

4-29-2009

Mechanism of Cellulose Hydrolysis by Inverting GH8 Endoglucanases: A QM/MM Metadynamics Study


Luis Petersen
Iowa State University

Albert Ardèvol
Parc Científic de Barcelona and Universitat de Barcelona

Carme Rovira
Universitat de Barcelona and Institució Catalana de Recerca i Estudis Avançats

Peter J. Reilly
Iowa State University, reilly@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/cbe_pubs

 Part of the [Biochemical and Biomolecular Engineering Commons](#), and the [Biological Engineering Commons](#)

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/cbe_pubs/14. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Article is brought to you for free and open access by the Chemical and Biological Engineering at Digital Repository @ Iowa State University. It has been accepted for inclusion in Chemical and Biological Engineering Publications by an authorized administrator of Digital Repository @ Iowa State University. For more information, please contact digirep@iastate.edu.

Mechanism of Cellulose Hydrolysis by Inverting GH8 Endoglucanases: A QM/MM Metadynamics Study

Luis Petersen,[†] Albert Ardèvol,^{‡,§} Carme Rovira,^{‡,§,¶} and Peter J. Reilly^{*,†}

Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa 50011, Computer Simulation and Modeling Laboratory (CoSMoLab), Parc Científic de Barcelona, 08028 Barcelona, Spain, Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, 08028 Barcelona, Spain, and Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

Received: December 29, 2008; Revised Manuscript Received: March 18, 2009

A detailed understanding of the catalytic strategy of cellulases is key to finding alternative ways to hydrolyze cellulose to mono-, di-, and oligosaccharides. Endoglucanases from glycoside hydrolase family 8 (GH8) catalyze the hydrolysis of β -1,4-glycosidic bonds in cellulose by an inverting mechanism believed to involve an oxacarbenium ion-like transition state (TS) with a boat-type conformation of the glucosyl unit in subsite -1. In this work, hydrolysis by *Clostridium thermocellum* endo-1,4-glucanase A was computationally simulated with quantum mechanics/molecular mechanics metadynamics based on density functional theory. Our calculations show that the glucosyl residue in subsite -1 in the Michaelis complex is in a distorted ${}^2S_0/{}^{2.5}B$ ring conformation, agreeing well with its crystal structure. In addition, our simulations capture the cationic oxacarbenium ion-like character of the TS with a partially formed double bond between the ring oxygen and C5' carbon atoms. They also provide previously unknown structural information of important states along the reaction pathway. The simulations clearly show for the first time in GH8 members that the TS features a boat-type conformation of the glucosyl unit in subsite -1. The overall catalytic mechanism follows a $D_N^*A_N$ -like mechanism and a β - ${}^2S_0 \rightarrow {}^{2.5}B$ [TS] $\rightarrow \alpha$ - 5S_1 conformational itinerary along the reaction coordinate, consistent with the anti-periplanar lone pair hypothesis. Because of the structural similarities and sequence homology among all GH8 members, our results can be extended to all GH8 cellulases, xylanases, and other endoglucanases. In addition, we provide evidence supporting the role of Asp278 as the catalytic proton acceptor (general base) for GH8a subfamily members.

1. Introduction

Glycoside hydrolases (GHs) catalyze the hydrolysis of glycosidic bonds that join either two carbohydrate moieties or a carbohydrate glycon and a noncarbohydrate aglycon. At present, GHs are divided into more than 100 families¹ based on primary sequence data. Known tertiary structures exist in most families, many complexed with their substrates or inhibitors. Despite this wealth of structure–function information, further supported by mechanistic studies from kinetic isotope effects, computational simulations, and enzyme kinetics,^{2–4} there are still several mechanistic aspects of GHs that are not fully understood and are being researched.³

There is overwhelming evidence that GHs distort the glycon sugar ring bearing the scissile glycosidic bond and further change the ring conformation along the reaction coordinate.^{5,6} The conformational interconversions that occur in the active sites of specific GHs have been the subject of many recent studies.^{7–14} Nevertheless, direct evidence of conformational itineraries is still missing for many GH families. Furthermore, detailed structural features of the enzymatic transition state (TS) in GHs are still fairly unknown. This issue has been difficult to address because of the extremely short life of an oxacarbenium ion-like TS, which makes it very difficult to obtain detailed

structural insight with experimental methods. Most of the information that we have about TS structures in GHs comes from kinetic isotope effect measurements and structural information of GHs complexed with TS mimics.^{12,15,16} Fortunately, reliable computational techniques can be exploited to explicitly model reactions inside GHs. To date, the reaction mechanism in GHs has been modeled only for GH22 lysozyme and GH2 β -galactosidase by first-principles approaches, providing evidence of a covalent glycosyl–enzyme intermediate for these retaining GHs.^{17–19}

Cellulases, found in many GH families, have received much attention for their connection with the biofuels industry. The slow degradation of cellulose, as well as that of other plant cell-wall components like hemicellulose and lignin, is one of the principal impediments to competitively produce fuel from biomass.^{20–22} Cellulase catalytic mechanisms have attracted interest because these enzymes have been optimized by many years of evolution to hydrolyze cellulose to mono-, di-, and oligosaccharides. A deep understanding of the catalytic strategy of these enzymes would be very valuable in pursuing viable alternatives to perform cellulose hydrolysis.

GH family 8 (GH8) enzymes catalyze the hydrolysis of β -1,4-glycosidic bonds present in cellulose, chitosan, and xylan. The GH8 active site is a cleft that allows substrate binding and subsequent bond cleavage at interior chain positions. GH8 belongs to clan M (GH-M), along with GH48, with their members having $(\alpha/\alpha)_6$ tertiary folds.^{1,23}

GH8 enzymes use the inverting mechanism (Figure 1). Two carboxylate amino acid residues, either aspartate or glutamate,

* Corresponding author: Telephone: +1-515-294-5968. Fax: +1-515-294-2689. E-mail: reilly@iastate.edu.

[†] Iowa State University.

[‡] CoSMoLab.

[§] IQTCUB.

[¶] ICREA.

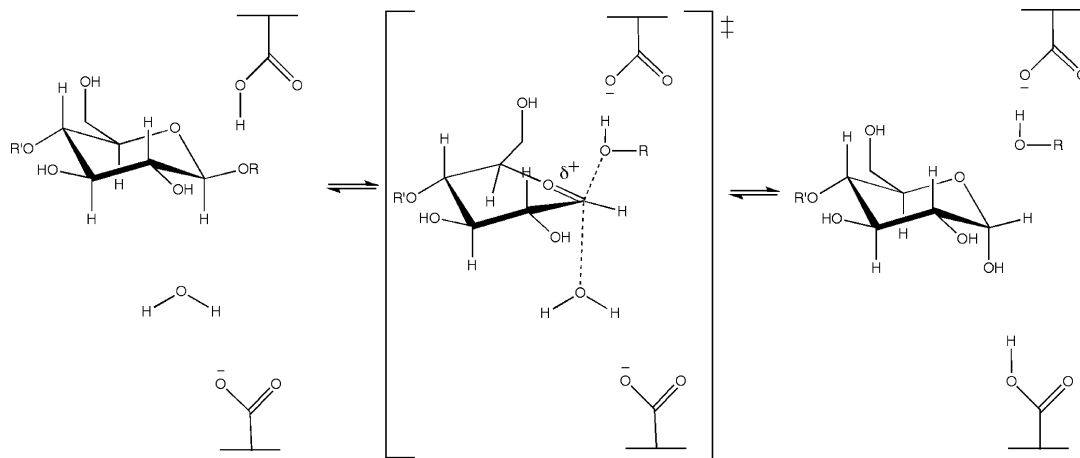


Figure 1. General reaction mechanism of inverting GHs.

