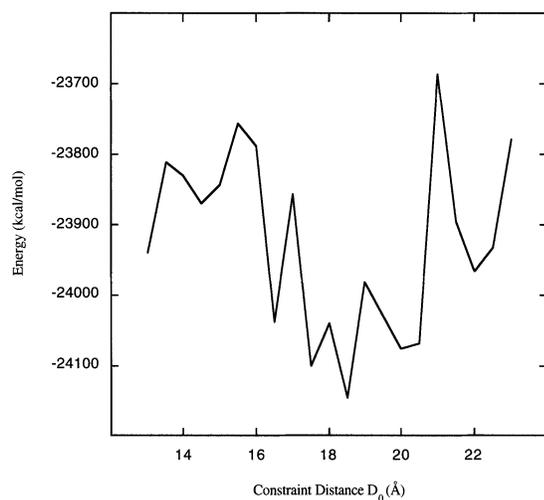
Fig. 6. Plot of the minimized energy of the E2–cellotetraose substrate complex, constrained to move the P1 and P2 lobes to various relative separation distances d_0, plotted as a function of d_0. The constraint energy is not included in the total energy plotted.

Fig. 7. Plot of the energy of the constrained E2–substrate complex as a function of the distance of the closest Asp79 oxygen atom to the glycosidic oxygen atom of the scissile bond. This distance is different from the constraint distance applied to the two lobes since the protein relaxes in a complex fashion at each value of the constraint distance, which is measured between the centers of mass of the two chains.

conformation into a $^2S_0$ skew boat (Zou *et al.*, 1999) (see below). The authors attributed the twisting of the ring to a hydrogen bond formed between CBH II Tyr169 (corresponding to the E2 Tyr73) and the exocyclic primary alcohol hydroxyl group of the glucose monomer in the –1 position. Such a hydrogen bond also exists in our E2–substrate complex model.

In order to investigate whether water molecules bridge between the putative catalytic acid Asp117 and the glycosidic oxygen atom of the bond being hydrolyzed, mediating proton transfer, the structure of the E2–cellotetrose complex was subjected to MD simulations in an aqueous environment, as described in the Computational methods section. Two simulations were performed, one in which Asp117 was unprotonated, followed by a subsequent simulation in which this acid residue was protonated. Early in the data collection phase of this simulation, the tilted glucose ring in the –1 position underwent a conformational change from the $^4C_1$ conformation to a $^2S_0$ skew boat, while remaining tilted out of the plane of the other three rings (see Figure 3). This transition arose spontaneously without any forcing and was subsequently stable throughout the remainder of the simulation. Following this transition, a series of small adjustments in the protein conformation and surrounding aqueous solvent occurred which produced a complex which had a water molecule bridging between the Asp117 residue and the linkage oxygen atom (Figures 4 and 5). The most significant of these conformational shifts brought Arg78 from its crystallographic position in the uncomplexed protein to a position between Tyr73 and Asp117, where the nitrogen atom of its guanidinium group was only 3.92 Å from the nearest carboxyl oxygen atom of the acid group. Concurrently, Tyr73 broke its hydrogen bond to the primary alcohol of the sugar ring in the –1 position and shifted to make a stronger bond with the Arg78 guanidinium group (Figure 3). In the new position, the aromatic ring of Tyr73 experienced increased steric clashes with the sugar primary alcohol group, which was apparently the cause of its ring conformational 'flip'. Following these changes, a water molecule occupied a site intermediate between all three of these residues and the glycosidic linkage, in which it was hydrogen bonded to all four groups (see Figure 5). It made one hydrogen bond as an acceptor to the protonated acid of Asp117, another
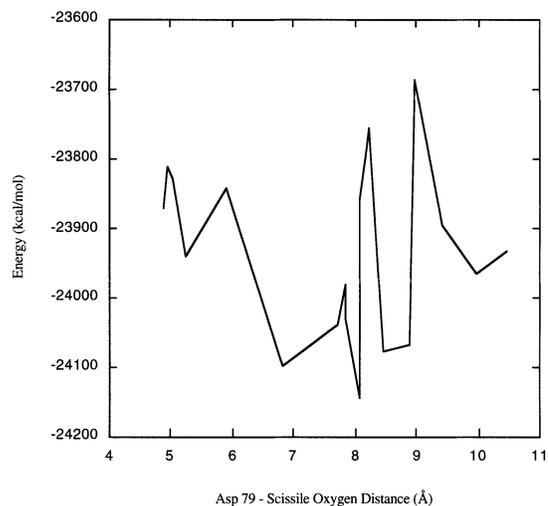
as an acceptor to the Arg residue, a third as a donor to the hydroxyl oxygen of Tyr73 and a fourth directly to the glycosidic oxygen atom of the scissile bond, as shown in Figure 5. The spontaneous appearance of this arrangement suggests that the complex has a role in the catalytic mechanism of the enzyme. The complex persisted throughout a subsequent MD simulation in which Asp117 was protonated, where the water donated a proton in a hydrogen bond to Tyr73 and was the acceptor in hydrogen bonds from Asp117 and Arg78. Each of these three residues is highly conserved in all family 6 proteins, including CBH II, CelA and CenA (Spezio, 1994). This complex could exist whether the Asp117 residue was protonated or unprotonated (Figure 4) because of the ability of the water to reorient and change its hydrogen bonding pattern and because of the capacity of the hydroxyl group on Tyr73 to serve as either a donor or acceptor in a hydrogen bond. It appears that this water molecule serves to transmit a proton from Asp117 to the glycosidic oxygen atom of the scissile bond, bridging the 5.5 Å gap between Asp117 and this oxygen. This water molecule may need to be tightly constrained to keep it properly positioned. The hydrogen bonds to the other two residues would serve to further constrain it even as other parts of the protein shift.

