

QM/MM Study on the Catalytic Mechanism of Cellulose Hydrolysis Catalyzed by Cellulase Cel5A from *Acidothermus cellulolyticus*

Jingli Liu, Xuemei Wang, and Dingguo Xu*

College of Chemistry, MOE Key Laboratory of Green Chemistry and Technology, Sichuan University, Chengdu, Sichuan 610064, PR China

Received: September 23, 2009; Revised Manuscript Received: November 11, 2009

Cellulase Cel5A from *Acidothermus cellulolyticus* is an endoglucanase which features the retention mechanism for the cleavage of the β -1,4-glycosidic bond. In this work, we investigated the detailed catalytic steps in the formation of two cellobiose units from the hydrolysis of a cellotetraose molecule using a combined QM/MM approach. The understanding of the catalysis process at the atomistic level may help further protein engineering research. Molecular dynamics, potentials of mean force (PMFs), and reaction path calculations confirmed that the hydrolysis of cellotetraose has a retention mechanism via oxocarbenium-like transition states for both the glycosylation and deglycosylation steps.

1. Introduction

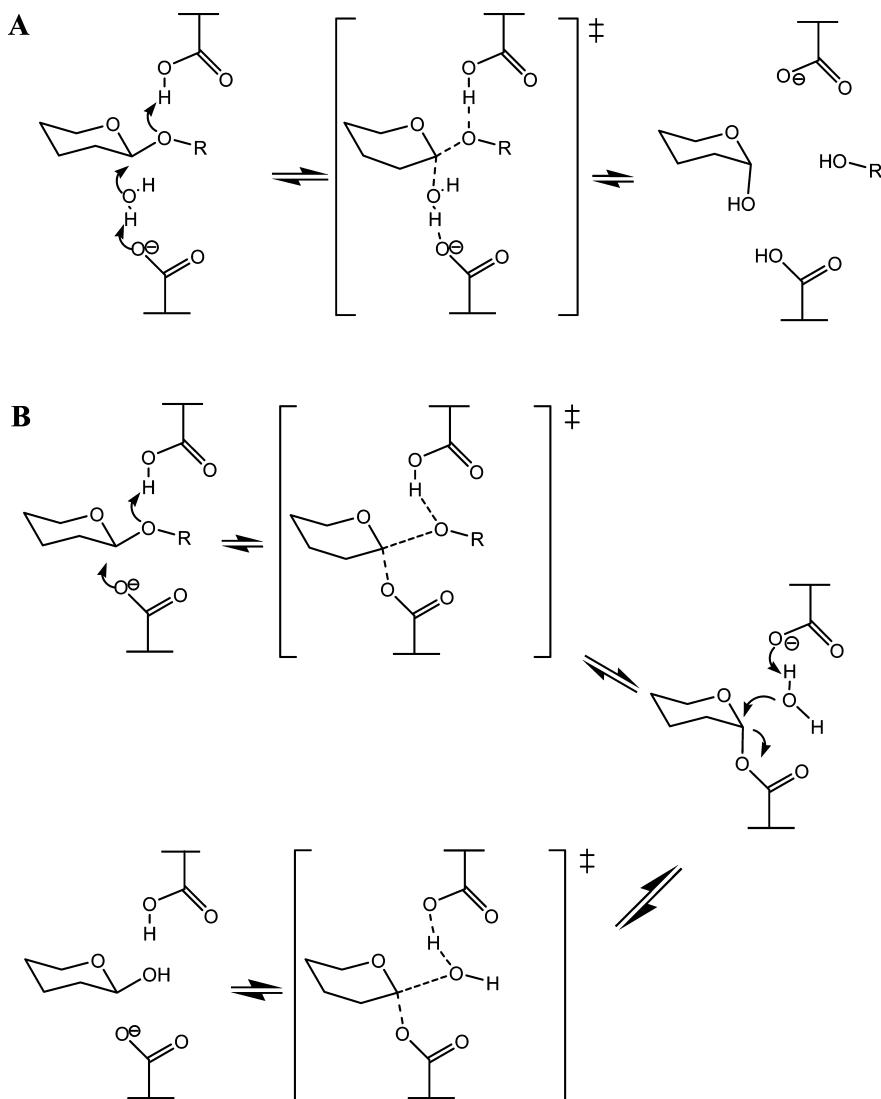
Cellulose, an insoluble glucose homopolymer linked by unbranched β -1,4-glycosidic bonds, is the primary source of renewable biofuel.^{1–5} Cellulases, on the other hand, are enzymes that are capable of biodegradation of cellulose. The low activity of cellulases to the crystalline cellulose represents a major bottleneck in the conversion of cellulose-based biomass to ethanol.⁵ Recently, protein engineering studies aimed at increasing the catalytic activity have received extensive attention. A better understanding of the catalytic mechanism of cellulase-catalyzed hydrolysis of cellulose is highly desired for further protein engineering research.

Cellulases belong to the glycoside hydrolase (GH) superfamily, which can catalyze the cleavage of the β -1,4-glycosidic bond. It is believed that enzymes in the GH superfamily, which contain over 110 subfamilies depending on their amino acid sequences, are one of the most efficient catalysts. The detailed subfamily or protein sequence information can be found at the Web site (http://www.cazy.org/fam/acc_GH.html).⁶ In 1953, Koshland⁷ provided a conceptual framework on how the glycoside hydrolases catalyze the cleavage of the glycosidic linkage. Two different stereospecific reaction mechanisms for the GH superfamily enzymes by different stereospecific manners were proposed: retention and inversion. A detailed description of the catalytic mechanism for GH enzymes can be found in several excellent reviews.^{8–11} Here we give a simple description in Scheme 1. In summary, for the retaining enzymes, the reaction follows a double displacement mechanism via a stable covalent glycosyl-enzyme intermediate, whereas for inverting enzymes, the reaction consists of a single nucleophilic substitution with only one transition state. Some basic geometrical characters for these two kinds of enzymes have been identified: e.g., for inverting enzyme, two carboxylate groups of two conserved Glu residues should be placed some 10.5 Å away^{9,12,13} to accommodate the substrate and water nucleophile; while for retaining enzymes, two carboxyl groups are only 5.5 Å apart.⁹ Since the proposal by Koshland, there have been numerous experimental studies on the enzymatic hydrolysis of the glycosidic bond. For example, Davies et al.^{14,15} reported a

series of X-ray structures along the putative reaction coordinates. Kinetic isotopic effects studies^{16–19} by several groups also provided evidence for the two possible mechanisms for the hydrolysis of the glycosidic bond. At the same time, a QM/MM study²⁰ on the glycosidic bond cleavage reaction of hen egg white lysozyme (HEWL) with its substrate has been reported. Another study using DFT method²¹ on the hydrolysis of glycosidic bond has been reported. In addition, the dynamics of the substrate in the *Bacillus* 1,3-1,4- β -glucanase active site has also been examined by first principles molecular dynamics simulations.²² Finally, an ab initio metadynamics study²³ was carried out on the conformational free energy landscape of β -D-glucopyranose. However, for the cellulases and its catalysis, there are few theoretical studies reported to date. In this publication, we present a detailed theoretical investigation on the hydrolysis mechanism of cellulose catalyzed by cellulase Cel5A from *Acidothermus cellulolyticus*.

