

Glycosidic-Bond Hydrolysis Mechanism Catalyzed by Cellulase Cel7A from *Trichoderma reesei*: A Comprehensive Theoretical Study by Performing MD, QM, and QM/MM Calculations

Jinghua Li,[†] Likai Du,[‡] and Lushan Wang^{*,†}

State Key Laboratory of Microbial Technology and Institute of Theoretical Chemistry, Shandong University, Jinan, 250100, P. R. China

Received: July 12, 2010; Revised Manuscript Received: September 8, 2010

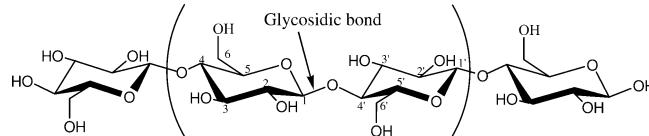
Cellulase Cel7A from *Trichoderma reesei* is one of the most abundant and effective cellulases. Structural studies have established that Cel7A is a retaining glycosidase and it can processively hydrolyze cellobiose units from the reducing end of a cellulose chain. Here, to elucidate the mechanism of enzymatic catalysis of cellulase Cel7A, we carried out a multisized level theoretical study by performing MD, QM, and QM/MM calculations. At the accurate level of theory, we showed the mechanism details of the catalytic cycle, which involves the configuration inversion of the anomeric center twice: the first results in the glycosidic bond cleavage and the formation of covalent glycosyl–enzyme intermediate, and the second restores the anomeric carbon to its original configuration. Calculated results have provided detailed structural and energetic information about these two processes, both of which proceed according to a S_N2 -type-like mechanism via loose transition state structures. It is clearly indicated that the glycosidic bond hydrolysis involves the formation of a covalent glycosyl–enzyme intermediate, which has been identified as the minimum on the potential energy surface. At the catalytic active region, hydrogen bond interactions exist throughout the whole process of the catalytic cycle, which are of special importance for stabilizing the glycosyl–enzyme intermediate. The present results provide a clear paradigm of the mechanisms of general glycosidases, which hydrolyze the glycosidic bonds with net retention of the anomeric configuration.

1. Introduction

Cellulose, as the major component of plant cell walls, is the most abundant and renewable biomass source on the planet,¹ which is a polysaccharide formed by the polymerization of a few hundred to over several thousand β -(1,4)-D-glucose units,^{2,3} as shown in Scheme 1. With the excessive consumption of nonrenewable resources, cellulose with renewability and biodegradability has been attracting more and more attention. If cellulose could be degraded into small polysaccharides or glucose units, it could be used as a cheap and nontoxic raw material to produce numerous useful products.^{4–7} However, there are still some open issues hindering the utilization of cellulose, in which one principal challenge is its high resistance to both chemical hydrolysis and enzymatic degradation due to the very stable β -glycosidic bonds between glucose units and the intra- and intermolecular hydrogen bonds between cellulose chains in the crystal.^{8,9}

As one kind of glycoside hydrolases (glycosidases), cellulases are of high diversity, belonging to different families of the carbohydrate-active enzymes. Synergistic mixtures of different cellulases produced by various microbes can completely degrade native cellulose to soluble sugars.¹⁰ At present, the biodegradation technology of crystalline cellulose, the cleanest way to degrade cellulose, is still not effective and economical enough and therefore does not meet the commercial-scale production cost and industrial application yet.¹¹ The major bottleneck for the efficient and economical use of cellulose is the low con-

SCHEME 1: Cellulose, a Polysaccharide Consisting of β -1,4 Linked D-Glucose Units (Four Are Shown)^a



^a The two glucose units in the parentheses consist of a basic block of cellulose.

version rate.¹² Molecular bioengineering technology is expected to be able to produce more efficient cellulases, once the detailed catalytic mechanism of cellulases is completely mastered. Therefore, understanding the catalytic mechanism by which cellulose degrades into simple sugar units for further fermentation into ethanol or other chemicals is vital for improving the enzymatic degradation efficiency.

One of the most abundant and effective cellulases is Cel7A from *Trichoderma reesei* (also known as CBHI, which stands for cellobiohydrolase I). Cel7A has received considerable interest over the past two decades, and relevant research has provided valuable information for understanding its structure and function.^{13,14} It is now generally accepted that Cel7A possesses a large catalytic domain containing tunnel-shaped active sites formed by long loops on the surface, and it hydrolyses a cellobiose units from the reducing end.^{15–17} Cel7A is described as a “processive” enzyme that is hypothesized to cut a cellobiose unit off per catalytic event continuously before the substrate is separated from it and the resultant product leaves the enzyme from the other side of the tunnel.⁷ Although a lot is known about Cel7A from previous studies, our knowledge of the

* Corresponding author. E-mail: lswang@sdu.edu.cn. Tel: 086-531-88366202. Fax: 086-531-88565610.

[†] State Key Laboratory of Microbial Technology.

[‡] Institute of Theoretical Chemistry.

catalytic mechanism is still far from complete, and many mechanistic aspects have not been elucidated yet. In particular, also as a common concern for generally enzymatic hydrolysis of glycosidic linkage, the structural details of intermediates and transition states involved during the hydrolysis of the glycosidic bond are still not fully understood at the atomic and molecular levels.