Our previous modeling and mutagenesis studies demonstrated the importance of Tyr73 to catalytic activity (Taylor *et al.*, 1995); replacing it with an Ala or Ser almost totally eliminated hydrolysis, nearly as efficiently as mutating the catalytic acid residue Asp117 (see Table I). In the starting structure of the enzyme–substrate complex, the hydroxyl group of this residue hydrogen bonds to the –1 residue of the substrate through the O6 hydroxyl group of its exocyclic primary alcohol. In the dynamic simulation, however, this residue shifts so as to hydrogen bond to the Arg78 side chain, placing it into steric conflict with the sugar ring. This steric stress is relieved, however, by the distortion of the ring conformation into the twist boat.

Zou *et al.* also found that Tyr169, the analog of this residue in CBH II, is essential for activity as well (Zou *et al.*, 1999). They hypothesized that the hydrogen bond from this residue
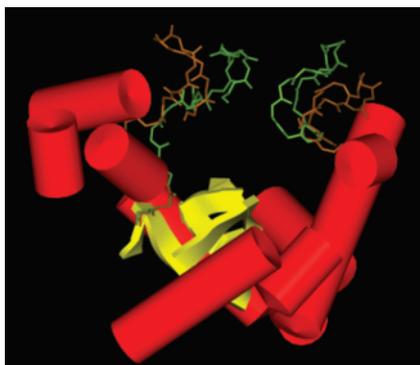
Fig. 8. A Richardson cartoon of E2, showing the secondary structural elements and indicating the initial positions of the two lobes P1 and P2 (see Figure 2) in orange and the lowest energy structure of these same loop chains in green, showing how the motion of the lobes has closed down the binding cleft into a tunnel such as is found in the exocellulase CBH II.

to the exocyclic hydroxymethyl alcohol is necessary to distort the sugar ring in the −1 position into a $^2S_0$ skew-boat. In their experimental studies, a synthetic substrate analog with a xylose residue in the −1 position was still tilted, as was observed for cellotetraose both in their studies and in our modeling, but this ring did not distort into the skew-boat conformation. This result is consistent with our observation that the ring tilt is caused by steric clashes with the Lys259 side chain.