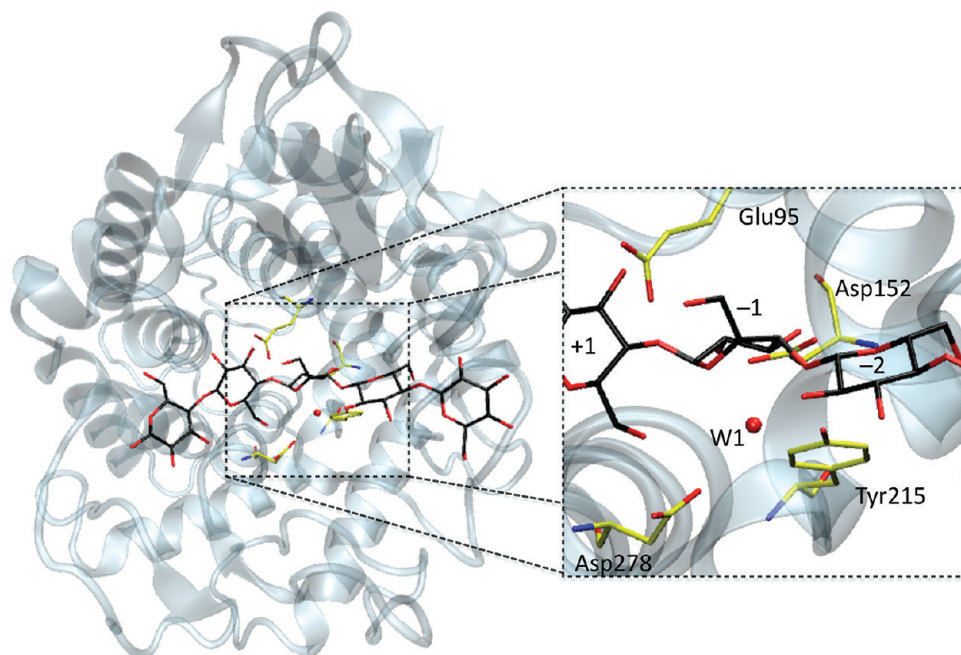


Figure 2. Active-site structure of *C. thermocellum* GH8 EG (PDB entry 1KWF). Red spheres are water molecules. The Gln95 residue of the crystal structure has been back-mutated to Glu. For the sake of clarity, the protein is oriented with the negative glucosyl subsites (−1, −2 and −3) at the right-hand side (i.e., different from the general scheme of Figure 1).

are employed by most inverting GHs to catalyze hydrolysis, one protonating the scissile glycosidic oxygen atom (O4) and the other coordinating the nucleophile, a water molecule, to assist its deprotonation to complete the reaction.²⁴ The proton donor and putative proton acceptor (general base) were identified based on structural information as Glu95 and Asp278, respectively,²⁵ in *Clostridium thermocellum*. However, the identification of Asp278 as the proton acceptor has been controversial in light of recent kinetic studies that show that some GH8 members, in which residues analogous to Asp278 have been mutated, still retain significant activity.^{26,27} Also, a phylogenetic study showed that the putative proton acceptor is not strictly conserved in GH8; instead its location shifts within the active site.²⁸

Guerin and co-workers published a high-resolution crystal structure, PDB 1KWF, of a mutated (E95Q) GH8 endoglucanase (EG) from the *C. thermocellum* cellulosome (Figure 2).²⁵ A cellopentaose molecule is bound in a groove-shaped active site, characteristic of EGs, spanning subsites −3, −2, −1, +1, and +2. The active site forces the cellopentaose chain to kink, and

it distorts the glucosyl residue in subsite −1 away from the ground-state 4C_1 conformation to a ${}^2S_0/{}^2.5B$ ring conformation, while the other four glucosyl residues remain in the relaxed 4C_1 conformation. The catalytic proton donor, Glu95, is anti positioned (i.e., pointing toward the O4 lone pair that is at the opposite side to the ring O5' atom).²⁹ The putative catalytic proton acceptor, Asp278, is located below the average plane of the glucosyl ring occupying subsite −1, hydrogen-bonding the nucleophilic water molecule (Figure 2). Tyr215 forms a second hydrogen bond with this water molecule, fixing it in place for optimal catalysis.^{25,26} Asp152 forms a bidentate interaction with the C2'−OH and C3'−OH hydroxyl groups of the glucosyl residue in subsite −1.

The finding of a ${}^2S_0/{}^2.5B$ type of distortion in the Michaelis complex of GH8 EG, together with that of a stable ${}^2.5B$ -glycosyl–enzyme intermediate in GH11,³⁰ led to the suggestion that these GHs operate via a ${}^2.5B$ -type TS.^{4,5} In this work, this prediction is tested by first-principles quantum mechanics/molecular mechanics (QM/MM) simulations of cellulose hydrolysis catalyzed by GH8 EGs. The simulations were meant

to gain insight into the enzyme catalytic mechanism, the structural features of the reaction TS, and the interconversions of the glucosyl unit occupying subsite -1 along the reaction coordinate. Also, further evidence that supports the role of Asp278 as the proton acceptor is presented. To the best of our knowledge, the hydrolysis reaction of an inverting GH has not previously been investigated by first-principles methods.

2. Computational Methods

2.1. Classical Molecular Dynamics. The structure of the GH8 EG-cellopentaose complex described above was submitted to classical molecular dynamics (MD) simulations with Amber9 software.³¹ The FF99SB Amber³² and GLYCAM06 force fields³³ were used to model the protein and cellopentaose substrate, respectively. The E95Q mutation of the crystal structure was manually changed back to glutamic acid. The crystal structure used, PDB 1KWF, contains two conformations. The conformation that contains the cellopentaose substrate spanning subsites -3 to +2 was used for this study. The protonation states of ionizable amino acids were selected based on a visual inspection of their microenvironments. The protonation states of Glu95 and Asp278 were chosen based on the common knowledge of the enzymatic inverting mechanism, where the proton donor begins the reaction in a neutral state and the proton acceptor must be deprotonated (Figure 1). Moreover, this choice led to good agreement between the computed and experimental structure at the active-site region. Asp152 was modeled in its deprotonated state to correctly simulate the two hydrogen bonds with hydroxyl groups of the substrate in subsite -1 (C2'-OH and C3'-OH) observed in the crystal structure. The rest of the aspartate and glutamate residues were deprotonated. All seven histidine residues, none of which are close to the enzyme active site, were modeled in their neutral states. The water molecules observed in the crystal structure were retained, and additional water molecules were added to fully solvate the protein, creating a water box with a 10 Å cushion around the protein surface. The water molecules were modeled with the TIP3P force field.³⁴ Seven sodium ions were added to neutralize the system. The MD simulation was carried out in several steps. First, the system was subjected to energy minimization with fixed protein and substrate, and then the constraints were removed and the whole system was minimized. To gradually reach the desired temperature of 300 K, weak spatial constraints were initially added to the protein and substrate while the water molecules and sodium ions were allowed to move freely. Finally, all constraints were removed, and the MD simulation was extended to 1 ns when the system reached equilibrium. The structure was analyzed with the VMD package.³⁵ A snapshot of the equilibrated system was used as a starting point for further QM/MM simulations.