Cellulase *A. cellulolyticus* Cel5A belongs to the GHs subfamily of GH5, which contains several other enzymes from different bacterial sources.^{24–26} The generally accepted catalytic mechanism for cellulase Cel5A is the so-called retention mechanism with two Glu residues as the nucleophile and general acid/base, respectively. Such a mechanism is further supported by several X-ray structures for this enzyme.^{14,15,24,26–28} Particularly, Sakon et al.²⁴ reported an enzyme–cellotetraose complex structure for Cel5A from *A. cellulolyticus*, which provides a good starting point to carry out further mechanism studies for this enzyme. On the basis of this, Baker et al.²⁹ found that the activity for *A. cellulolyticus* Cel5A can be increased by about 20% with the Y245G mutant, due to the fact that the produced cellobiose binding affinity is reduced by the mutation. The engineering of the enzyme is not just restricted to the catalytic domain, but also in the carbohydrate binding modules (CBM). The mutagenesis studies on some cellulose surface-binding residues were also carried out by McCarter et al.³⁰ on this enzyme. Another site-directed mutagenesis study and CBM engineering of *Thermotoga maritima* Cel5A shows 14–18-fold higher hydrolytic activity.³¹ Several reviews on the rational design for cellulases can be found in ref 5 and the literature cited therein. The primary difficulty for an effective rational design for the

* To whom correspondence should be addressed: dgxu@scu.edu.cn.

SCHEME 1: General Inverting (A) and Retaining (B) Mechanisms for the Hydrolysis of β -1,4-Glycosidic Bond

cellulases is that there are no experimental crystal structures reported for cellulase–crystalline cellulose complexes.

Our current understanding of the hydrolysis mechanism for Cel5A or other cellulases is based on the experimental X-ray structures or kinetic studies.²⁶ In this work, we will examine how cellobiose molecule is converted into two cellobiose units catalyzed by Cel5A from *A. cellulolyticus*. Although its catalytic mechanism has been discussed extensively,^{24,29} some basic questions still await answers. For example, which step is the rate-limiting step, glycosylation or deglycosylation? Should the hydrolysis product of cellobiose be released before the nucleophilic water molecule comes in for the deglycosylation? The detailed understanding of the catalysis mechanism at the atomistic level is obviously necessary for a successful rational design of the enzyme. In this article, a hybrid quantum mechanical and molecular mechanical approach (QM/MM)³² is used for all of molecular simulation computations.

2. Computational Methods

The combined quantum mechanical/molecular mechanical approach is considered as a useful tool to understand characteristics of macromolecular systems, such as enzyme systems.^{32–36} The basic idea is to divide the whole system into two parts. The QM region consists of the atoms which are involved in

bond-forming and bond-breaking processes, and thus it is treated with an accurate QM method. On the other hand, the MM region containing most of protein residues and solvent waters is approximated using a classical force field. The QM/MM method is a simulation approach which combines accuracy of QM and efficiency of MM.

Semiempirical approaches, such as PM3,³⁷ AM1,³⁸ etc., have been employed in the literature^{39,40} for the electronic structure calculations in the QM region. Recently, hybrid ab initio QM/MM or DFT/MM studies have also been reported,^{41–43} but their applications are mostly restricted to the minimum energy path calculations or single-point energy computation. In this work, we employed a recent developed semiempirical approach, namely the self-consistent charge-density functional tight binding (SCC-DFTB) method,^{44,45} to perform the electronic structure calculations for QM region. This method has shown some encouraging results for several enzyme systems.^{36,46–50} The MM region was described by the CHARMM22 all-atom force field.⁵¹

The initial structure for molecular simulation was obtained from Protein Data Bank (PDB code 1ECE²⁴), which is the X-ray structure of Cel5A from *A. cellulolyticus* complexed with the cellobiose substrate molecule. Such an enzyme–substrate (E-S) complex structure provides a good starting point for subsequent simulations. Before the simulations, we need to

identify the correct protonation states for all histidine residues, which is done by carefully examining how the histidine residues contact nearby residues or solvent. In this work, Glu162 was simulated as a protonated state judged by the distance of about 2.52 Å from O_{e2} to O4(2) atom (Scheme 1). The nomenclature of atoms for the polysaccharide in this work follows the IUPAC-IUB convention.⁵²

We first solvated the entire system with a 25 Å radius sphere of preequilibrated TIP3P⁵³ water molecules with the origin located at C1(3) atom and removed all of those water molecules within 2.8 Å of heavy protein atoms. Subsequently, a 30 ps molecular dynamics was then applied with all protein, substrate and X-ray water molecules fixed at their original positions to relax all of added solvent molecules. Such a procedure was repeated several times to ensure that the system is sufficiently solvated. Then, stochastic boundary conditions⁵⁴ were applied to reduce the computational costs. All atoms of outside a 25 Å radius away from the origin were removed, while atoms in the buffer zone (22 Å < r < 25 Å) were subjected to Langevin dynamics⁵⁵ with friction and random forces. In the inner reaction zone (r < 22 Å), the atoms follow Newtonian dynamics on the hybrid QM/MM potential energy hypersurface. A group-based switching scheme was used for nonbonded interactions.⁵⁶

The selection of the QM region is critical for the accuracy of QM/MM calculations. In this work, the QM region consists of the cellobiose, side-chain groups of E162 and E282 residues for the enzyme–substrate (E-S) complex molecular dynamics (MD) simulation and the glycosylation reaction. For the subsequent deglycosylation step in which the cellobiose product obtained from the glycosylation half-reaction is removed, the QM region contains the E282–cellobiose complex, E162 and the nucleophilic water molecule which forms a hydrogen bond with E162 side chain. The link atom approach³³ was applied to describe the covalent bond connection between the QM and MM regions. Once the QM region has been determined, the subsequent calculations including structure optimizations and molecule dynamics simulations were carried at the QM/MM level of theory. The adopted basis Newton–Raphson (ABNR) method⁵⁷ was then used for the geometry optimization. The SCC-DFTB/MM optimizations is considered converged when the rms gradient is less than 0.01 kcal/(mol·Å²) if not stated in other place.

MD simulations were carried out for three different model structures: the E-S complex, enzyme–intermediate (E-I) complex with the cellobiose leaving group binding in the active site, and E-I complex without this cellobiose. For each structure, the system was slowly heated to room temperature (300 K) in 30 ps, and followed by 70 ps MD for sufficient equilibration at 300 K, the rest 450 ps MD at room temperature was used for analysis. The MD integration time step was set as 1 fs, and SHAKE algorithm⁵⁸ was applied to maintain all of covalent bonds involved hydrogen atoms.