In order to examine the effects of a serine mutation in residue 73 on the enzyme–substrate dynamics, a stochastic boundary simulation of the E2–cellotetraose complex with the Tyr73 Ser mutation was set up and executed like the previous simulation, with a 20 ps equilibration period followed by a 40 ps data collection period. Throughout the course of the simulation, the substrate remained in its starting conformation and in particular the sugar ring in the −1 position did not undergo any ring conformational inversions as observed previously. Because the serine side chain is so much smaller than the tyrosine that it replaced, a layer of water molecules occupied the void previously filled by the phenolic ring in the wild-type structure, effectively shielding the primary alcohol of the −1 sugar ring from direct interaction with the side chain. One water molecule was observed to be strongly localized by the serine side chain approximately where the phenolic group would have otherwise been in the wild-type. It seems likely that this absence of steric clashes in the serine mutant contributes to the drastic decrease in the activity of the mutant observed experimentally (see Table I). Similar mutant simulations were also conducted for the Tyr73 Phe and Tyr73 Glu mutant enzymes. The Tyr73 Glu mutant enzyme also exhibited no sugar ring distortion. This smaller acid side chain did not form a salt bridge with Arg78, in spite of its charge, as it was apparently too short, and instead surrounded itself with water molecules hydrating its acid group. In the Tyr73 Phe mutant simulation there also was no distortion of the −1 sugar ring chair conformation. Perhaps surprisingly, the Tyr73 Phe mutant was found to retain 8% of the wild-type activity (see Table I). The bulky phenylalanine aromatic ring may occasionally clash with the sugar ring enough to distort its conformation, thereby helping to promote cleavage, even though such a ring conformation change was not observed in our short simulation. Without the hydroxyl group of tyrosine the Phe side chain is not drawn toward Arg78 and is therefore less often brought into steric conflict.

Relaxation of the loop containing Asp79

Figure 6 displays the results of the reaction path analysis for shifting the polypeptide segment containing Asp79 toward the catalytic site under the influence of a series of distance constraints as described in the Methods section. Initially the conformational deformation led to a large increase in the total potential energy of the complex, as the molecule shifts away from the local minimum associated with the crystallographic structure. However, after passing over a barrier, the molecule enters into a broad region of conformational space in which the energy of the complex is actually much lower than for the enzyme–tetrasaccharide complex using the crystal conformation. A global minimum along this pathway is reached for $d_0 = 18.5$ Å, giving a complex in which the distance from the closest aspartate oxygen to the glycosidic oxygen of the scissile bond is 8.47 Å. As the constraint distance d_0 is further reduced, the complex goes over another, higher energy barrier before finding another low-energy conformation at 13 Å, which places the Asp79 residue only 5.24 Å away from the scissile glycosidic linkage. This complex, however, is 220 kcal/mol higher in energy than the conformation at $d_0 = 18.5$ Å.

Figure 7 illustrates the same process of constrained movement of these two loops, but with the energy plotted as a function of the distance from the oxygen atom of the scissile glycosidic bond to the closest carboxylic oxygen atom of the Asp79 side chain. The steady decrease in the constraint distance d_0 did not correspond to a uniform and monotonic decrease in the distance from Asp79 to the reaction site, as the overall protein conformation relaxed in various complex ways in response to the constraints. In addition, the energy as a function of this more relevant distance exhibits a different functional dependence. There is an alternative conformation for the enzyme–substrate complex, which is separated from the crystal structure by a barrier, which has a much lower energy than the crystal structure, but this conformation is found in a broad minimum centered around 8.5 Å and its energy climbs steeply as the acid group is brought closer than 6 Å to the reaction site. A distance of 8.5 Å is too far to be bridged by a single water molecule and would require a chain containing two hydrogen-bonded water molecules. The energy again decreases when the separation distance reaches 4.88 Å, but the energy is still 272 kcal/mol higher than the global minimum energy structure and is in fact 57 kcal/mol higher in energy than the complex with the crystallographic conformation. It should be remembered, however, that minimization calculations achieve only a very limited sampling of configuration space owing to the limitations of the multiple-minimum problem (Scheraga, 1983) and the most important result here is that there are many alternative conformations of E2 which bring Asp79 closer to the active site while significantly lowering the energy.