It is well-known that enzyme-catalyzed reactions are so fast that the received information from experimental studies regarding enzyme functioning mechanism is limited and indirect. Theoretical studies using quantum mechanics (QM) calculations and molecular dynamics (MD) simulations based on molecular mechanics (MM) provide one of the most direct ways to reveal mechanism details at the molecular level. As far as we know, theoretical studies on cellulases are very limited, and only a few MD studies on this subject can be found in the literature.^{18–21} Because of the lack of QM calculations, the structural information (bond breaking and bond formation) regarding intermediates and transition states involved in the catalytic cycle of cellulases, which importantly contribute to the refinement of mechanistic details, remains unclear. Of course, accurate QM calculations are only available for small systems typically containing several tens of atoms and not practical for large systems such as cellulases due to the limitation of current computer capabilities. In recent years, combined quantum mechanics/molecular mechanics (QM/MM) methods have been established as a valuable tool for large systems consisting of up to a few thousands of atoms, where QM methods are employed to describe the active site of chemical reactions and MM approaches are used to mimic the environmental effect on the active site.^{21–23} A lot of successful applications of QM/MM methods have been achieved for studying enzymatic reactions,^{24–26} indicating the reliability of the existing QM/MM methods for describing real-world chemical reactions.

Here, we present a comprehensive theoretical study by performing MD, QM, and QM/MM calculations on the representative cellulase Cel7A from *Trichoderma reesei*, for which there has been only very limited theoretical investigation. Our prime goal is to shed some light over the catalytic mechanism at the molecular, atomic, and electronic levels, which are not accessible by experimental approaches currently. We will elucidate the binding behavior of cellulose on the enzyme, structural and energetic details of intermediates and transition states during the catalytic cycles.

2. Computational Details

2.1. MD Simulation. The simulation system was prepared on the basis of the X-ray crystal structure of TrCel7A containing a cellulose chain¹⁷ (PDB ID: 8CEL) obtained from Protein Data Bank (<http://www.pdb.org>) (see Figure 1). Before simulation, residue Glu217 was protonated and the mutated residue Gln212 was altered to Glu212, which reveals the normal condition in enzymes of the wild type. Hydrogen atoms not included in the PDB structure were added to the system using the VMD 1.8.6 program.²⁷ The enzyme–substrate complexes were solvated in a box of explicit water with the dimension of $80 \times 74 \times 90 \text{ \AA}^3$, containing NaCl at a concentration of 50 mmol/L. Molecular dynamics simulations were performed using NAMD 2.6 program.²⁸ The length of each time step was set to 1 fs and the temperature was kept at 310 K, the optimum temperature for the catalysis of CBHI. The system was first energetically minimized by performing a MD simulation of 5000 steps, which was then equilibrated thermally for 2 ns to reach the equilibration state of the system.

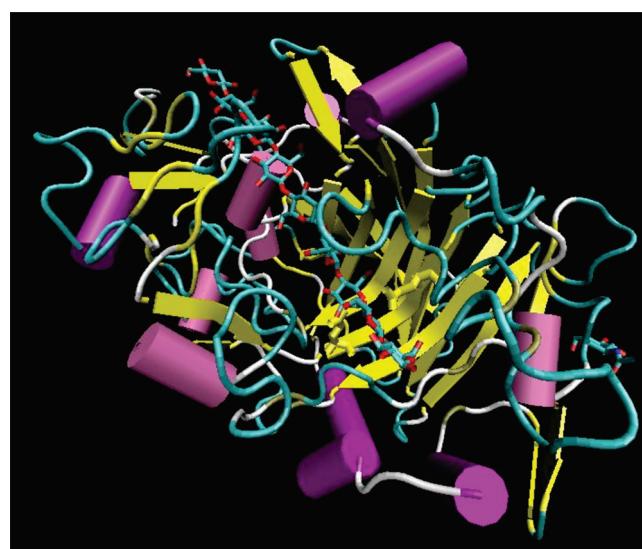


Figure 1. X-ray structure of TrCel7A containing a cellulose chain obtained from the Protein Data Bank (<http://www.pdb.org>), where water molecules are not included for clarity.

2.2. QM Calculations. The model systems for QM calculations that would mimic the detailed glycosidic-bond hydrolysis mechanism were built based on the NAMD equilibrium state. In order to be computationally feasible, the model only included the two catalytic residues (protonated Glu217 and deprotonated Glu212) and one basic unit of cellulose composed of two sugar rings, Glc453 and Glc454. The calculations were carried out by the Gaussian 03 program package,²⁹ using the hybrid B3LYP functional³⁰ combined with the standard 6-31G(d,p) basis set. Full geometry optimizations of minima and transition states were performed at the selected level of theory. Frequency calculations at the same level of theory have also been carried out to verify all stationary points as minima (zero imaginary frequencies) or first-order saddle points (one imaginary frequency). The intrinsic reaction coordinates (IRC)^{31,32} were used with the aim of confirming that all located transition states actually connected two related reactants and products.

2.3. Combined QM/MM Study. The equilibrium system from NAMD simulations was minimized and equilibrated again with the CHARMM program by setting a 14 Å nonbonded cutoff using the CHARMM22 force field.³³ The protein and the substrate were first kept fixed, while the water molecules and sodium ions were allowed to move freely, and then the constraints were gradually removed. The equilibrated system was taken as the starting point for QM/MM calculations, where the residues, Glu212 and Glu217, and sugar rings, Glc453 and Glc454, were included in the QM region. The total charge of the system is -1 . The QM part was treated at the B3LYP/6-31G(d) level and the MM part was described by the CHARMM22 force field. For the QM/MM boundary, we adopted an electronic embedding scheme³⁴ incorporating the MM charges into the one-electron Hamiltonian of the QM treatment and hydrogen link atoms with charge shift model (Figure 2).³⁵ The calculations were performed with the ChemShell³⁶ package integrating TURBOMOLE^{37–39} and DL-POLY⁴⁰ programs. The default convergence criteria have been employed in the QM/MM geometry optimization with the HDLC optimizer⁴¹ for the reactant and product. In the transition state search, an elaborate initial structure from the potential energy surface scan has been adopted. Finally, the single-point QM/MM calculations were performed using the larger 6-311++G(d,p) basis set to obtain accurate energies.