The potential of mean force (PMF) is the key to discuss the catalysis mechanism of an enzyme-catalyzed reaction. In this work, we first calculated the reaction paths for both the glycosylation and deglycosylation steps using an adiabatic mapping method (also sometimes referred to reaction coordinate driving method; details can be seen in ref 59). These structures along the putative reaction coordinates were then used as the initial structures for the computation of the PMFs. In particular, the reaction coordinates were chosen as follows: $d_1 = d_{O_{e1}-C_{1(3)}} - d_{C_{1(3)}-O_{4(2)}}$ for the glycosylation step; and $d_2 = d_{O_{e1}-C_{1(3)}} - d_{C_{1(3)}-O_w}$ for the deglycosylation step. An umbrella sampling method⁶⁰ was applied with harmonic constraints ranging from

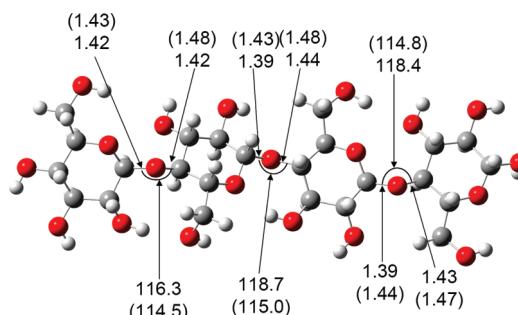


Figure 1. Comparison of selected geometries between B3LYP/6-31G* and SCC-DFTB (parentheses) optimized cellobiose structures. Units for distances are given in angstroms and angles in degrees.

100 to 200 kcal/(mol·Å²). A total 17 windows were for glycosylation and 20 windows for deglycosylation. A 100 ps constrained molecular dynamics simulation was performed in each window, with 60 ps heating and equilibration and 40 ps data collection for further analysis. Finally, the weighted histogram analysis method (WHAM)⁶¹ was used to obtain the final PMF. All of our molecular simulation calculations were carried out using the CHARMM suite of programs.⁵⁷

3. Results and Discussion

3.1. Validation of SCC-DFTB. The validity for the SCC-DFTB method has been assessed by several previous studies. Particularly, the SCC-DFTB-optimized geometries can be compared with those obtained at the B3LYP/6-31++G(d,p) level of theory.⁴⁶ For the particular system in this work, we have fully optimized the cellobiose structure at the B3LYP/6-31G* level of theory and compared it in Figure 1 with that obtained from the SCC-DFTB optimization with the convergence criterion that the rms gradient is less than 0.01 kcal/(mol·Å²). The corresponding bond distances and bond angles are given in parentheses for comparison in Figure 1. Clearly, the typical geometry differences between the SCC-DFTB and B3LYP results are less than 0.1 Å for bond distances, and 5° for bond angles. Such small differences indicate that the SCC-DFTB method is reliable for describing geometries for relevant molecules. The density functional theory (DFT) calculations were performed using Gaussian03 suite of program.⁶²

3.2. Enzyme–Substrate Complex. The understanding of the enzyme–substrate binding pattern is essential for the further study of the catalysis mechanism, as well as rational design of the protein. In this work, we performed a 550 ps MD simulation to examine the binding mode of the enzyme–cellobiose complex. Although there are two substrate binding domains for Cel5A, namely the catalytic domain and carbohydrate binding module (CBM), we have in this work investigated only the binding mode of cellobiose with the catalytic domain.

The Michaelis complex is quite stable during the MD simulation, demonstrated by the root-mean-square deviation (rmsd) of about 0.66 ± 0.03 Å as shown by Figure 2. The selected geometrical parameters are listed in Table 1. A snapshot from the MD trajectory is given in Figure 3. The geometries from our simulation are in good agreement with the X-ray structure. Clearly, the shape of the substrate is different from the geometry optimized by DFT in the gas phase, in which the four rings are in a planar conformation. The geometrical arrangement of the substrate in the enzyme active site is in favor of the in-line attack by the nucleophile. At the same time, it also leaves space for the hydrogen bond from Glu162 to the glycosidic oxygen atom of O4(2). As pointed out by ref 9, for

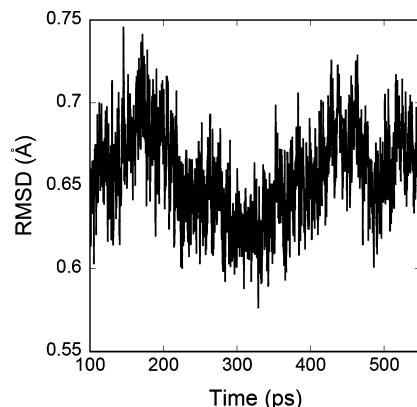


Figure 2. Rmsd for the backbone atoms of the substrate–enzyme complex.

a retaining glycosidase, the distance between two conserved carboxylate groups should be about 5.5 Å away. Indeed, the distance between two C_δ atoms of E162 and E282 is about 4.85 ± 0.50 Å, which can be compared with the experimental value of 4.85 Å. In addition, during our simulation, the distance between E282 O_{ε1} and C1(3) is about 3.07 ± 0.19 Å, which places the nucleophile E282 in a perfect near-attack position. At the same time, several hydrogen bonds from the enzyme provide additional stabilization of the nucleophile, thus facilitating the nucleophilic attack. For example, the distance between O_{ε1}(E282) and HH(Y240) is about 1.65 ± 0.09 Å, and the H21(Q161)–O_{ε2}(E282) distance is 2.45 ± 0.53 Å. And the position of the nucleophile is also stabilized by another hydrogen bond provided by substrate molecule itself, evidenced by the HO2(3)–O_{ε2}(E282) distance of 1.70 ± 0.13 Å.

On the other hand, as the general acid in the first half-reaction, the Glu162 carboxylate hydrogen forms a hydrogen bond with the O4(2) atom, with a distance of 2.21 ± 0.49 Å throughout the dynamics. Additionally, Glu162 is also stabilized by the hydrogen bond to HO6(2) with a distance of 2.20 ± 0.43 Å. At the same time, for the binding of the substrate cellobiose molecule to the enzyme, hydrogen bonds between cellobiose hydroxyl group and protein side-chain groups can be found, e.g., 2.20 ± 0.41 Å for O2(1)–H21(Q247), and 2.46 ± 0.44 Å for HO3(3)–N_{ε2}(H116). One of the important residues is Y245,

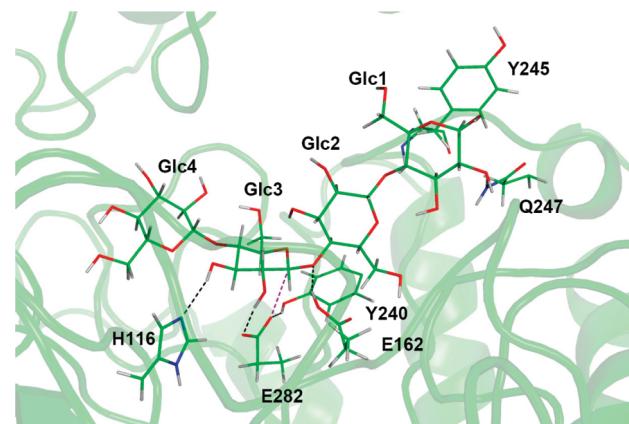


Figure 3. Snapshot of the enzyme–substrate complex. Black dashed lines represent the hydrogen bonds and purple dashed lines the nucleophilic attack direction.

which is believed to be involved in the binding of the cellobiose leaving group with hydrophobic interactions between its phenol group and the Glc1 sugar ring. Our simulation confirms that the phenol group of Y245 is almost parallel to the sugar ring as shown in Figure 3.