The effect of the restraining forces applied across the binding groove is to draw the two lobes P1 and P2 as shown in Figure 2 closer together. Figure 8 shows a hybrid Richardson cartoon illustration, with these loops shown in atomic detail, of the final minimized conformation corresponding to the lowest energy conformation found. As can be seen, these two lobes have been drawn together to close off the binding groove in much the same way as is seen in the exocellulases such as CBH II, which have a binding tunnel formed by these loops, which are longer in that enzyme. Apparently as a result of these deletions in E2, these loops are only stable in this closed
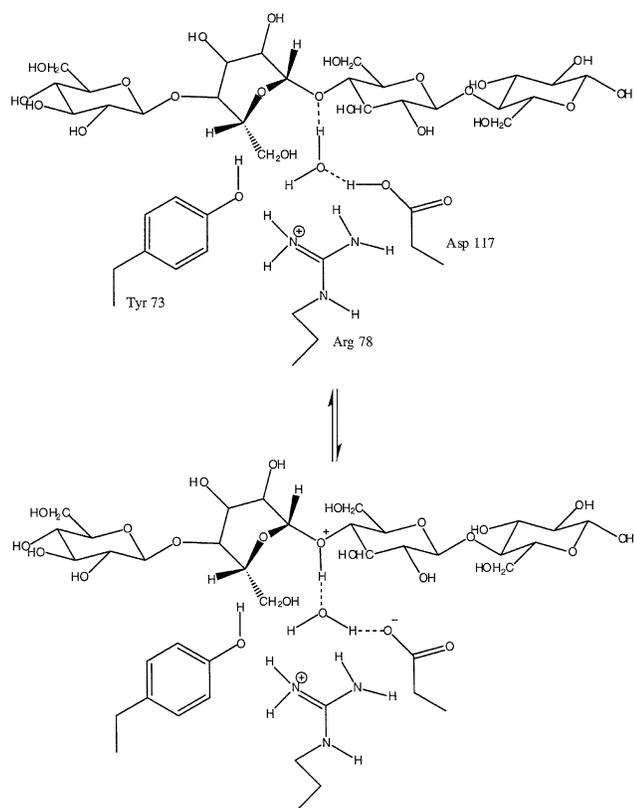
Fig. 10. View of the relative orientation of cellotetraose and several important amino acid side chains for a final dynamics snapshot of a hydrated MD simulation conducted on the lowest energy structure found by pulling the P1 and P2 lobes together. The side chains shown are Asp117 (red), Asp265 (orange), Tyr73 (purple), Arg78 (green) and Asp79 (blue). A water molecule is also shown, hydrogen bonded to Asp79 and the O3 hydroxyl group of the sugar ring in the +1 binding site. Note that this water molecule is positioned below the aliphatic proton of the anomeric carbon of the sugar in the –1 site. If the scissile bond had actually been cleaved in this simulation, this carbon would be an achiral carbocation and this water molecule would be ideally positioned to donate a hydroxyl group to invert the stereochemistry, as is observed in E2. In this model, then, Asp79 would serve as the catalytic base.

Fig. 9. Schematic representation of the proposed mechanism of E2 cellulase, showing cellotetraose with the sugar in the –1 binding site distorted into a skew-boat. The Tyr73, Arg78 and Asp117 side chains are also indicated. In the proposed mechanism, following the shift in conformation of Arg78 and the ring flip in the substrate, a water molecule can comfortably hydrogen bond to the glycosidic oxygen atom, the catalytic acid group Asp117, the hydroxyl of Tyr73 and also to Arg78. This precise positioning allows a facile transfer of a proton from Asp117 to the water and from there to the glycosidic linkage. This step is then followed by the scission of the C1–O glycosidic bond of the sugar in the –1 site, producing an oxocarbonium ion and, in this example, a cellobiose leaving group. The ion would then be attacked by a hydroxyl group produced by the interaction of a water molecule with Asp79, which is positioned on the appropriate side of the binding cleft for this purpose (see Figure 10). The catalytic cycle would then be completed by the transfer of a proton from Asp79 to Asp117 via another bridging water molecule.

down conformation in the presence of substrate, while the longer segments in CBH II keep the active site tunnel closed down even without substrate.