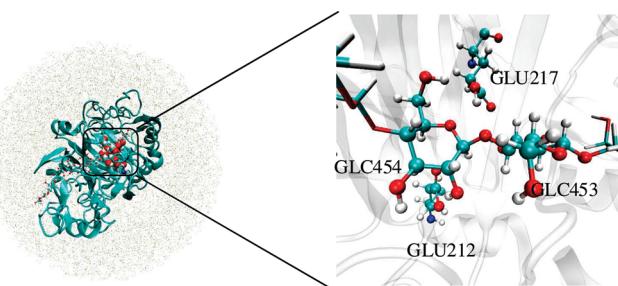


Figure 2. Prepared model for the QM/MM calculations, where the region enclosed in the rectangle is the catalytic active site which is highlighted by the vdW structure on the right side for the sake of clarity.

3. Results and Discussion

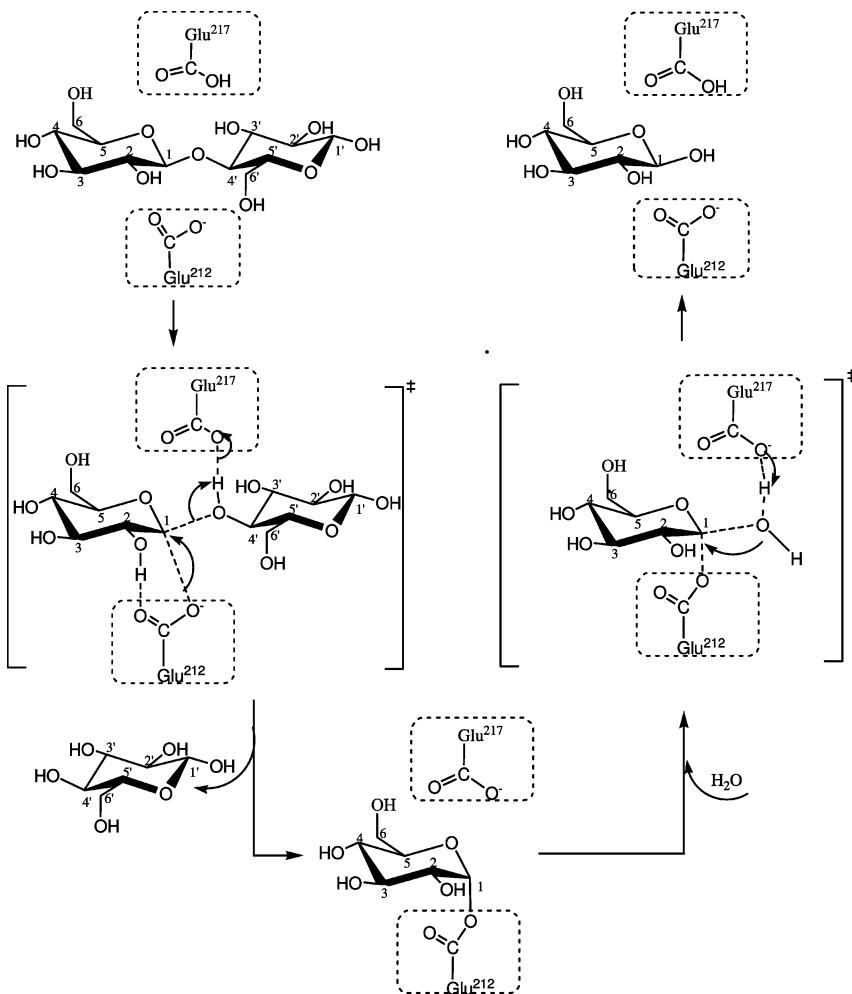
3.1. General Consideration of the Catalytic Mechanism.

Like other glycoside hydrolases, cellulases are generally divided into a retaining set and an inverting set, which involve different catalytic mechanisms giving rise to either an overall retention or an inversion of the anomeric carbon configuration depending on the nature of the nucleophile. Cel7A is known as a configuration-retaining cellulase,^{42,43} which hydrolyzes the β -1,4-glycosidic bond via a general acid/base catalysis that requires two critical residues: one of which acts as a proton donor and the other as a nucleophile/base. These two catalytic residues were identified as Glu217 and Glu212 for TrCel7A.¹⁷ As usually

known, the glycosidic-bond hydrolysis catalyzed by a retaining cellulase takes place via a double displacement mechanism,⁴⁴ where the β -conformation of the anomeric carbon (C_1) in cellobiose is maintained after a catalytic cycle. The double displacement mechanism imagines two elementary steps:⁴⁵ (1) Glu217 acts as a general acid to protonate the glycoside oxygen, and at the same time Glu212 as a nucleophile attacks the anomeric carbon to form a covalent glycosyl–enzyme intermediate, and (2) an incoming water molecule, which is being deprotonated by Glu217, attacks the anomeric carbon and replaced the Glu212. Thus, in each catalytic cycle Cel7A cuts a cellobiose unit off from the cellulose chain. These two steps can be described as “the formation of covalent glycosyl–enzyme intermediate” and “the recovery of the anomeric carbon”, respectively. According to these fundamental suppositions, which have been used as initial guesses for our QM calculations, we schematically illustrate the glycosidic-bond hydrolysis mechanism on the active site of Cel7A in Scheme 2. The structural details of intermediates and transition states involved in Scheme 2 will be discussed in the following sections.

3.2. Enzyme–Substrate Complex. During MD simulations, the enzyme–substrate complex is constantly disturbed by thermal fluctuations. Throughout the whole process of simulation, however, no significant displacement for either of the enzyme molecule or the cellulose chain was observed. In other words, in the thermally equilibrated state of system, TrCel7A

SCHEME 2: Schematic Illustration of the Glycosidic-Bond Hydrolysis Mechanism on the Active Site of TrCel7A According to the Double Displacement Mechanism Proposed by Koshland⁴⁶



almost remains at the original location and the cellulose chain is bound within the catalytic tunnel of TrCel7A at seven sites from -6 to $+2$. A supplemental movie file showing the thermal fluctuations of the enzyme–substrate complex during the MD simulations can be found in the Supporting Information.