3.3. Enzyme–Intermediate Complexes. The direct evidence to support the double-displacement mechanism is a stable covalent glycosyl–enzyme intermediate. To gain further insights into the catalytic mechanism, we performed two MD simulations for the glycosyl–enzyme intermediate complex, with and without the cellobiose leaving group binding in the active site.

We first examined the model structure in the presence of the cellobiose leaving group. The initial structure for this model was obtained from our reaction path calculation after glycosylation. The covalent glycosyl–enzyme intermediate is quite stable during our MD simulation. Throughout the 550 ps simulation, the covalent bond between the O_{ε1}(E282) atom and substrate anomeric center of C1(3) atom is well maintained with a bond length of 1.49 ± 0.04 Å. A snapshot given in Figure 4 shows that the cellobiose molecule stays inside the active site quite stably. The distance between C1(3) and O4(2) is about 3.38 ± 0.37 Å. A strong hydrogen bond with the E162 carboxylate group can be also found from Table 1; i.e., the

TABLE 1: Selected Geometric Parameters for Stationary Points along the Glycosylation Path of the *A. cellulolyticus* Cel5A Catalyzed Hydrolysis of Cellobiose Calculated Using the SCC-DFTB/MM Methods^a

distances (Å), angles (deg)	glycosylation step					
	expt ²⁴	ES	TS1	EI1	MD (ES)	MD (EI1)
C1(3)…O _{ε1} (E282)	3.46	3.07	2.19	1.51	3.07 ± 0.19	1.49 ± 0.04
C1(3)…O4(2)	1.45	1.47	2.09	2.78	—	3.38 ± 0.37
C _ε (E162)…C _ε (E282)	4.85	4.64	4.06	4.00	4.85 ± 0.50	4.40 ± 0.30
HO2(3)…O _{ε2} (E282)	3.12 ^a	1.65	1.61	2.55	1.70 ± 0.13	2.92 ± 0.41
HO2(3)…O _{ε2} (E162)	—	3.28	3.10	1.82	3.16 ± 0.33	1.93 ± 0.46
O _{ε1} (E282)…HH(Y240)	2.54 ^a	1.58	1.66	1.90	1.65 ± 0.09	2.02 ± 0.26
H _{ε2} (E162)…O4(2)	2.52 ^a	1.82	1.53	1.03	2.21 ± 0.49	1.05 ± 0.05
H _{ε2} (E162)…O _{ε2} (E162)	—	1.00	1.05	1.68	—	2.61 ± 0.23
H _{ε2} (E162)…O _{ε1} (E162)	—	—	—	—	—	1.62 ± 0.18
O _{ε1} (E162)…HO6(2)	3.38 ^a	1.89	1.90	1.82	2.20 ± 0.43	2.10 ± 0.43
O _{ε1} (E162)…H _{δ1} (H238)	2.74 ^b	4.14	4.06	4.26	4.52 ± 0.85	4.68 ± 0.50
O2(1)…H21(Q247)	2.99 ^b	2.03	2.01	2.01	2.20 ± 0.41	2.24 ± 0.56
HO3(3)…N _{ε2} (H116)	—	2.29	2.11	2.12	2.46 ± 0.44	2.47 ± 0.50
C1(3)…O4(2)…C4(2)	114.6	119.4	—	—	118.3 ± 3.1	—
C5(3)–O5(3)–C1(3)–C2(3)	−41.9	−55.4	−5.68	−39.1	−55.4 ± 7.0	−55.5 ± 7.4

^a O–O distances. ^b O–N distances, and the experimental values are taken from chain A (1ECE). ^a For comparison, the average geometries from our QM/MM MD simulation for ES and glycosyl–enzyme intermediate were also listed. The names of protein atoms followed CHARMM convention, while those for cellobiose were named with IUPAC conventions.

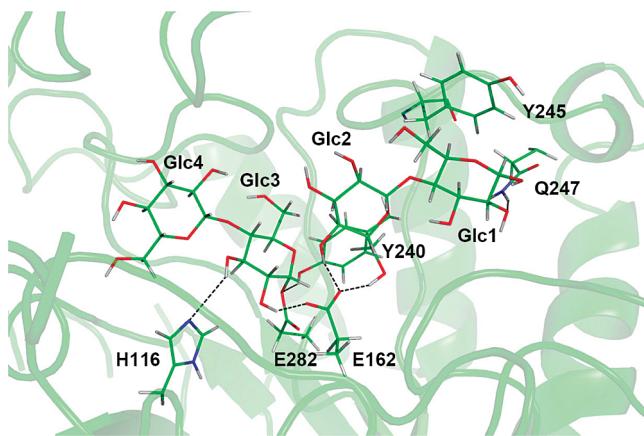


Figure 4. Snapshot of the glycosyl–enzyme intermediate with the cellobiose molecule in the active site. Black dashed lines represent hydrogen bonds.

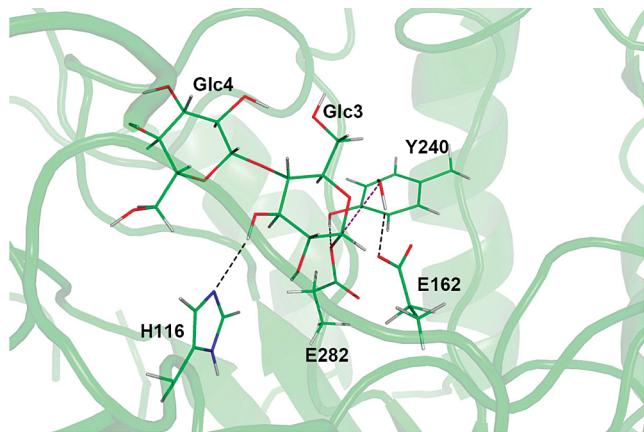


Figure 5. Snapshot of the glycosyl–enzyme intermediate without the cellobiose molecule in the active site. Black dashed lines represent the hydrogen bonds and purple dashed line the nucleophilic attack direction.

distance between $H_{\epsilon 2}$ and $O_{\epsilon 1}(E162)$ is 1.62 ± 0.18 Å. The position of E162 is further stabilized by a strong hydrogen bond between $HO2(3)$ and $O_{\epsilon 2}$ (E162), with a distance of 1.93 ± 0.46 Å. Throughout our simulation, no water molecule was found to enter the active site. At the same time, there are no other Glu or Asp residues in the active site that could serve the candidate for a general base for the next deglycosylation step. These observations suggest that the cellobiose product has to depart before deglycosylation.

To understand the functional role of the Glu162 in the deglycosylation step, we have also performed MD studies of the enzyme–intermediate complex without the cellobiose leaving group. The setup protocol is essentially the same as described above. Due to the absence of the product cellobiose, there is at least one water molecule in the active site, which forms a hydrogen bond with E162, and the distance between water oxygen to C1(3) atom is about 4.0 Å (see Supporting Information for details). A snapshot is given in Figure 5. This observation further suggests that E162 could act as the general base for the subsequent deglycosylation step, with the water serving as the nucleophile. At the same time, the covalent glycosyl–enzyme bond between E282 and cellobiose is well maintained, judging by the distance of 1.49 ± 0.04 Å for the C1(3)– $O_{\epsilon 1}(E282)$ bond.