2.2. QM/MM Molecular Dynamics. The method developed by Laio et al.³⁶ was used. This method combines the first-principles MD method of Car and Parrinello (CPMD)³⁷ with a force-field MD methodology (i.e., QM/MM CPMD). In this approach, the system is partitioned into QM and MM regions. The dynamics of the atoms in the QM region depends on the electronic density, $\rho(r)$, computed with density functional theory (DFT), whereas the dynamics of the atoms in the MM region is ruled by an empirical force field. The QM/MM interface is modeled by the use of link-atom pseudopotentials that saturate the QM region.³⁸ The electrostatic interactions between the QM and MM regions are handled via a fully Hamiltonian coupling scheme,³⁶ where the short-range electrostatic interactions between the QM and the MM regions were explicitly taken into account for all atoms. An appropriately modified Coulomb potential was used to ensure that no unphysical escape of the

electronic density from the QM to the MM region occurs. The electrostatic interactions with the more distant MM atoms are treated via a multipole expansion. Bonded and van der Waals interactions between the QM and the MM regions are treated with the standard Amber force field. Long-range electrostatic interactions between MM atoms are described with P3M implementation,³⁹ using a $64 \times 64 \times 64$ mesh.

An accurate description of energetic, dynamic, and structural features of biological systems has been previously obtained with this methodology, confirming its reliability (see refs 40 and 41 and references therein).

The QM region was confined in an isolated $17.3 \times 19.1 \times 13.3$ Å supercell that included 66 atoms, consisting of the side chains of Glu95 and Asp278, capped at the C α with a link-atom pseudopotential, and the glucosyl residues in subsites -1 and +1 capped at the C1 of subsite -2 and C4 of subsite +2. The nucleophilic water molecule was also included. Kohn-Sham orbitals were expanded in a plane-wave basis set with a kinetic energy cutoff of 70 Ry. Ab initio pseudopotentials generated by the Troullier-Martins scheme⁴² were used. The Perdew-Burke-Ernzerhoff functional⁴³ in the generalized gradient-corrected approximation of DFT was selected. This functional was chosen based on its reliability in describing hydrogen bonds,⁴⁴ and it is the one that we used in our previous QM/MM work on a retaining GH.⁸ A constant temperature of 300 K was reached by coupling the system to a Nosé-Hoover thermostat⁴⁵ of 3500 cm⁻¹ frequency.

Structural optimizations were done using QM/MM MD with annealing of electronic velocities until the maximal component of the nuclear gradient was 10^{-4} au. The time step used for the annealing simulations was 0.072 fs, and the fictitious mass of the electrons was 1000 au. For the MD simulation coupled with a thermostat, a time step of 0.12 fs and a fictitious electron mass of 700 au were used. The TS electronic structure was analyzed in terms of localized orbitals according to the Wannier functions methodology.⁴⁶

Structural optimizations to determine the identity of the catalytic proton acceptor were carried out in a larger $19.0 \times 21.7 \times 20.8$ Å QM box. This included the side chains of Glu95 and Asp278; Tyr215, which coordinates the nucleophilic water molecule; Asp152, which interacts with the C2'-OH and C3'-OH hydroxyl groups; seven water molecules; and the glycon and aglycon. All the amino acids were capped with link-atom pseudopotentials at the C α position.

2.3. QM/MM Metadynamics. One of the main advantages of ab initio MD is the possibility of modeling chemical reactions. Nevertheless, present computational power allows for only picoseconds of ab initio QM/MM MD simulations of large systems such as proteins, not long enough to observe spontaneous reactions with high-energy barriers. The metadynamics technique⁴⁷ is a novel methodology that can be used to induce the reaction by selecting a set of collective variables that include the relevant modes of the TS of the simulated reaction. The collective variables can be any function of the nuclear coordinates of the system as long as they are able to distinguish the different states of the system (i.e., reactants, TS, and products). In this technique, small repulsive Gaussian-like potentials are added to the regions of space that were already visited, preventing the system from exploring the same regions and, thereby, allowing the system to overcome significant energy barriers. The small repulsive potential terms are used to construct the free-energy surface of the system, which can be estimated as the negative of the sum of the Gaussian potential terms.⁴⁷⁻⁴⁹

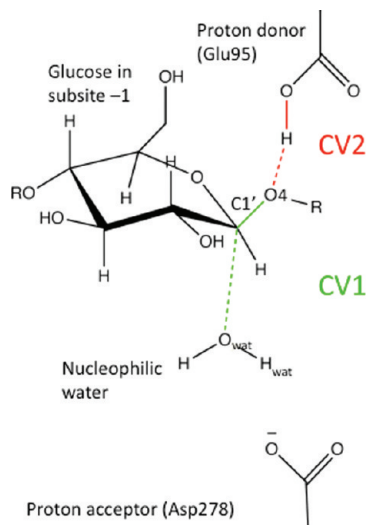


Figure 3. Collective variables (CV1 in green and CV2 in red) used to model hydrolysis of the glycosidic bond catalyzed by *C. thermocellum* GH8 EG (see definition in the text).

This method has recently been applied to a variety of problems in biophysics, chemistry, and material science (see e.g. refs 50–54).

Two collective variables, CV1 and CV2, were chosen to model the hydrolysis of the β -1,4-glycosidic bond between two glucosyl residues; each of them was a difference of coordination numbers (CNs). The CN between two atoms, CN_{ij} , indicates whether a covalent bond exists (1 = bond, 0 = no bond). CV1 is the difference $CN_{C1',O4} - CN_{C1',O_{wat}}$ (Figure 3). Thus, CV1 measures the formation of the α -glycosidic bond between the C1' atom and the incoming water molecule's oxygen atom with the simultaneous rupture of the β -glycosidic bond. CV2 is the difference $CN_{H_{Glu95},O_{Glu95}} - CN_{H_{Glu95},O4}$ (Figure 3). Thus, CV2 measures the proton transfer between the catalytic proton donor and the O4 atom. The CN between two atoms i and j is defined⁴⁷ by

$$CN_{ij} = \frac{1 - (d_{ij}/d^0)^p}{1 - (d_{ij}/d^0)^{p+q}} \quad (1)$$

where d_{ij} is the internuclear distance of the atoms involved, d^0 is the threshold distance for bonding, and p and q are exponents that determine the steepness of the decay of CN_{ij} with respect to d_{ij} . The selected values were $d^0 = 3.03$ au, $p = 12$, and $q = 2$ for CV1 and $d^0 = 2.08$ au, $p = 16$, and $q = 3.6$ for CV2. It is important to note that this choice of collective variables does not introduce any bias concerning the type of nucleophilic substitution mechanism. In other words, it does not dictate whether the mechanism is A_ND_N (fully concerted mechanism), $D_N^*A_N$ (dissociative mechanism, i.e., departure of the leaving group prior to attack by the nucleophile), or $D_N + A_N$ (dissociative mechanism with the presence of an intermediate carbocation).^{55,56}

The metadynamics simulations were done within the Car–Parrinello approach.^{37,47} In this scheme, the Car–Parrinello (CP) Lagrangian is extended by extra terms describing the fictitious dynamics of the collective variables. These additional fictitious particles are coupled through a harmonic potential to the value of the selected collective variables in the real system.⁴⁷ The mass of this fictitious particle and the force constant of the coupling potential were tested to ensure that the coupled particle naturally follows the value of the associated collective variable in the real system. The selected mass values of the fictitious

particles were 15 and 5 amu for CV1 and CV2, respectively, while those of the force constant were 1.5 and 1 au for CV1 and CV2. The height of the Gaussian terms was $1.5 \text{ kcal mol}^{-1}$, which ensures sufficient accuracy for the reconstruction of the free-energy surface. The width of the Gaussian terms was 0.05 Å, according to the oscillations of the selected collective variables observed in a free dynamics. A new Gaussian-like potential was added every 400 MD steps.