In the thermally equilibrated system, it is observed that there are a considerable number of intermolecular hydrogen bonds between TrCel7A and the sugar chain in the catalytic tunnel, where all hydroxyl groups of the sugar substrate various amino acid residues are engaged to stabilize the enzyme–substrate complex. In particular, at the active center, the sugar substrate units at sites -1 and $+1$ are gripped by two catalytic residues Glu212 and Glu217, which are separated by a distance of ~ 6.0 Å, which is slightly larger than an suitable distance⁴⁶ of ~ 5.5 Å in a retaining enzyme for breaking the glycosidic bond via the double displacement mechanism. Furthermore, it should be noted that in the enzyme–substrate complex, because of the strong hydrogen bond interactions between the enzyme and substrate, the sugar ring at site -1 has been distorted into a half-chair or sofa conformation, which is favorable form in energy for the coming glycosidic bond hydrolysis and consistent with that found in lysozyme.^{47,48} This implies that the glycosidic bond hydrolysis catalyzed by different glycosyl hydrolases could follow a similar mechanism. Thus, it is expected that the present study using TrCel7A as a representative of glycosyl hydrolases is of general significance to understand the catalytic mechanism of glycosyl hydrolases.

3.3. Formation of Covalent Glycosyl–Enzyme Intermediate. We first focus on the first step of the catalytic cycle, the formation of covalent glycosyl–enzyme intermediate, which involves the glycosidic bond cleavage and is a crucial step of the enzymatic hydrolysis. By performing potential energy surface scan calculations for the model system chosen at the B3LYP/6-31G(d,p) level, we have located two minima, denoted as structures **I** and **II** in Figure 3, which correspond to the structures before and after the glycosidic bond cleavage, as well as the transition state breaking glycosidic bond, denoted as **TS_{I-II}** in Figure 3.

Structure **I** in Figure 3 is considered as the starting point of the reaction, which is a precursor complex consisting of the substrate and two catalytic residues, where protonated Glu217 and deprotonated Glu212 establish close contact with the substrate via hydrogen-bond interactions, as indicated by calculated H-bond distances of 1.741 and 1.538 Å. In this precursor, Glu212 lies on the back of the glycosidic bond of the anomeric carbon with a C–O distance of 3.083 Å and is ready for attacking the anomeric carbon by the carboxyl oxygen. Along the reaction coordinate, via transition state **TS_{I-II}** structure **I** evolves into structure **II**. In structure **II**, the glycosidic bond has been broken and a covalent glycosyl–enzyme intermediate has been formed, as indicated by the broken C–O distance of 3.894 Å and the formed C–O bond of 1.512 Å. The formation of structure **II** is regarded as a result of the proton migration from Glu217 to the glycosidic oxygen accompanied by the attack of Glu212 to the anomeric carbon.

The glycosidic bond cleavage process from structure **I** to structure **II** can be understood best by analyzing the structural details of **TS_{I-II}**, which has been confirmed by running IRC calculations to connect structures **I** and **II** with an imaginary frequency of 132*i* cm⁻¹. As shown in Figure 3, we have marked several crucial geometrical parameters in **TS_{I-II}**. Note that in this structure the glycosidic bond (2.052 Å) and the O–H bond (1.395 Å) in Glu217 are breaking, and the bond between the anomeric carbon and the carboxyl oxygen in Glu212 (2.605 Å)

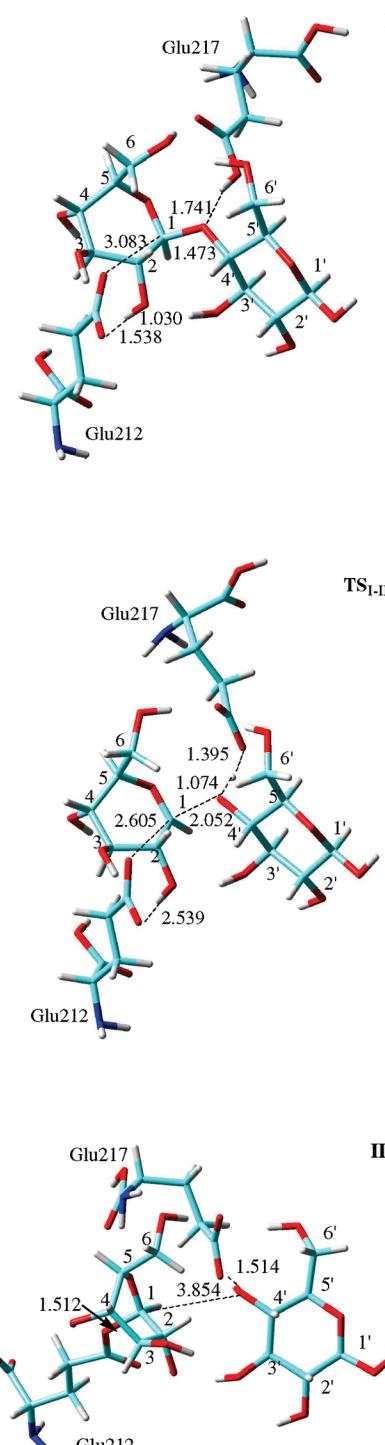


Figure 3. Structures involved in the glycosidic bond cleavage, calculated at the B3LYP/6-31g(d,p) level. The distances shown are in angstroms.

and the bond between the glycosidic oxygen and the hydroxyl hydrogen in Glu217 (1.074 Å) are forming. From these crucial geometrical parameters, it is clear that at the transition state stage the glycosidic bond is almost broken while the covalent bond between the nucleophile and the anomeric carbon is far from formed, indicating the proton transfer from Glu217 to the glycosidic oxygen is prior to the attack of the carboxyl oxygen in Glu212 to the anomeric carbon. Comparing the configurations of the anomeric carbon in structures **I**, **TS_{I-II}**, and **II**, we observe the inversion of configuration of the anomeric carbon along the reaction coordinate, which is an inherent characteristic of the