3.4. Reaction Paths. Calculated reaction pathways for both the glycosylation and deglycosylation steps are depicted in

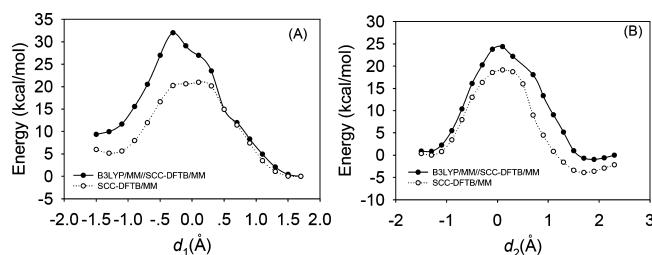


Figure 6. Minimum energy paths for both glycosylation (A) and deglycosylation(B) steps using the SCC-DFTB/MM method. For comparison, energies obtained from B3LYP/MM//SCC-DFTB/MM single-point calculations are also shown for both pathways.

Figure 6. There are six stationary states for the reactions, as displayed in Figure 7. Starting from a snapshot extracted from the MD simulation trajectory for the E-S complex, we first computed the minimum energy path (MEP) for the formation of the glycosyl–enzyme intermediate using an adiabatic mapping approach along the putative reaction coordinate of d_1 . As Figure 6A shows, there is only one transition state for this reaction, which is further investigated by a B3LYP/MM//SCC-DFTB/MM single-point calculation. Our B3LYP/MM calculations were performed using CHARMM interfaced with GAMESS-UK,⁶³ and the basis set used in the calculation is 6-31G(d). The results are consistent with the formation of a stable covalent glycosyl–enzyme intermediate, although some 11.0 kcal/mol energy barrier height difference can be seen. The selected geometry properties for the stationary structures are listed in Table 1. For the retention mechanism, one of the major characteristics is that the C5, O5, C1, and C2 atoms adopt a planar geometry¹¹ at the transition state. For a better view of the conformational change, we superimpose three stationary structures in Figure 8A. Indeed, we can observe the change of the C5–O5–C1–C2 dihedral angle from -55.4° to near planar (-5.7°) at the transition state. A clear chair–half-chair–chair conformational change can be seen from ES to the glycosyl intermediate. Particularly, a near-planar geometry for C5–O5–C1–C2 can be seen at the transition state. Such planarity indicates the formation of an oxocarbenium-like ion. At the same time, the proton $H_{\epsilon 2}$ is also transferred from E162 $O_{\epsilon 2}$ atom to $O4(2)$ to facilitate the cleavage of the C1(3)–O4(2) glycosidic bond. After formation of the glycosyl–enzyme intermediate, the proton is then fully transferred to $O4(2)$ to form the cellobiose molecule with a deprotonated E162 residue. In this part of the glycosylation reaction, the E162 plays the role of a general acid and the change of protonation state of E162 makes it a candidate for the general base in the subsequent deglycosylation half-reaction. The cellobiose leaving group still stays in the active site stably, which is consistent with our MD simulations.

After the departure of the cellobiose product, a water molecule should be able to come into the active site and act as the nucleophile in the second half-reaction with a Glu residue serving as the general base to activate this water molecule. To simulate the deglycosylation step, we first removed the cellobiose product from the active site and allowed some water molecules to fill the cavity. Those water molecules were further relaxed by a MD simulation as described in the Computational Methods section. Starting from a snapshot extracted from the MD simulation, we calculated the reaction path for the deglycosylation process. Figure 6B displays the MEP for this half-reaction, which again features a single transition state. The reaction path was also confirmed by B3LYP/MM//SCC-DFTB/MM single-point calculations with the standard 6-31G(d) basis

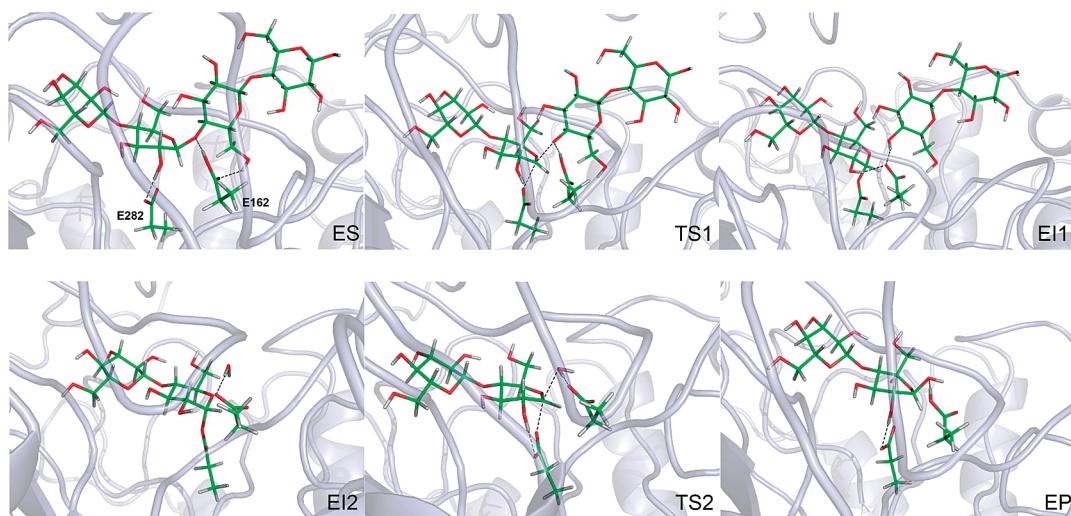


Figure 7. Snapshots of the QM/MM stationary states along the reaction path for Cel5A-catalyzed hydrolysis of cellotetraose. ES, TS1, and EI1 are for the glycosylation subreaction, while EI2, TS2, and EP are for deglycosylation subreaction.

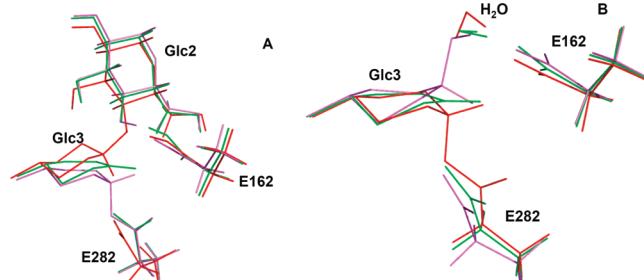


Figure 8. Overlay of structures of the conformational change of the Glc3 sugar ring. The structures are from the snapshots of stationary states obtained from reaction path calculations. Panel A is for the glycosylation step, and panel B is for the deglycosylation step. To obtain a better view, we keep the carbon atoms and O5 atom on the Glc3 sugar ring, and H1 atom is also kept to show the conformational change. Red color is for ES or EI2, green color is for transition states, while purple color is for EI1 and EP.

set. Like the process of glycosylation, the carboxylate group of the Glu residue (E162) accepts a proton from the water nucleophile. The conformation at the anomeric center C1(3) changes from chair to half-chair again to facilitate the attacking of nucleophile. Such a distortion of the sugar ring is once again confirmed by the change of the C5—O5—C1—C2 dihedral angle as shown in Table 2. A clearer view of the conformational change can also be found in Figure 8B. With the elongation of the C1(3)—O_{e1}(E282) bond from 1.50 to 2.99 Å, the covalent glycosyl–enzyme bond is finally broken, and the E162 is also protonated again to complete the catalytic cycle.