3. Results and Discussion

3.1. Classical and QM/MM Molecular Dynamics Simulations. A model of the fully solvated wild-type enzyme–substrate complex at its natural conditions (300 K and 1 atm) was obtained by performing a classical MD simulation. The overall structure after the simulation is very similar to the original crystal structure, with an overall rmsd of 0.60 Å. The length of the hydrogen bond between Glu95, the catalytic proton donor, and the glycosidic oxygen atom is 1.87 Å. Hydrogen bonds with Asp278, the putative proton acceptor, and Tyr215 fix the nucleophilic water molecule in place, and the bonds are 2.2 Å and 3.7 Å long, respectively. The oxygen atom of the nucleophilic water molecule is 3.79 Å below the C1' atom. The glucosyl ring in subsite –1 remains in its distorted $^2S_0/^{2.5}B$ conformation, and the nucleophilic water molecule stays underneath the C1' atom all through the classical MD run. The simulation was then continued by QM/MM for 1.5 ps; during this time no significant conformational change was observed.

3.2. QM/MM Metadynamics Simulation. The system behavior during the metadynamics simulation is displayed in Figure 4, which shows the variation of the main distances characterizing the reaction. First explored was the reactant state with the proton donor in its neutral (protonated) state. During the first 7 ps of the simulation, the proton moved back and forth from Glu95 to the O4 atom several times (the black line). At ~ 7.2 ps, with the O4 atom protonated, the nucleophilic attack occurred. The β -glycosidic bond (the blue line) lengthened while the distance between the C1' atom and the oxygen atom of the incoming water molecule (the red line) shortened to within bonding distance. Also, one of the hydrogen atoms of the water molecule transferred to Asp278 (the green line).

After the forward reaction occurred, the product state was explored for ~ 24 ps. At ~ 15 ps, Arg281 pulled the deprotonated Glu95 toward it. This interaction is probably a strategy of the enzyme to prevent the reverse reaction by positioning the catalytic proton donor in an unproductive local minimum while it is deprotonated. Therefore, a constraint in the CP Lagrangian⁵⁷ forcing the Arg281...Glu95 distance to increase was applied for < 1 ps, after which Glu95 returned to its original position. While exploring the product state, the activated proton moved between the O4 atom and Glu95 several times (the gray and black lines) until the reverse reaction occurred at ~ 31 ps. The simulation was continued until ~ 33 ps when no further significant change in the free-energy surface (described below) was observed.

3.3. Reaction Free-Energy Landscape and Minimal-Energy Pathway. Figure 5 shows the free-energy surface of the GH8 EG reaction. The axes in the figure represent the two reaction coordinates used (CV1 and CV2). The minimal-energy pathway from the reactants to the products traverses the reactant state, the protonated glycosidic oxygen atom, the TS, and the product state. When the system is in the reactant state the values of CV1 and CV2 are ~ 0.8 each. The values of CV1 and CV2 for the product state are ~ -0.8 each.

The reactant well is approximately 36 kcal mol^{-1} below the TS conformational free energy, which is the maximal-energy

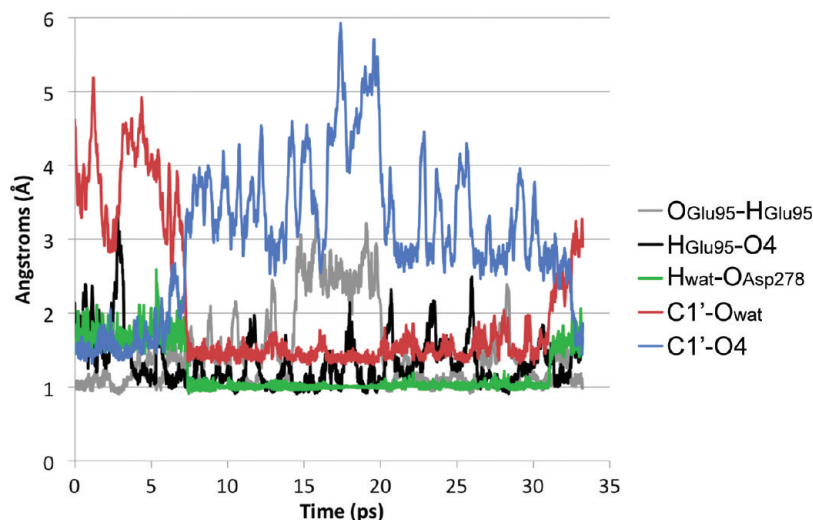


Figure 4. Time evolution of some important distances along the metadynamics simulation. Glu95 is the proton donor. O4 is the glycosidic oxygen atom. H_{wat} is the nearer proton of the nucleophilic water molecule to Asp278, the putative proton acceptor. C1' is the glycon anomeric oxygen atom. The forward reaction occurs at ~ 7.2 ps, and the reverse reaction occurs at ~ 31 ps.

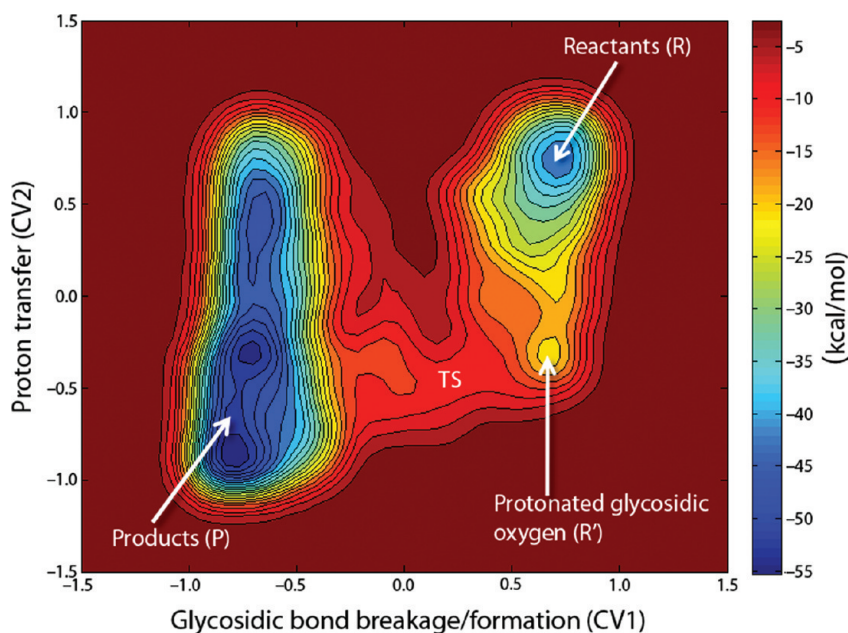


Figure 5. Free-energy surface of cellopentaose hydrolysis by a GH8 EG. CV1 measures the rupture of the β -glycosidic bond with the simultaneous formation of the α -glycosidic bond. CV2 measures the proton donation between Glu95 and the oxygen atom of the scissile β -glycosidic bond. Contours are separated by 3 kcal mol^{-1} .

point in the reaction pathway. When the Glu95 proton is transferred to the β -glycosidic oxygen atom (O4), the free energy increases to about 12 kcal mol^{-1} below the TS energy. Our calculations support cellulose hydrolysis as an exothermic reaction, as the products are $\sim 12 \text{ kcal mol}^{-1}$ below the reactants.