S_N2 mechanism. However, the glycosidic bond cleavage and the formation of covalent glycosyl–enzyme intermediate are not completely consistent with a classical S_N2 mechanism picture, which involves a bimolecular collision between the nucleophile and the leaving group to reach the transition state and results in a configuration inversion of the center carbon in product. Here, the enzymatic reaction is a trimolecular collision process, where Glu 217 works as a proton donor on the glycosidic oxygen and the other as a nucleophile attacking the anomeric carbon and these two residues concerted but asynchronously catalyze the glycosidic bond cleavage. We refer to the configuration inversion mechanism as the S_N2 -type-like mechanism. From structure **I** to structure **II** via **TS_{I-II}**, the main conformational change of the system is the rearrangement of the sugar ring Glc454, where the dihedral angle $O_{ring}C_1C_3C_4$ is 2.8° in structure **I**, 8.6° in **TS_{I-II}**, and 45.4° in structure **II**. In other words, with the weakening of the glycosidic bond, the deformation extent of the sugar ring Glc454 increases, while the sugar ring Glc453 presents a relative small rearrangement. The energy required to reach the transition state is calculated to be $14.13 \text{ kcal mol}^{-1}$, and the reaction is exothermic by $1.44 \text{ kcal mol}^{-1}$. Note that the active barrier calculated here for the glycosidic-bond cleavage is remarkably smaller than previous theoretical results obtained by Bottoni⁴⁹ ($22.7 \text{ kcal mol}^{-1}$ from the B3LYP functional with a locally dense basis set) and Bras⁵⁰ ($28.5 \text{ kcal mol}^{-1}$ from the B3LYP functional with the 6-31G(d) basis set). Both of them mimicked the substrate using a β -methylglucoside, where the methyl replaced the realistic six-membered sugar ring to reduce computational cost.^{49,50} The present calculations used a cellobiose, the basic building block unit of a cellulose chain, to mimic the realistic glycosidic bond, and were performed using a larger 6-31G(d,p) basis set, and thus are expected to give more reliable results.

To consider the influence of the surrounding around the active site on the enzymatic reaction in a more realistic system, we further performed a combined QM/MM calculation study for the formation of covalent glycosyl–enzyme intermediate. Figure 4 schematically shows the resultant structures in the QM region at three stages of reaction corresponding to structures **I**, **TS_{I-II}**, and **II** shown in Figure 3. The imaginary frequency of the transition state is $143i \text{ cm}^{-1}$, which is comparable with that ($132i \text{ cm}^{-1}$) from the gas-phase B3LYP calculation. It is noted that the geometrical characteristics for all three structures are very similar to those obtained using the pure QM method on the whole. So the mechanism details for the glycosidic bond cleavage will not be discussed again, and only the main geometrical parameters are given in Figure 4 for the sake of comparison. The most important difference between the combined QM/MM results and the pure QM calculations is that the hydrogen-bond interaction between the catalytic residues and the substrate seems to be weaker in the more realistic QM/MM system than that in the pure QM system, as shown by the calculated larger H-bond distances in Figure 4 than the corresponding those shown in Figure 3. Accordingly, the barrier breaking the glycosidic bond is verified to increase to $23.9 \text{ kcal mol}^{-1}$ and the reaction process is calculated to be endothermic by $2.60 \text{ kcal mol}^{-1}$, in contrast to $14.13 \text{ kcal mol}^{-1}$ and exothermic by $1.44 \text{ kcal mol}^{-1}$ in the pure QM calculations. The higher barrier from the QM/MM calculation may be explained as following: in the QM/MM method, the space and orientation of the catalytic amino acid residues are restricted by peptide bonds which link them to other residues in the enzyme. In contrast, the two catalysts in the QM calculations are allowed to move freely to search for the minima and the

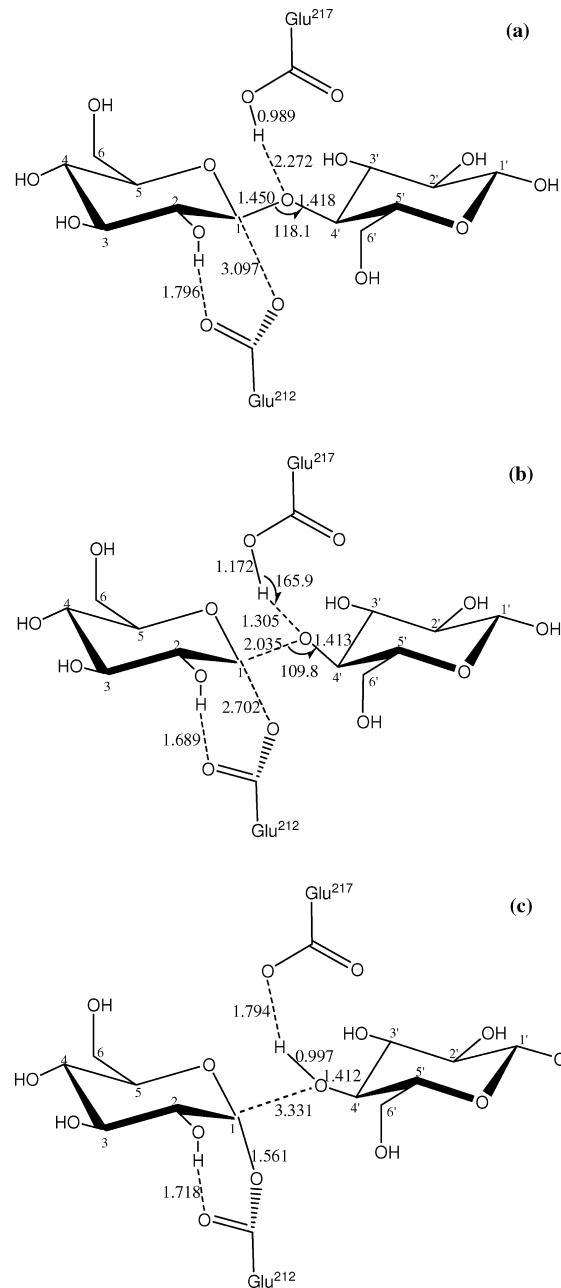


Figure 4. Schematic representation for the QM/MM optimized structures in the QM region: (a), (b), and (c) correspond to structures **I**, **TS_{I-II}**, and **II** shown in Figure 3, respectively. The distances shown are in angstroms.