3.5. Potentials of Mean Force. Although the reaction paths can provide some useful information for the reaction mechanisms, it does not include entropic contributions. To include the effect of protein fluctuations and solvent effect, we have computed the potentials of mean force (PMFs) along two putative reaction coordinates, respectively.

The calculated PMFs are displayed in Figure 9: the left panel is for glycosylation and the right panel is for deglycosylation. Consistent with our previous reaction path calculations, only one transition state can be identified for the glycosylation or deglycosylation step. The free energy barrier is about 25.4 kcal/mol for the first step and 29.7 kcal/mol for the second step. The experimental free energy barrier estimated from a recent kinetic study on the hydrolysis of cellotetraose catalyzed by *Bacillus agaradherans* Cel5A is about 19.4 kcal/mol (derived

TABLE 2: Selected Geometric Parameters for Stationary Points along the Deglycosylation Path of the *A. cellulolyticus* Cel5A Catalyzed Hydrolysis of Cellotetraose Calculated Using the SCC-DFTB/MM Method^a

distances (Å), angles (deg)	deglycosylation step			
	MD(EI2)	EI2	TS2	EP
C1(3)…O _{e1} (E282)	1.49 ± 0.04	1.50	2.10	2.99
O _w …C1(3)	— ^b	3.37	2.20	1.45
H _w …O _{e2} (E162)	— ^b	1.75	1.27	1.00
H _w …O _w	— ^b	1.01	1.17	1.85
HO3(3)…N _{e2} (H16)	2.23 ± 0.36	2.12	2.09	2.12
O _{e1} (E282)…HH(Y240)	1.88 ± 0.19	1.77	1.65	1.62
HO2(3)…O _{e2} (E162)	4.00 ± 0.77	1.80	3.55	3.52
HO2(3)…O _{e2} (E282)	2.99 ± 0.36	2.69	1.70	1.79
H1(3)…C1(3)…C2(3)…H2(3)	45.6 ± 9.7	61.5	89.8	126.8
C2(3)…C1(31)…O5(3)…C5(3)	−56.0 ± 6.5	−52.7	−10.8	31.8

^a Geometries from the QM/MM MD simulation of EI2 were also listed in the table for comparison. ^b Due to high mobility of water molecules, it is impossible to give the statistical average values for these distances. Corresponding changes of distances along the time trajectory are given in the Supporting Information.

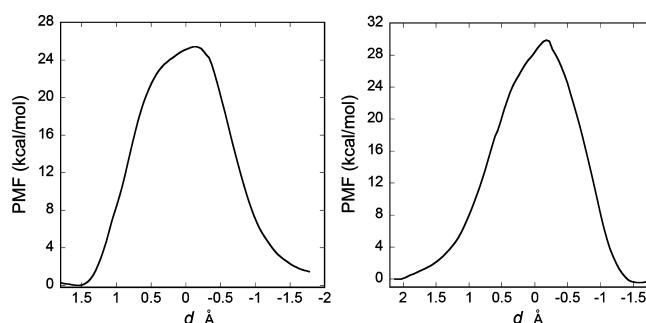
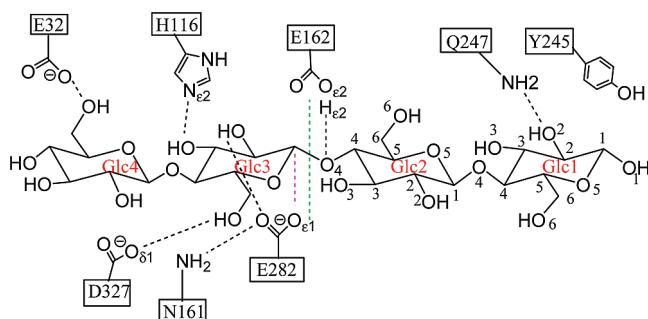


Figure 9. Calculated potentials of mean force for both glycosylation (left panel) and deglycosylation (right panel) processes.

from $k_{\text{cat}} = 0.05 \text{ s}^{-1}$).²⁶ Since the enzyme studied in this work is from a different bacteria resource, *A. cellulolyticus*, the different amino acid sequences could affect the free energy barrier for the enzymatic reaction. Additionally, the semiempirical SCC-DFTB method used for the QM region could also introduce significant errors. However, the mechanism of the catalyzed reaction seems to be firmly established by our QM/MM studies.

3.6. Discussion. In this work, we have investigated the transformation from a cellotetraose substrate molecule into two

SCHEME 2: Atomic Definitions (Following the IUPAC-IUB Convention for Cellotraose, and CHARMM for Protein Atoms)^a



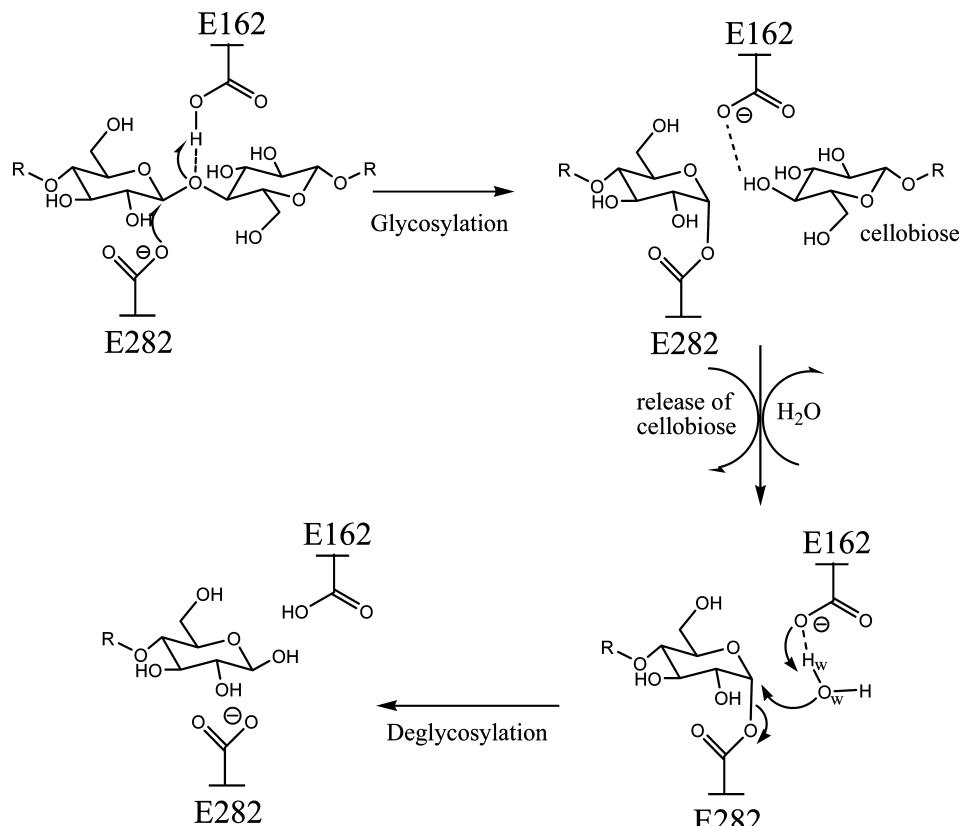
^a Some important hydrogen bonds are labeled by black dashed lines for the substrate binding. The green dashed line separates two cellobiose units. The purple dashed line represents the possible nucleophilic attack direction.