Snapshots of representative states along the reaction pathway are depicted in Figure 6. Table 1 shows average distances computed from all configurations falling into a small region around the stationary point of the free-energy surface (Figure 5) for structures that correspond to each of its minima (reactants, protonation of the glycosidic oxygen atom, and products) and to the TS.

The reactant conformation is shown in Figure 6A. Here, protonated Glu95 forms a hydrogen bond of 1.98 \AA with the O4 atom (Table 1). Asp278 and Tyr215 form hydrogen bonds with the nucleophilic water molecule and position it 3.82 \AA below the C1' atom. The glycon ring in the reactant state (the Michaelis complex) is maintained in an intermediate $\beta\text{-}^2S_0/^{2.5}B$

conformation, in good agreement with the experimental structure.²⁵ This ring conformation pre-activates the substrate for catalysis by moving the glycosidic bond toward an axial position, increasing the charge of the C1' atom¹⁰ and removing steric conflict between the H1' atom and the incoming nucleophilic water molecule.

The next step in the reaction pathway is proton transfer from Glu95 to the O4 atom (Figure 6B), thereby increasing the aglycon's leaving group ability. Glycosidic oxygen protonation also increases the length of the β -glycosidic bond by 0.11 \AA (Table 1) and the energy of the system, facilitating the reaction (Figure 5).

After the glycosidic oxygen atom is protonated, the dissociation (D_N) step takes place. The β -glycosidic bond lengthens by a further 0.83 \AA until it disengages sufficiently for the complex to reach the TS (Figure 6C). Simultaneously, the glycon becomes an oxacarbenium ion-like species, where the C1'–O5' bond shortens by 0.13 \AA compared to its reactant state (Table

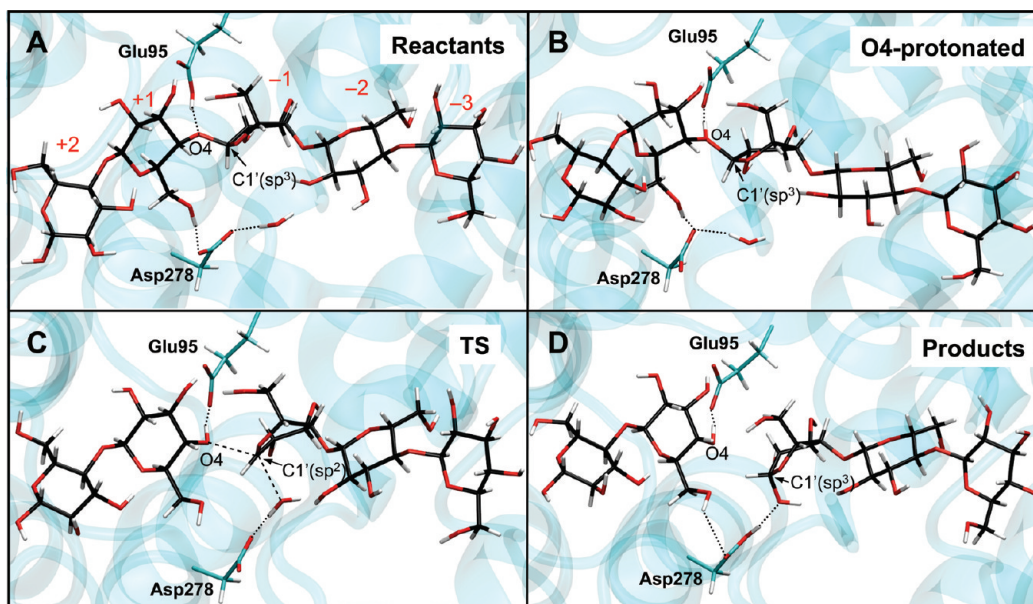


Figure 6. Snapshots of important states along the reaction pathway. Subsite numbering is indicated in the reactants panel.

TABLE 1: Average Bond Distances and Their Standard Deviations at Different Stages of the Reaction Pathway (Å)

bond	reactants	O4 protonated	TS	products
C1'–O4	1.51 ± 0.07	1.62 ± 0.07	2.45 ± 0.17	3.83 ± 0.29
C1'–O _{wat}	3.82 ± 0.34	4.14 ± 0.19	3.16 ± 0.38	1.41 ± 0.04
O _{Glu95} –H _{Glu95}	1.02 ± 0.04	1.34 ± 0.05	1.47 ± 0.06	2.31 ± 0.38
H _{Glu95} –O4	1.98 ± 0.20	1.14 ± 0.04	1.08 ± 0.04	0.98 ± 0.03
O _{Asp278} –H _{wat}	1.79 ± 0.26	1.77 ± 0.12	1.69 ± 0.11	1.01 ± 0.03
O _{wat} –H _{wat}	1.01 ± 0.02	1.01 ± 0.02	1.02 ± 0.03	2.01 ± 0.19
O _{Tyr215} –H _{wat2}	1.91 ± 0.24	1.91 ± 0.25	2.03 ± 0.24	2.17 ± 0.33
C1'–O5'	1.42 ± 0.04	1.37 ± 0.02	1.29 ± 0.02	1.45 ± 0.03

1) and becomes sp^2 hybridized, forcing the C2', C1', O5', and C5' ring atoms into a plane. Calculated Wannier orbitals⁴⁶ show the formation of a partial double C1'–O5' bond at the TS (Figure 7). Upon formation of the oxacarbenium ion-like TS, the C1' atom moves closer to the oxygen atom of the nucleophilic water molecule by changing the glucopyranosyl ring conformation from a ${}^2S_0/{}^{2.5}B$ into a ${}^{2.5}B$ (Figure 6C).

At the TS, the glycosidic oxygen atom is 2.45 Å from the C1' atom, while the nucleophilic water molecule is 3.16 Å from

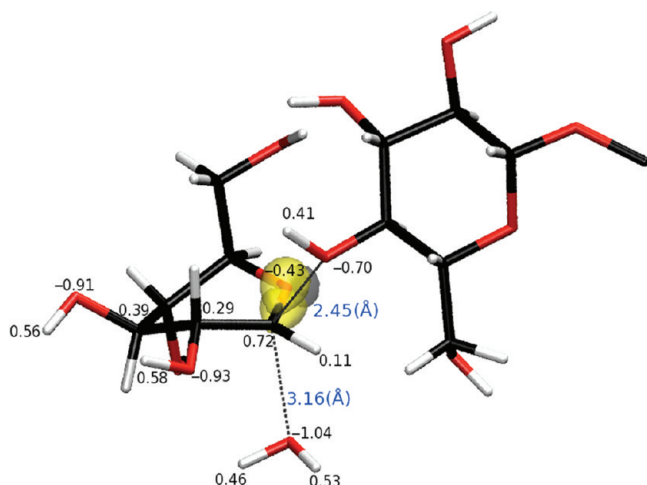


Figure 7. Structural and electronic features of the oxacarbenium ion-like TS. RESP partial charges of the more relevant atoms are shown as black labels. Wannier orbitals are shown for the sp^2 -hybridized C1'–O5' bond (yellow) and the ring oxygen lone pair (gray). The distances between the C1', O4, and O_{wat} are shown in blue.