transitional state on the potential energy surface. These results imply that the surrounding around the active site has an important influence on the thermodynamics and kinetics of reaction, but almost does not influence the intrinsic natures of species involved in reaction process, for example, the most vivid characteristic of the transition state structure is that both the glycosidic oxygen and the carboxyl oxygen of the nucleophile are loosely bound to the anomeric carbon. The larger barrier from the QM/MM calculations is attributed to the fact that both the two catalytic residues Glu212 and Glu217 are polar and possess strong dipole interactions with the water molecules in surroundings, which reduce the catalytic activity of the residues.

It should be emphasized that the present result for the effect of the solvent environment on the reactivity is in contrast to early findings,^{49,50} where the inclusion of solvent environment using the solvent continuous model approach rather than the

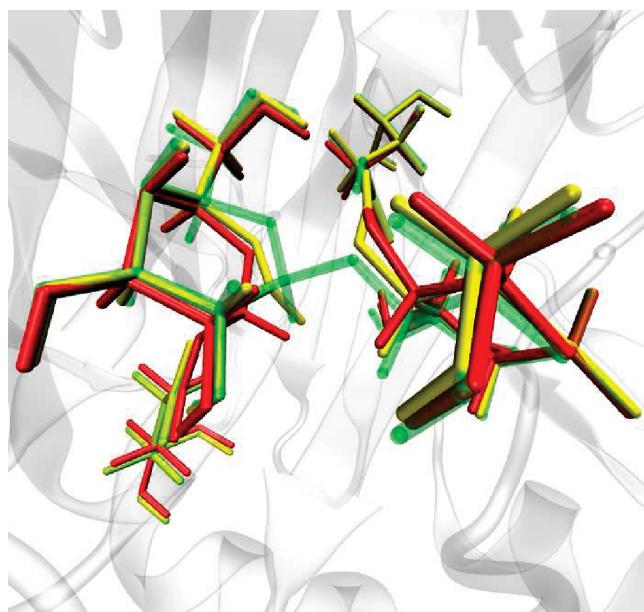


Figure 5. Superposition of structures at the active center during the glycosidic bond cleavage: the green, yellow, and red correspond to the substrate–residues complex, the transition state, and the glycosyl–enzyme intermediate, respectively.

realistic solvent environment used here was not found to change significantly the barrier obtained with the gas-phase model.

Similar to the findings from pure QM calculations, the present QM/MM results also show a remarkable conformation rearrangement of sugar ring Glc454 during the glycosidic bond cleavage. Figure 5 shows the superposition of three structures described by panels (a), (b), and (c) in Figure 4, where the green, yellow, and red ones denote the substrate–residues complex, the transition state, and the glycosyl–enzyme intermediate, respectively. Clearly, the anomeric carbon (C_1) moves the longest distance to trap the nucleophilic oxygen atom in Glu212 during the glycosidic bond cleavage, making the conformation of sugar ring Glc454 change from the chairlike form in the substrate–residues complex to the half-chair form in the transition state and to the chairlike form again in the glycosyl–enzyme intermediate.

3.4. Recovery of the Anomeric Carbon. It is worth to point out that the configuration of the anomeric carbon in structure **II** is inverted. In the subsequent step, the nucleophilic attack of a water molecule on the anomeric carbon causes its second configuration inversion, thus restoring its configuration to the original state.

From the discussion above, we see that the present QM results provide reliable information for understanding the formation mechanism of covalent glycosyl–enzyme intermediate, although the relative energies for the species established at the three stages are different from the QM/MM calculations, which is not our main concern in the present work. So in the following section, using the gas-phase model system, we further perform QM calculations to understand the recovery of the anomeric carbon, which is the second step of the catalytic cycle and is achieved by the attack of a water molecule on the glycosyl–enzyme intermediate.

Figure 6 shows the optimized structures involved in this process, where structures **III** and **IV** correspond to two minima before and after the recovery of the anomeric carbon, and structure **TS_{III-IV}** is the transition state connecting these two minima. In structure **III**, a water is located at between Glu212

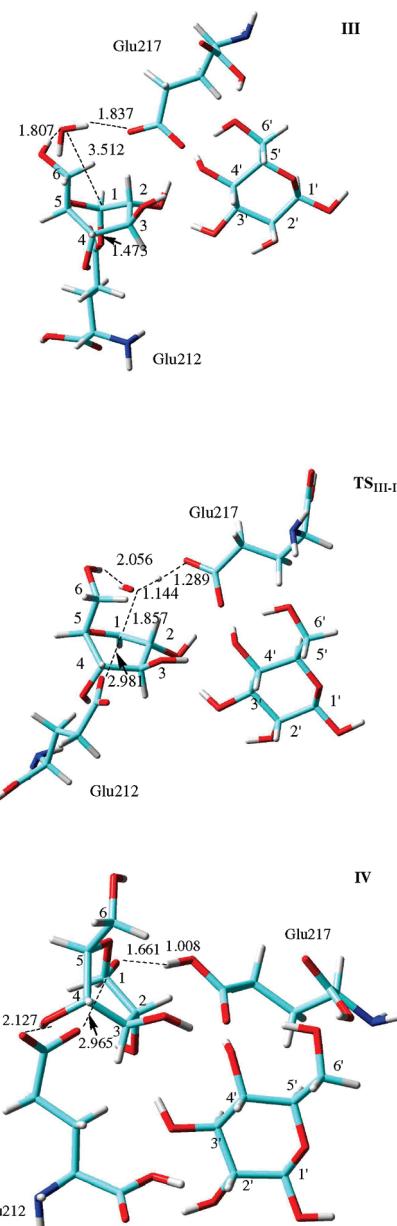


Figure 6. Structures involved during the recovery of the anomeric carbon, calculated at the B3LYP/6-31g(d,p) level. The distances shown are in angstroms.