cellobiose units catalyzed by *A. cellulolyticus* cellulase Cel5A. Our QM/MM study provides strong support for Koshland's double-displacement mechanism for the cleavage of the β -1,4-glycosidic linkages, consistent with the conclusions of early experimental studies.^{24,30} Our simulations can provide more details on this catalysis process. Specifically, the E282 plays the role of nucleophile. A protonated E162 serves as the general acid for the formation of covalent glycosyl–enzyme intermediate, and the general base for the subsequent deglycosylation reaction. An oxocarbenium-like ion transition state is associated with each subreactions. As shown in Figure 8, the conformational change at the anomeric center of C1(3) atom for two

separate subreactions has the same character, which features the chair–half-chair–chair change along the reaction coordinate. The detailed reaction mechanism for the hydrolysis of cellotetraose catalyzed by Cel5A is presented in Scheme 2. The calculated free energy barriers for both parts of reaction are 25.7 kcal/mol for glycosylation, and 29.4 kcal/mol for deglycosylation. On the other hand, as Figure 6 shows, there are large energy differences between SCC-DFTB/MM and the high-level DFT/MM single-point calculations. Therefore, the 3.7 kcal/mol difference between the barriers in the two reaction steps may not be meaningful. To obtain more quantitative values for the reaction barriers, it might be desirable to carry out further free energy computations using higher-level method like ab initio/MM free energy computations. In addition, in the kinetic study on the hydrolysis of cellooligosaccharides catalyzed by the *B. agaradherans* Cel5A,²⁶ the catalysis rate increases when the degrees of polymerization (DP) increase accordingly. For example, it was observed that the free energy barriers for cellooligosaccharides are about 19.2 kcal/mol for DP = 4, 15.7 kcal/mol for DP = 5, and 15.3 kcal/mol for DP = 6 (estimated from the k_{cat} values using transition state theory) at pH = 7.5. The same behavior can be found in a recent kinetic study on the β -mannanase Man5A.⁶⁴ Since we do not know the impact of the glucose chain length on the reaction, it might be desirable to carry out further simulations using the cellooligosaccharides with larger DPs.

The oxocarbenium-like transition state is not unique for the retaining glycosidases. A very recent QM/MM study on the inverting GH8 endoglucanase also revealed an oxocarbenium transition state structure during the reaction.⁶⁵ Interestingly, the existence of covalent glycosyl–enzyme intermediate is not just restricted in the retaining β -glycosidases. In the hen egg white

SCHEME 3: Proposed Catalytic Mechanism for the Hydrolysis of Cellotetraose Catalyzed by Cel5A from *A. cellulolyticus*^a



^a For a clear view of the mechanism, we plot the E282 and E162 residues on different sides of the substrate.

lysozyme (HEWL), a recent QM/MM study²⁰ clearly showed that the formation of an oxocarbenium transition state and a stable glycosyl–enzyme intermediate during the hydrolysis of its β -glycosidic bond, which is in contrast to a long-standing mechanism proposed by Philips.⁶⁶ The mechanism proposed therein is quite close to the Koshland’s double displacement mechanism. Obviously, the formation of oxocarbenium ion transition state could be the major characteristics for the hydrolysis of β -glycosidic linkage.

Another important issue related with *A. cellulolyticus* Cel5A is that the mutation of Y245G can increase catalysis rate when combined with Cel7A.²⁹ In their study, Baker et al. found that the cellobiose binding free energy is about 7.9 kcal/mol (derived from the binding constant of $K_i = 1.88$ mM). It was claimed that the binding energy of cellobiose product is reduced for the Y245G mutant. In our simulation, the phenol group of Y245 does form the near parallel interaction with the pyranose ring at Glc1 subsite. That means the mutation of Y245 to G245 should facilitate the releasing of cellobiose leaving group. However, the free energy barrier for the celldextrin with DP = 6 in *B. agaradherans* Cel5A is estimated about 15.3 kcal/mol. The releasing of product cellobiose might not be the rate-limiting step for the whole reaction. On the other hand, since we do not know the catalysis rate for the cellulose substrate hydrolysis, that makes it very hard to believe that the reduction of the binding affinity of cellobiose could be the reason for the increased catalysis rate. In addition, the hydrolysis of cellulose catalyzed by cellulases is a synergistic process, which generally involves with three different kinds of cellulases. Therefore, for successful protein engineering to increase the rate of hydrolysis of cellulose, it might be desirable to consider the synergistic effects. Indeed, such kind of work has recently been carried out by Zhang and Lynd using a kinetic model.⁶⁷

Another possible way to improve the catalytic efficiency of this enzyme is to perform site-directed mutagenesis studies on the carbohydrate binding module (CBM). The CBM plays essential role of cellulose binding before the catalytic domain begins its work. Combined with with mutagenesis around the catalytic domain, mutation studies on CBM for Cel5A from *Thermotoga maritima* did show a 14–18-fold increase of the catalytic efficiency.³¹ Therefore, for a successful protein engineering for this particular enzyme or other cellulases, it might not be enough to consider the catalytic domain only.

4. Conclusions

Given the importance of cellulose-based biomass, a detailed understanding of the hydrolysis mechanism for cellulases is important for designing new and more powerful catalysts. In this work, we studied the hydrolysis mechanism catalyzed by Cel5A from *A. cellulolyticus* using cellobotetraose as the substrate molecule. A QM/MM approach was applied to simulations of enzyme–substrate and glycosyl–enzyme intermediate, and potentials of mean force for both glycosylation and deglycosylation processes. Due to the geometrical characteristics of the active site, we did not pursue the possibility of inversion mechanism for this enzyme. Our computational results indicated that the full reaction process via a retention mechanism is reasonable, with E282 as the nucleophile, and E162 as the general acid/base. Although the Y245G mutant was shown to increase the rate of hydrolysis, it is not clear that this mutation could be the only reason to lead that. Further studies should include the binding mode of cellulose to CBM, and the linker region between CBM and the catalytic domain.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (No. 20803048).