To explore further the stability of these altered conformations, an MD simulation was performed on the minimized conformation found for the structure with the smallest value of d_0, 13Å. In this simulation, the tetrasaccharide substrate used in the earlier search was replaced by the oligosaccharide with the second glucose ring in the $^2S_0$ conformation which arose spontaneously in the simulations of the complex of the crystal structure with substrate. This oligosaccharide was superimposed on to the positions of the previous substrate, which was then removed and the complex was minimized with the harmonic backbone dihedral angle constraints for 100 steepest descent steps. This complex was then subjected to 5 ps of equilibration and 15 ps of dynamics at room temperature, using the same previous harmonic constraints on the backbone dihedral angles, to explore the nearby conformation space. The final coordinate frame from this simulation was then minimized as before and found to have the lowest energy of
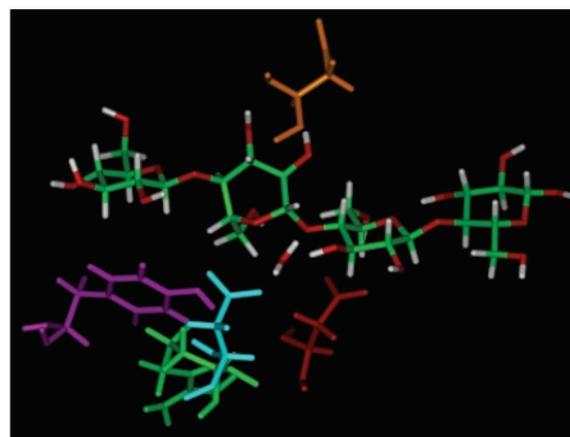
any structure found in these studies, 349 kcal/mol lower than the minimum on the reaction path curve in Figure 6 and 561 kcal/mol lower than the minimized crystal structure. When a completely unconstrained standard MD simulation at room temperature (300 K) was performed on this minimized enzyme–substrate complex, this 'closed' conformation was found to be stable during 100 ps of simulation time.

Activities of the Arg78 and Ser85 mutant enzymes

The activities of the Arg78 mutant enzymes are given in Table I. Mutation of Arg78 to Ala caused a significant loss of activity on all substrates, while mutation to Lys only reduced the activity from 2- to 6-fold. These results are consistent with the modeling results and show that Arg78 is a very important, but not absolutely essential, residue. Ser85 is the equivalent of Ser181 in CBH II (*Trichoderma reesei* Cel6A), which has been suggested as a possible residue that might hydrogen bond to the catalytic water (Zou *et al.*, 1999). Ser85 and the adjacent Ser84 in E2 were mutated to Ala both separately and together and all of the mutant enzymes retained wild-type activity, as tested by CMC and HEC overlay assays. These results show that these two residues have no catalytic function in E2 and possibly do not function in catalysis in other family 6 enzymes.

Proposed mechanism

The results presented here from both the simulations and the mutagenesis experiments allow us to propose a relatively complete model for the E2 mechanism that appears to explain all of the principal experimental facts. This model is based on the assumption of a general conformational change in the protein upon binding of substrate, which brings Asp79 close to the catalytic site. The mechanism may involve a two-step process with the conformational change being the second step in the process or the entire sequence may happen as a concerted and basically simultaneous process. For the sake of clarity, the mechanism will be presented here as a two-step sequence.

This approach also parallels the calculations that were performed, since we separately studied the structuring of water around the substrate-bound complex and the energetics of moving Asp79 towards the active site. The proposed mechanism is illustrated schematically in Figure 9.

In the proposed mechanism, Lys259 serves to twist the sugar ring in the –1 position out of the plane of the cellulosic ribbon, but without necessarily distorting the sugar ring away from the $^4C_1$ conformation. From this role, one would predict that mutation of this residue to a less bulky side chain would relieve steric clashes and significantly reduce activity. In fact, the Ala259 mutant enzyme has less than 1% of wild-type E2 activity. The Lys ε-amino group makes three hydrogen bonds to the substrate in the model enzyme–substrate complex and elimination of these hydrogen bonds by mutation produces a drastic reduction in binding.