and C_6 –OH group to form two O –H \cdots O H-bonds with distances of 1.807 and 1.837 Å, respectively, and it lies above the anomeric carbon with a distance of 3.512 Å between C_1 and the O atom of water. Structure **III** is considered as a water-stabilizing glycosyl–enzyme intermediate, which is a momentary state that the glycosyl–enzyme intermediate is just formed and not disturbed yet. This structure is geometrically favorable for the sequent attack of the –OH group in the water molecule to C_1 with the simultaneous transfer of the proton in water molecule to Glu 217, which leads structure **III** to structure **VI**, a product-like complex. The transition state connecting structures **III** and **IV** is structure **TS_{III-IV}**, the geometrical parameters of which a picture for the recovery of the anomeric carbon emerges clearly. As seen from Figure 6, in **TS_{III-IV}** Glu212 is moving away from the anomeric carbon upon the attack of the –OH group, and at the same time the proton in the water molecule is migrating to the carboxyl oxygen. The forming C–O and O–H distances are 1.857 and 1.289 Å, and the breaking C–O and O–H distances are 2.981 and 1.144 Å, respectively. From these

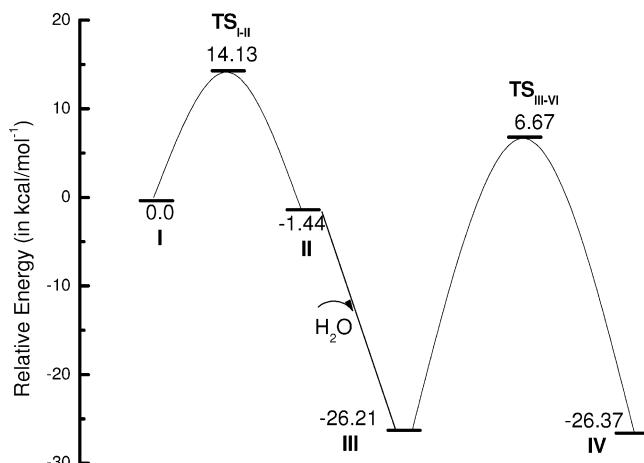


Figure 7. Calculated energy profile calculated at the B3LYP/6-31g(d,p) level for two elementary steps of the glycosidic bond hydrolysis.

geometrical parameters, it is clear that like **TS_{I-II}**, **TS_{III-IV}** is also very loose structurally, where the C–O bond in the glycosyl–enzyme intermediate has almost ruptured while the restoring hydroxyl group is still far from the anomeric carbon. The imaginary frequency for **TS_{III-IV}** is $419i\text{ cm}^{-1}$, and its normal mode clearly corresponds to the desired the bond-forming and bond-breaking motions. The calculated barrier from structure **III** to **TS_{III-IV}** is $32.9\text{ kcal mol}^{-1}$, which is much higher than that for the glycosidic bond cleavage, $14.1\text{ kcal mol}^{-1}$. This case is in contrast to the early study of Bottone et al.⁴⁹ For the lysozyme hydrolysis mechanism, the second elementary step of the reaction, the attack of the water molecule on the anomeric carbon, was found to be energetically less required than that of the first elementary step of the reaction, the glycosidic bond cleavage. Once the barrier from structure **III** to **TS_{III-IV}** is crossed, structure **IV** is formed subsequently and thereby a catalytic cycle is completed with the second configuration inversion of the anomeric center. In structure **IV**, catalytic residues 217 and 212 are now protonated and deprotonated, respectively, and thus are ready for acting as the active center of next catalytic cycle. From the calculated geometrical parameters shown in Figure 6, it is obvious that the second configuration inversion also proceeds according to the S_N2 -type-like mechanism.

The energy profile for the two elementary steps discussed above is shown in Figure 7. We see that the whole reaction is calculated to be exothermic by $26.4\text{ kcal mol}^{-1}$. From this energy profile, it is obvious that the elementary step barrier for the second step is larger than that for the first step, seemingly implying that the second step could be a bottleneck step of the enzymatic catalysis. As generally known, however, glycosidases accomplish the hydrolysis of cellulose in an up to 1000^{-1} rate,⁴⁹ a so fast process that the energy released from forming water-stabilizing covalent glycosyl–enzyme intermediate can compensate partly the subsequent energy requirement accomplishing the second configuration inversion rather than releasing it into the surroundings. In other words, the newly formed water-stabilizing covalent glycosyl–enzyme intermediate may be facile to perform the second configuration inversion to complete the catalytic cycle with a net energy requirement of 6.7 kcal mol^{-1} . In this sense, the first step of the reaction should be considered as the rate-determining step for the enzymatic hydrolysis of cellulose with a barrier of $14.1\text{ kcal mol}^{-1}$, and the life of the covalent glycosyl–enzyme intermediate is expected to be very short, which explains the difficulty in experimentally detecting this intermediate, although it is cal-

culated to be a global minimum on the potential energy surface. As shown in Figure 7, **TS_{I-II}** is a global maximum on the potential energy surface, and once it is crossed the hydrolysis reaction will be facile to accomplish.