Supporting Information Available: Time evolution of the distances between oxygen atoms of water molecules and C1(3) atom and the corresponding distances between hydrogen atoms of these waters and the Glu162 carboxylate group. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Demain, A. L.; Newcomb, M.; Wu, J. H. D. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 124.
- (2) Galbe, M.; Zacchi, G. *Appl. Biochem. Biotechnol.* **2002**, *59*, 618.
- (3) Lynd, L. R.; Cushman, J. H.; Nichols, R. J.; Wyman, C. E. *Science* **1991**, *251*, 1318.
- (4) Wang, D. I. C.; Avgerinos, G. C.; Biocic, I.; Wang, S.-D.; Fang, H.-Y. *Philos. Trans. R. Soc. London B* **1983**, *300*, 323.
- (5) Zhang, Y.-H. P.; Himmel, M. E.; Mielenz, J. R. *Biotechnol. Adv.* **2006**, *24*, 452.
- (6) Henrissat, B.; Bairoch, A. *Biochem. J.* **1996**, *316*, 695.
- (7) Koshland, D. E. *J. Biol. Rev.* **1953**, *28*, 416.
- (8) Vasella, A.; Davies, G. J.; Bohm, M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 619.
- (9) Zechel, D. L.; Withers, S. G. *Acc. Chem. Res.* **2000**, *33*, 11.
- (10) Zechel, D. L.; Withers, S. G. *Curr. Opin. Chem. Biol.* **2001**, *5*, 643.
- (11) Vocadlo, D. J.; Davies, G. J. *Curr. Opin. Chem. Biol.* **2008**, *12*, 539.
- (12) McCarter, J.; Withers, S. G. *Curr. Opin. Struct. Biol.* **1994**, *4*, 885.
- (13) Wang, Q.; Graham, R. W.; Trimbur, D.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1994**, *116*, 11594.
- (14) Davies, G. J.; Mackenzie, L.; Varrot, A.; Dauter, M.; Brzozowski, A. M.; Schulein, M.; Wither, S. G. *Biochemistry* **1998**, *37*, 11707.
- (15) Varrot, A.; Davies, G. J. *Acta Crystallogr.* **2003**, *D59*, 447.
- (16) Sinnott, M. L.; Souchard, I. J. *Biochem. J.* **1973**, *133*, 89.
- (17) Bause, E.; Legler, G. *Biochim. Biophys. Acta* **1980**, *626*, 459.
- (18) Wither, S. G.; Street, I. P.; Bird, P.; Dolphin, D. H. *J. Am. Chem. Soc.* **1987**, *109*, 7530.
- (19) Uitdehaag, J. C. M.; Mosi, R.; Kalk, K. H.; van der Veen, B. A.; Dijkhuizen, L.; Wither, S. G.; Dijkstra, B. W. *Nat. Struct. Biol.* **1999**, *6*, 432.
- (20) Bowman, A. L.; Grant, I. M.; Mulholland, A. J. *Chem. Commun.* **2008**, 4425.
- (21) Bottone, A.; Miscione, G. P.; De Vivo, M. *Proteins—Struct. Funct. Bioinform.* **2005**, *59*, 118.
- (22) Biarnes, X.; Nieto, J.; Planas, A.; Rovira, C. *J. Biol. Chem.* **2006**, *281*, 1432.
- (23) Biarnes, X.; Ardevol, A.; Planas, A.; Rovirat, C.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2007**, *129*, 10686.
- (24) Sakon, J.; Adney, W. S.; Himmel, M. E.; Thomas, S. R.; Karplus, P. A. *Biochemistry* **1996**, *35*, 10648.
- (25) Chhabra, S. R.; Shockley, K. R.; Ward, D. E.; Kelly, R. M. *Appl. Environ. Microbiol.* **2002**, *68*, 545.
- (26) Davies, G. J.; Dauter, M.; Brzozowski, A. M.; Bjornvad, M. E.; Anderson, K. V.; Schulein, M. *Biochemistry* **1998**, *37*, 1926.
- (27) Varrot, A.; Schulein, M.; Pipelier, M.; Vasella, A.; Davies, G. J. *J. Am. Chem. Soc.* **1999**, *121*, 2621.
- (28) Varrot, A.; Schulein, M.; Fruchard, S.; Driguez, H.; Davies, G. J. *Acta Crystallogr.* **2001**, *D57*, 1739.
- (29) Baker, J. O.; McCarley, J. R.; Lovett, R.; Yu, C. H.; Adney, W. S.; Rignall, T. R.; Vinzant, T. B.; Decker, S. R.; Sakon, J.; Himmel, M. E. *Appl. Biochem. Biotechnol.* **2005**, *121*–124, 129.
- (30) McCarter, S. L.; Adney, W. S.; Vinzant, T. B.; Jennings, E.; Eddy, F. P.; Decker, S. R.; Baker, J. O.; Sakon, J.; Himmel, M. E. *Appl. Biochem. Biotechnol.* **2002**, *98*–100, 273.
- (31) Mahadevan, S. A.; Wi, S. G.; Lee, D.-S.; Bae, H.-J. *FEMS Microbiol. Lett.* **2008**, 287, 205.
- (32) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
- (33) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
- (34) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298.
- (35) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425.
- (36) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G.; Xu, D.; Guo, H.; Elstener, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458.
- (37) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.
- (38) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
- (39) Alhambra, C.; Corchado, J. C.; Sanchez, M. L.; Gao, J.; Truhlar, D. G. *J. Am. Chem. Soc.* **2000**, *122*, 8197.

- (40) Xu, D.; Wei, Y.; Wu, J.; Dunaway-Mariano, D.; Guo, H.; Cui, Q.; Gao, J. *J. Am. Chem. Soc.* **2004**, *126*, 13649.
- (41) Xu, D.; Guo, H. *J. Phys. Chem. B* **2008**, *112*, 41.
- (42) Friesner, R.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389.
- (43) Corminboeuf, C.; Hu, P.; Tuckerman, M. E.; Zhang, Y. *J. Am. Chem. Soc.* **2006**, *128*, 4530.
- (44) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. *Phys. Status Solidi* **2000**, *B217*, 357.
- (45) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.
- (46) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565.
- (47) Xu, D.; Guo, H.; Cui, Q. *J. Phys. Chem. A* **2007**, *111*, 5630.
- (48) Xu, D.; Guo, H.; Cui, Q. *J. Am. Chem. Soc.* **2007**, *129*, 10814.
- (49) Xu, D.; Xie, D.; Guo, H. *J. Biol. Chem.* **2006**, *281*, 8740.
- (50) Xu, D.; Zhou, Y.; Xie, D.; Guo, H. *J. Med. Chem.* **2005**, *48*, 6679.
- (51) MacKerell, A. D., Jr.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586.
- (52) IUPAC-IUB. *Eur. J. Biochem.* **1983**, *131*, 5.
- (53) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (54) Brooks III, C. L.; Brünger, A.; Karplus, M. *Biopoly.* **1985**, *24*, 843.
- (55) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University: Oxford, UK, 1986.
- (56) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667.
- (57) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
- (58) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- (59) Woodcock, H. L.; Hodosek, M.; Brooks, B. R. *J. Phys. Chem. A* **2007**, *111*, 5720.
- (60) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
- (61) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
- (62) Frisch, M. J.; et al. Gaussian03; Gaussian, Inc.: Pittsburgh, PA, 2003.
- (63) Guest, M. F.; Bush, I. J.; van Dam, H. J. J.; Sherwood, P.; Thomas, J. M. H.; van Lenthe, J. H.; Havenith, R. W. A.; Kendrick, J. *J. Mol. Phys.* **2005**, *103*, 719.
- (64) Larsson, A. M.; Anderson, L.; Xu, B.; Munoz, I. G.; Uson, I.; Janson, J.-C.; Stalbrand, H.; Stahlberg, J. *J. Mol. Biol.* **2006**, *357*, 1500.
- (65) Petersen, L.; Ardèvol, A.; Rovira, C.; Reilly, P. J. *J. Phys. Chem. B* **2009**, *113*, 7331.
- (66) Phillips, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 483.
- (67) Zhang, Y.-H. P.; Lynd, L. R. *Biotechnol. Bioeng.* **2006**, *94*, 888.

JP909177E