The Tyr73 residue plays two important roles in the proposed mechanism, consistent with the experimental observation that replacement of this residue with either Ser or Ala decreases activity by as much as mutation of the catalytic acid residue (see Table I). The first of these roles is the same as that proposed for Tyr169 of CBH II, to distort the sugar ring of the substrate residue in the –1 position into a twist-boat conformation (Zou *et al.*, 1999). Such a conformational distortion has long been proposed in generalized glycosidase mechanisms and has been observed experimentally in lysozyme (Strynadka and James, 1991). The second, less crucial, function of Tyr73 is to help position a water molecule between the catalytic acid and the glycosidic oxygen atom of the scissile bond by hydrogen bonding to it, thus helping to prevent its escape into the bulk solution before it has transferred its newly acquired proton to the glycosidic oxygen. These two roles seem to be consistent with the observed mutagenesis data. Changing Tyr73 to Phe reduces activity somewhat but does not eliminate it, while replacing this residue with Ser or Glu substantially decreases activity (Table I). These results imply that the bulk of the aromatic ring is required for simple steric repulsion to distort the sugar ring, but hydrogen bonding between residue 73 and the substrate is not critical. Similarly, the hydrogen bond to the water molecule helps to facilitate proton transfer by precisely positioning it, even though this molecule can occupy the necessary position without the additional hydrogen bond. The serine is too far from the proper position to help localize the water molecule and could even make matters worse by offering the hydronium ion an alternative site to form a hydrogen bond that is too far from the scissile bond for efficient proton transfer.

Asp117 is the catalytic acid, in spite of its distance from the glycosidic oxygen atom, since the mutagenic studies clearly indicate that it is essential. It accomplishes proton donation to the scissile bond by transferring the proton to the water molecule that bridges between Asp117 and the glycosidic oxygen. The precise positioning of this water molecule is facilitated by its other hydrogen bonds to Arg78 and Tyr73. Before the proton is transferred, this water molecule is hydrogen bonding as an acceptor with Asp117 and as a donor to the glycosidic oxygen. Since Arg78 can only serve as a hydrogen bond donor, the water molecule must be the acceptor in that bond and must therefore be the donor in the hydrogen bond to the hydroxyl group of Tyr73. The proposed mechanism in part requires Tyr73 and Arg78 to keep the water molecule precisely positioned for proton transfer, although their role is not as essential as that of Asp117. The presence of Arg78 also facilitates the transfer of the proton to the water molecule by shifting the pK_a of Asp117, since its positive charge stabilizes the carboxyl group, which is only 3.92 Å away, after the proton is lost. The experimental result that mutating Arg78 to Lys only reduces activity would be consistent with this model, since neither function would be absolutely essential and lysine could partially fill both roles, particularly that of shifting the pK_a.

After protonation of the glycosidic oxygen atom, bond scission occurs, producing an sp^2-hybridized carbocation in the –1 ring and a cellobiose leaving group in the +1 and +2 subsites. Since modeling of alternative low-energy conformations for the protein finds a position for the enzyme–substrate complex in which Asp79 is only about 5.5 Å away from the scissile bond of the uncleaved oligosaccharide and, most importantly, positioned above the cleft and 'below' the sugar ring of the carbocation in the –1 position (see Figure 10), it is reasonable to postulate that this residue is the catalytic base. Its location would explain how E2 inverts the product anomeric configuration and this role would be consistent with the experimental observation of a reduction in catalytic activity when it is mutated. In this model, then, the carboxyl group abstracts a proton from yet another water molecule located 'below' the carbon ion and the resulting hydroxyl group is captured by the sugar ion. The cycle is then completed and the enzyme primed for another bond scission by the transfer of the proton from Asp79 to Asp117 via water molecules.