the C1' atom (Table 1). The O_{wat}–C1'–O4 angle is $\sim 133 \pm 8^\circ$. Calculation of restrained electrostatic potential (RESP) partial charges⁵⁸ indicates a highly cationic oxacarbenium ion-like TS (Figure 7) bearing a substantial positive charge at the C1' atom (0.72, compared to 0.50 and 0.44 in the reactants and products, respectively). One of the lone pairs of the ring oxygen atom helps to stabilize the positive charge buildup by forming a partial double bond. The nucleophilic water molecule is polarized by its interactions with the putative proton acceptor Asp278, with the O_{Asp278}...H_{wat} distance decreasing by 0.1 Å from the reactants to the TS (Table 1). The interactions of the C2'–OH and C3'–OH groups of the glucosyl residue in subsite –1 with Asp152 (Figure 2) are likely to relieve the electron deficiency of the TS and provide a stabilizing effect.

In the following association (A_N) step to form the product, the glycon changes from the ${}^{2.5}B$ TS to an α - 5S_1 conformation by moving its C1' atom further and simultaneously trapping the water molecule. In a concerted manner, a proton from the water molecule is transferred to Asp278 to complete the reaction (Figure 6D). The product well is very long and is characterized by three minima (Figure 5), two of which represent the aglycon protonated state ($CV2 \approx -0.4$ and -0.9). Another minimum ($CV2 \approx 0.4$), much less stable than the other ones, represents reprotonation of the proton donor.

In summary, our results show that the glucosyl residue occupying subsite –1 follows a β - ${}^2S_0 \rightarrow {}^{2.5}B$ [TS] $\rightarrow \alpha$ - 5S_1 conformational itinerary along the reaction coordinate, confirming previous predictions of a ${}^{2.5}B$ TS in GH8-catalyzed hydrolyses.^{4,5} A second and less rigorous analysis was done by automated docking to test the ability of the GH8 EG to bind

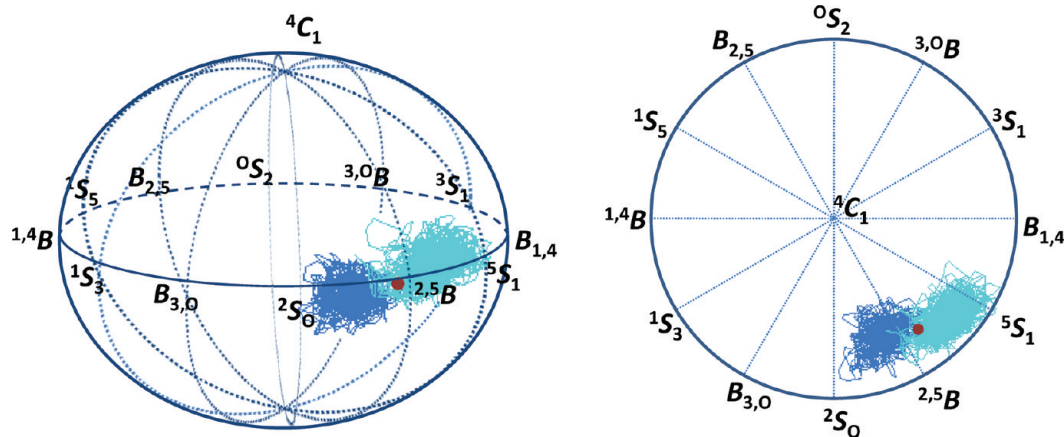


Figure 8. Conformational itinerary of the glycon glucosyl ring along the reaction coordinate. All ring conformations of the metadynamics simulation are mapped onto a Cremer–Pople sphere. Blue is the reactant ring conformations visited before the TS. Cyan is the product ring conformations visited after the TS. The red dot indicates the TS ring conformation.

possible TS conformations. The docking results, available as Supporting Information, further support a $^{2,5}B$ transition state.

3.4. Stereoelectronic Effects on the β - $^2S_0 \rightarrow ^{2,5}B$ [TS] $\rightarrow \alpha$ - 5S_1 -like Itinerary. Stereoelectronic effects are generally considered, but not without opposing views,⁵⁹ to play an important role in anomeric substitution reactions (such as hydrolysis) of pyranoses.⁶⁰ The ring oxygen atom facilitates the reaction by donating electrons to stabilize the positive charge that develops in the anomeric center at the TS. Positioning the oxygen lone pairs in an optimal orientation with respect to the C1' atom facilitates this electron transfer. For pyranosyl rings with a β -glycosidic bond, stereoelectronic theory dictates that the ring must undergo a conformational change away from the most stable 4C_1 chair form into a skew-boat conformation that allows the overlap between the anti-periplanar lone pair of the ring oxygen atom and the antibonding orbital of the β -glycosidic bond. This requirement prepares the ring for the substitution reaction. As the β -glycosidic bond elongates, the anti-periplanar lone pair in the O5' atom proceeds to form the partial C1'–O5' double bond at the TS (Figure 7). This is known as the anti-periplanar lone pair hypothesis (ALPH).

The pyranosyl ring conformations observed in GH active sites usually comply with the ALPH requirement. In the GH8 EG, the itinerary of the glucosyl ring in subsite -1 along the reaction coordinate follows a pathway similar to a β - $^2S_0 \rightarrow ^{2,5}B$ [TS] $\rightarrow \alpha$ - 5S_1 . Figure 8 shows two views of the Cremer–Pople sphere⁶¹ with the pyranosyl ring conformations visited throughout the metadynamics simulation. This conformational itinerary is consistent with ALPH as the pre-TS β - 2S_0 conformation allows the overlap of the anti-periplanar lone pair of the ring oxygen atom with the antibonding orbital of the β -glycosidic bond. Nevertheless, a $^{2,5}B$ conformation has a C5–H2 flagpole interaction and has eclipsed substituents in the C3–C4 positions of the pyranosyl ring that increase the energy of this conformation.

Substrate distortion during catalysis in GHs has been often discussed in relation to the relative movement of the anomeric carbon atom with respect to the nucleophile.^{4,62,63} Specifically, the C1' atom becomes electrophilic as the glycosidic bond breaks and migrates toward the nucleophile to trap it, forming the new bond. In the case of GH8 EG, the β - 2S_0 conformation of the glucosyl ring in subsite -1 in the Michaelis complex has its C1' atom placed higher relative to the $^{2,5}B$ [TS] conformation, which is the next step in the reaction coordinate (Figure 9B). Therefore, a simple C1' movement from the β - 2S_0 conformer reaches the $^{2,5}B$ conformation and places the C2', C1', O5', and

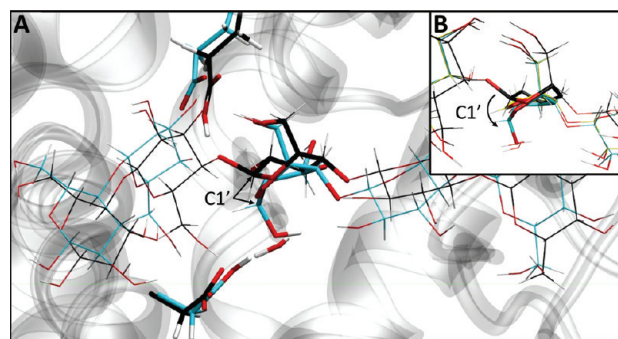


Figure 9. (A) Superposition of reactant (black) and product (blue) states. The C1' atom moves the longest distance to trap the nucleophilic water molecule. (B) Glycon ring conformations of the reactant state (black), TS (yellow), and product state (blue).