Conclusions

The mechanism proposed here explains the results of previously reported site-directed mutagenesis experiments which identify Asp79 as an important residue even though it is so far from the active site that it is difficult to develop a mechanistic explanation based upon the crystal structure. Since this observation forced the conclusion that the enzyme must undergo a conformational shift, the identification of alternative conformations for the enzyme–substrate complex which are lower in energy than the crystal structure and which place the acid group close to the active site are indeed encouraging. Because of the limitations of the multiple minimum problem, it is unlikely that these minimization and limited MD simulations have identified the lowest energy structure for this protein–substrate complex, but the results clearly demonstrate that the postulated conformational transitions for the Asp79 loop are plausible. The simulations of the enzyme–substrate complex in an aqueous environment confirm how the other residues identified as important can play their roles via hydrogen-bonded water molecules even though they are too far from the scissile bond to donate directly a proton or hydroxyl group. Furthermore, the observation that the –1 sugar ring of the substrate is distorted is consistent with the mechanisms of several other hydrolases.

Acknowledgments

References

Aabloo,A., French,A.D., Mikelsaar,R.-H. and Pertsin,A.J. (1994) *Cellulose*, **1**, 161–168.

Barr,B.K., Wolfgang,D.E., Piens,K., Claeyssens,M. and Wilson,D.B. (1998) *Biochemistry*, **37**, 9220–9229.

Brooks,C.L., Brünger,A. and Karplus,M. (1985) *Biopolymers*, **24**, 843–865.

Brünger,A., Brooks,C.L. and Karplus,M. (1984) *Chem. Phys. Lett.*, **105**, 495–500.

Damude,H.G., Withers,S.G., Kilburn,D.G., Miller,R.C. and Warren,R.A.J. (1995) *Biochemistry*, **34**, 2220–2224.

Duan,Y. and Kollman,P.A. (1998) *Science*, **282**, 740–744.

Ghangas,G.S. and Wilson,D.B. (1988) *Appl. Environ. Microbiol.*, **54**, 2521–2526.

Ghose,T.K. (1987) *Pure Appl. Chem.*, **59**, 257–268.

Henrissat,B., Teeri,T.T. and Warren,R.A.J. (1998) *FEBS Lett.*, **425**, 352–354.

Himmel,M.E., Ruth,M.F. and Wyman,C.E. (1999) *Curr. Opin. Biotechnol.*, 10, 358–364.

Irwin,D.C., Spezio,M., Walker,L.P. and Wilson,D.B. (1993) *Biotechnol. Bioeng.*, **42**, 1002–1013.

Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) *J. Chem. Phys.*, **79**, 926–935.

Koivula,A. *et al.* (1996) *Protein Eng.*, **9**, 691–699.

MacKerell,A.D. *et al.* (1998) *J. Phys. Chem. B*, **102**, 3586–3616.

Palma,R., Zuccato,P., Himmel,M.E., Liang. G. and Brady,J.W. (2000) In Himmel,M.E. (ed.), *Glycosyl Hydrolases in Biomass Conversion*. American Chemical Society, Washington, DC, pp. 112–130.

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

Scheraga,H.A. (1983) *Biopolymers*, **22**, 1–14.

Spezio,M.L.R. (1994) PhD Thesis, Cornell University.

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

Strynadka,N.J.C. and James,M.N.G. (1991) *J. Mol. Biol.*, **220**, 401–424.

Taylor,J.S.,B. Teo, Wilson,D.B. and Brady,J.W. (1995) *Protein Eng,*, **8**, 1145–1152.

van Gunsteren,W.F. and Berendsen,H.J.C. (1977) *Mol. Phys.*, **34**, 1311–1327.

Wolfgang,D.E. (1998) PhD Thesis, Cornell University.

Wolfgang,D.E. and Wilson,D.B. (1999) *Biochemistry*, **38**, 9746–9751.

Wood,P.J., Erfle,J.D. and Teather,R.M. (1988) *Methods Enzymol.*, **160**, 59–74.

Zhang,S. and Wilson,D.B. (1997) *J. Biotechnol.*, **57**, 101–113.

Zhang,S., Barr,B.K. and Wilson,D.B. (2000a) *Eur. J. Biochem.*, **267**, 244–252.

Zhang,S., Irwin,D.C. and Wilson,D.B. (2000b) *Eur. J. Biochem.*, **267**, 3101–3115.

Zou,J.-Y., Kleywegt,G.J., Ståhlberg,J., Driguez,H., Nerinckx,W., Claeyssens,M., Koivula,A., Teeri,T.T., Jones,T.A. (1999) *Structure*, **7**, 1035–1045.