C5' ring atoms in a plane, a requirement imposed by the oxacarbenium ion-like, sp^2 -hybridized C1'–O5' bond. The α - 5S_1 product conformation is also a mere C1' movement down from the TS conformation. Figure 9A shows a superimposition of the reactant and product states in the GH8 EG, demonstrating that the C1' atom migrates furthest to trap the nucleophilic water molecule. Therefore, the β - $^2S_0 \rightarrow ^{2,5}B$ [TS] $\rightarrow \alpha$ - 5S_1 itinerary is consistent with the so-called electrophilic migration mechanism.

3.5. The GH8 Endoglucanase Proton Acceptor. As mentioned above, there has been some controversy about the identity of the GH8 EG proton acceptor. This is because the position of this residue in the active site is not conserved among GH8 members, which is unusual within GH families. This led to a subdivision of GH8 into three subfamilies (GH8a, GH8b, and GH8c) depending on the location of the proton acceptor.²⁸ Moreover, kinetic studies of GH8 members that have residues corresponding to Asp278 show significant retention of activity upon mutation of the putative catalytic base.^{26,27}

The crystal structure of the GH8 EG has a water molecule underneath the C1' atom in a good position to proceed with the substitution. Asp278 and Tyr215 form hydrogen bonds with the nucleophilic water molecule. The crystallographic model²⁵ suggests that the proton acceptor is Asp278, as most GHs employ a pair of carboxylate-containing amino acids as proton donor and proton acceptor. Although Tyr215 also hydrogen-bonds the nucleophilic water molecule, it seems unlikely that it could be the catalytic proton acceptor given that its pK_a would have to be considerably lower to assume this role (there is no positively charged residue in the vicinity of Tyr215 that could

justify a drastic pK_a change of this residue). Tyr215 is strictly conserved within GH8 and even further within clan GH-M. The mutation of this residue to phenylalanine leads to an almost complete loss of activity in GH8 members. Even though tyrosine can act as nucleophile in some retaining GHs,⁶⁴ the role of Tyr215 in GH8 is apparently to fix the water molecule in a correct position for catalysis.²⁶

To confirm the identity of the proton acceptor in GH8 *C. thermocellum* EG, we calculated where the proton released from the nucleophilic water molecule after hydrolysis was likely to go by performing a series of QM/MM geometric optimizations for three different situations: (1) after Asp278 takes the proton; (2) after Tyr215, which forms a second hydrogen bond with the nucleophilic water molecule, takes the proton leading to the tyrosine hydroxyl group becoming positively charged; (3) after the solvent becomes the proton acceptor, forming a H_3O^+ cation. The GH8 active site is fairly solvent exposed with several water molecules underneath the substrate, but only one is coordinated by Asp278 and Tyr215 and positioned correctly for catalysis. The rest of the water molecules could be responsible for stabilizing the proton.

The first case is the result from the metadynamics simulation. In the second case, the proton on Tyr215 rapidly transfers to Asp278 via the $C1'-OH$ group. In the third case, the geometric optimization that starts with an H_3O^+ cation is unstable, and the proton also migrates back to Asp278. Therefore, our results support Asp278 as the proton acceptor of the GH8 *C. thermocellum* EG. However, the enzymatic mechanism and the identity of the proton acceptor in the unusual situation where Asp278 is not present, such as in GH8 EG mutants or in GH48 members, remain unknown and await study.

4. Conclusions

Enzymatic hydrolysis of cellulose by GH8 *C. thermocellum* EG was explicitly modeled with DFT QM/MM methods. The reaction was activated by using the metadynamics approach, and the free energy of the reaction with respect to two reaction coordinates was calculated. The Michaelis complex (i.e., the bound reactants) shows that the glucosyl unit occupying subsite -1 is in a distorted $^2S_0/^{2.5}B$ ring conformation, in good agreement with experiments.²⁵ The minimal-energy pathway from reactants to products passes through a transient state characterized by protonation of the glycosidic oxygen (O4) and a substantial lengthening of the glycosidic bond. The glucosyl residue occupying subsite -1 takes a $^{2.5}B$ conformation at the TS, whose structural features are consistent with an oxacarbenium ion-like structure (i.e., an sp^2 -hybridized $C1'-O5'$ bond that forces planarity of the $C2'$, $C1'$, $O5'$, and $C5'$ ring atoms). This is the first time that a boat-type transition state has been found in GHs by first-principles calculations, confirming previous hypothesis of a $^{2.5}B$ -type TS in GH8 enzymes.^{4,5} The conformation of the glucosyl residue occupying subsite -1 is 5S_1 in the product state. Therefore, our calculations predict that catalysis by GH8 members takes place via a $\beta\text{-}^2S_0 \rightarrow ^{2.5}B$ [TS] $\rightarrow \alpha\text{-}^5S_1$ itinerary of the glycon ring along the reaction coordinate, consistent with the anti-periplanar lone pair hypothesis. The reaction path calculations, as well as additional QM/MM calculations on alternative scenarios (i.e., Tyr215 or solvent water as a catalytic proton acceptor), provide evidence that Asp278 is the catalytic proton acceptor of GH8 *C. thermocellum* EG. In view of the sequence homology and structural similarity among all GH8a members, our results also apply to cellulases, xylanases, and other EGs belonging to subfamily GH8a.²⁸

Acknowledgment. The authors thank the National Research Initiative of the U.S. Department of Agriculture Cooperative State Research, Education and Extension Service (grant number 2007-35504-18252), the Generalitat de Catalunya (grant 2005SGR-00036), and the Ministerio de Ciencia e Innovación (MICINN) (grant FIS2008-03845) for supporting this work. A.A. acknowledges a FPU research fellowship from MCINN. We are also deeply grateful to Wim Nerinckx (University of Ghent) for his continued advice. We are very grateful to Antoni Planas (Universitat Ramon Llull) and Pedro Alzari (Institut Pasteur) for discussions on the GH8 catalytic mechanism, to Mercedes Alfonso-Prieto for her help with CPMD, and to the Iowa State University High Performance Computation System for making its facilities available to us.

Supporting Information Available: Automated docking analysis of a cellotriose molecule with the middle glucose ring distorted to oxacarbenium ion-like compliant conformations is included. In addition, structures of the active site corresponding to the stationary points of the free-energy surface (Figure 5) are provided. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Coutinho, P.; Henrissat, B. Carbohydrate-Active Enzymes Server 1999. <http://www.cazy.org/>.
- (2) Rye, C. S.; Withers, S. G. *Curr. Opin. Chem. Biol.* **2000**, *4*, 573–580.
- (3) Vasella, A.; Davies, G. J.; Bohm, M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 619–629.
- (4) Vocadlo, D. J.; Davies, G. J. *Curr. Opin. Chem. Biol.* **2008**, *12*, 539–555.
- (5) Davies, G. J.; Ducros, V. M.; Varrot, A.; Zechel, D. L. *Biochem. Soc. Trans.* **2003**, *31*, 523–527.
- (6) Nerinckx, W.; Desmet, T.; Claeysens, M. *ARKIVOC* **2006**, *xiii*, 1–27.
- (7) Ducros, V. M.; Zechel, D. L.; Murshudov, G. N.; Gilbert, H. J.; Szabo, L.; Stoll, D.; Withers, S. G.; Davies, G. J. *Angew. Chem., Int. Ed.* **2002**, *41*, 2824–2827.
- (8) Biarnes, X.; Nieto, J.; Planas, A.; Rovira, C. *J. Biol. Chem.* **2006**, *281*, 1432–1441.
- (9) Mulakala, C.; Nerinckx, W.; Reilly, P. J. *Carbohydr. Res.* **2006**, *341*, 2233–2245.
- (10) Biarnes, X.; Ardevol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2007**, *129*, 10686–10693.
- (11) Fushinobu, S.; Mertz, B.; Hill, A. D.; Hidaka, M.; Kitaoka, M.; Reilly, P. J. *Carbohydr. Res.* **2008**, *343*, 1023–1033.
- (12) Tailford, L. E.; Offen, W. A.; Smith, N. L.; Dumon, C.; Morland, C.; Gratien, J.; Heck, M. P.; Stick, R. V.; Bleriot, Y.; Vasella, A.; Gilbert, H. J.; Davies, G. J. *Nat. Chem. Biol.* **2008**, *4*, 306–312.
- (13) Greig, I. R.; Zahariev, F.; Withers, S. G. *J. Am. Chem. Soc.* **2008**, *130*, 17620–17628.
- (14) Soliman, M. E. S.; Ruggiero, G. D.; Ruiz Pernía, J. J.; Greig, I.; Williams, I. H. *Org. Biomol. Chem.* **2009**, *7*, 460–468.
- (15) Zechel, D. L.; Withers, S. G. *Acc. Chem. Res.* **2000**, *33*, 11–18.
- (16) Vocadlo, D. J.; Wicki, J.; Rupitz, K.; Withers, S. G. *Biochemistry* **2002**, *41*, 9727–9735.
- (17) Bottoni, A.; Miscione, G. P.; De Vivo, M. *Proteins: Struct., Funct., Genet.* **2005**, *59*, 118–130.
- (18) Bowman, A. L.; Grant, I. M.; Mulholland, A. J. *Chem. Commun. (Cambridge, U.K.)* **2008**, 4425–4427.
- (19) Brás, N. F.; Moura-Tamames, S. A.; Fernandes, P. A.; Ramos, M. J. *J. Comput. Chem.* **2008**, *29*, 2565–2574.
- (20) Bayer, E. A.; Lamed, R.; Himmel, M. E. *Curr. Opin. Biotechnol.* **2007**, *18*, 237–245.
- (21) Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804–807.
- (22) Sticklen, M. B. *Nat. Rev. Genet.* **2008**, *9*, 433–443.
- (23) Alzari, P. M.; Souchon, H.; Dominguez, R. *Structure (Cambridge, MA, U.S.)* **1996**, *4*, 265–275.
- (24) Koshland, D. E. *Biol. Rev. Cambridge Philos. Soc.* **1953**, *28*, 416–436.
- (25) Guerin, D. M.; Lascombe, M. B.; Costabel, M.; Souchon, H.; Lamzin, V.; Beguin, P.; Alzari, P. M. *J. Mol. Biol.* **2002**, *316*, 1061–1069.
- (26) Collins, T.; De Vos, D.; Hoyoux, A.; Savvides, S. N.; Gerday, C.; Van Beeumen, J.; Feller, G. *J. Mol. Biol.* **2005**, *354*, 425–435.

- (27) Yao, Q.; Sun, T.; Chen, G.; Liu, W. *Biotechnol. Lett.* **2007**, *29*, 1243–1247.
- (28) Adachi, W.; Sakihama, Y.; Shimizu, S.; Fukazawa, T. S. T.; Suzuki, M.; Yatsunami, R.; Nakamura, S.; Takénaka, A. *J. Mol. Biol.* **2004**, *343*, 785–795.
- (29) Hightman, T. D.; Vasella, A. T. *Angew. Chem., Int. Ed.* **1999**, *38*, 750–770.
- (30) Sabini, E.; Sulzenbacher, G.; Dauter, M.; Dauter, Z.; Jorgensen, P. L.; Schülein, M.; Dupont, C.; Davies, G. J.; Wilson, K. S. *Chem. Biol.* **1999**, *6*, 483–492.
- (31) Pearlman, D. A.; Case, D. A.; Caldwell, J.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
- (32) Hornak, V.; Simmerling, C. *Proteins: Struct., Funct., Genet.* **2003**, *51*, 577–590.
- (33) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.
- (34) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (35) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (36) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941–6947.
- (37) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- (38) Von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. *J. Chem. Phys.* **2005**, *122*, 014113.
- (39) Hünenberger, P. *J. Chem. Phys.* **2000**, *113*, 10464–10476.
- (40) Carloni, P.; Rothlisberger, U.; Parrinello, M. *Acc. Chem. Res.* **2002**, *35*, 455–464.
- (41) Dal Peraro, M.; Ruggerone, P.; Raugei, S.; Gervasio, F. L.; Carloni, P. *Curr. Opin. Struct. Biol.* **2007**, *17*, 149–156.
- (42) Troullier, N.; Martins, J. L. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 1993–2006.
- (43) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (44) Ireta, J.; Neugebauer, J.; Scheffler, M. *J. Phys. Chem. A* **2004**, *108*, 5692–5698.
- (45) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (46) Marzari, N.; Vanderbilt, D. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *56*, 12847–12865.
- (47) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302.
- (48) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (49) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. *J. Phys. Chem. B* **2005**, *109*, 6676–6687.
- (50) Oganov, A. R.; Martonak, R.; Laio, A.; Raiteri, P.; Parrinello, M. *Nature (London)* **2005**, *438*, 1142–1144.
- (51) Blumberger, J.; Ensing, B.; Klein, M. L. *Angew. Chem., Int. Ed.* **2006**, *45*, 2893–2897.
- (52) Fiorin, G.; Pastore, A.; Carloni, P.; Parrinello, M. *Biophys. J.* **2006**, *91*, 2768–2777.
- (53) Martonak, R.; Donadio, D.; Oganov, A. R.; Parrinello, M. *Nat. Mater.* **2006**, *5*, 623–626.
- (54) Rodriguez-Forte, A.; Iannuzzi, M.; Parrinello, M. *J. Phys. Chem. B* **2006**, *110*, 3477–3484.
- (55) Guthrie, R. D.; Jencks, W. P. *Acc. Chem. Res.* **1989**, *22*, 344–349.
- (56) Guthrie, R. D. *Pure Appl. Chem.* **1989**, *61*, 23–56.
- (57) Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737–7744.
- (58) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (59) Sinnott, M. *Biochem. J.* **1984**, *224*, 817–821.
- (60) Deslongchamps, P. *Pure Appl. Chem.* **1993**, *65*, 1161–1178.
- (61) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.
- (62) Vocadlo, D. J.; Davies, G. J.; Laine, R.; Withers, S. G. *Nature (London)* **2001**, *412*, 835–838.
- (63) Schramm, V. L.; Shi, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 657–665.
- (64) Watts, A. G.; Damager, I.; Amaya, M. L.; Buschiazzi, A.; Alzari, P.; Frasch, A. C.; Withers, S. G. *J. Am. Chem. Soc.* **2003**, *125*, 7532–7533.

JP811470D