4. Conclusive Remarks

In this paper, a comprehensive theoretical study of the catalytic mechanism of TrCel7A has been carried out by performing MD, QM, and QM/MM calculations. On the basis of the discussion above, we summarize the most significant conclusions drawn out from the present work as follows:

(1) At the thermal equilibrated state of system (the initial stage of the catalytic cycle), sugar ring Glc 454 containing the scissile glycosidic bond has distorted into a half-chair configuration mainly via hydrogen bond interactions between catalytic residues Glu212 and Glu217 with hydroxyl group at sugar rings Glc 453 and Glc 454.

(2) The mechanism details of the catalytic cycle have been shown at the accurate level of theory, which are on the whole consistent with generally accepted assumption of double-displacement mechanism that involves the configuration inversion of the anomeric center twice: the first results in the glycosidic bond cleavage and the formation of covalent glycosyl–enzyme intermediate, and the second restores the anomeric carbon to its original configuration.

(3) Both the configuration inversions proceed via the S_N2 -type-like mechanism involving loose transition state structures. The first step of reaction is considered as the rate-determining step for the enzymatic hydrolysis of cellulose with a barrier of $16.5\text{ kcal mol}^{-1}$, which is significantly smaller than previous theoretical results of glycosidic bond cleavages from gas-phase model calculations.^{49,50}

(4) The inclusion of solvent surroundings remarkably increases the barrier due to reducing hydrogen bond forces between catalytic residues and the substrate. This case is in contrast to early findings,^{49,50} where the protein environment mimicked using the solvent continuous model approach rather than the realistic solvent environment used here and was not found to significantly change the barrier obtained with the gas-phase model.

(5) At the catalytic active region, hydrogen bond interaction exists throughout the whole process of the catalytic cycle, which is of special importance for stabilizing the glycosyl–enzyme intermediate.

(6) The mechanism picture described here for the glycosidic bond hydrolysis is expected to present a versatile paradigm of the mechanisms of general glycosidases with net retention of the anomeric configuration.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (31070063, 20773078), the Scientific Research Reward Fund for Excellent Young and Middle-Aged Scientists in Shandong Province (BS2009SW021), and the Independent Innovation Foundation of Shandong University (2009JC006). We greatly acknowledge Professor Dongju Zhang from the College of Chemistry and Chemical Engineering of Shandong University for useful and helpful discussions. The authors also acknowledge the Institute of Theoretical Chemistry at Shandong University for providing the computational resources.

Supporting Information Available: Cartesian coordinates of all structures optimized from QM and QM/MM calculations and a video file showing the thermal fluctuations of the

enzyme–substrate complex during the MD simulations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) 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.
- (2) Himmel, M. E. *Biomass recalcitrance*; Wiley-Blackwell: London, 2008.
- (3) Updegraff, D. M. *Anal. Biochem.* **1969**, *32*, 420–424.
- (4) Wang, L. S.; Zhang, Y. Z.; Gao, P. *J. Sci. China Ser. C—Life Sci.* **2008**, *51*, 620–629.
- (5) Edgar, K. J.; Buchanan, C. M.; Debenham, J. S.; Rundquist, P. A.; Seiler, B. D.; Shelton, M. C.; Tindall, D. *Prog. Polym. Sci.* **2001**, *26*, 1605–1688.
- (6) Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J., Jr.; Hallett, J. P.; Leak, D. J.; Liotta, C. L.; Mielenz, J. R.; Murphy, R.; Templer, R.; Tschaplinski, T. *Science* **2006**, *311*, 484–489.
- (7) Lynd, L. R.; Laser, M. S.; Bransby, D.; Dale, B. E.; Davison, B.; Hamilton, R.; Himmel, M.; Keller, M.; McMillan, J. D.; Sheehan, J.; Wyman, C. E. *Nat. Biotechnol.* **2008**, *26*, 169–172.
- (8) Wolfenden, R.; Yuan, Y. *J. Am. Chem. Soc.* **2008**, *130*, 7548–7549.
- (9) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Soc. Chem.* **2003**, *125*, 14300–14306.
- (10) Teeri, T. T. *Trends Biotechnol.* **1997**, *15*, 160–167.
- (11) Schubert, C. *Nat. Biotechnol.* **2006**, *24*, 777–784.
- (12) André, G.; Kanchanawong, P.; Palma, R.; Cho, H.; Deng, X.; Irwin, D.; Himmel, M. E.; Wilson, D. B.; Brady, J. W. *Protein Eng.* **2003**, *16*, 125–134.
- (13) Kraulis, P. J.; Clore, G. M.; Nilges, M.; Jones, T. A.; Pettersson, G.; Knowles, J.; Gronenborn, A. M. *Biochemistry* **1989**, *28*, 7241–7257.
- (14) Divine, C.; Ståhlberg, J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J. K. C.; Teeri, T. T.; Jones, A. *Science* **1994**, *265*, 524–528.
- (15) Rouvinen, J.; Bergfors, T.; Teeri, T.; Knowles, J. K.; Jones, T. A. *Science* **1990**, *249*, 380–386.
- (16) Barr, B. K.; Wolfgang, D. E.; Piens, K.; Claeysse, M.; Wilson, D. B. *Biochemistry* **1998**, *37*, 9220–9229.
- (17) Divine, C.; Ståhlberg, J.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1998**, *275*, 309–325.
- (18) Mulakala, C.; Reilly, P. J. *Proteins* **2005**, *60*, 598–605.
- (19) Zhao, X.; Rignall, T. R.; McCabe, C.; Adney, W. S.; Himmel, M. E. *Chem. Phys. Lett.* **2008**, *460*, 284–288.
- (20) Koivula, A.; Ruohonen, L.; Wohlfahrt, G.; Reinikainen, T.; Teeri, T. T.; Piens, K.; Claeysse, M.; Weber, M.; Vasella, A.; Becker, D.; Sinnott, M. L.; Zou, J.; Kleywegt, G. J.; Szardenis, M.; Ståhlberg, J.; Jones, T. A. *J. Am. Chem. Soc.* **2002**, *124*, 10015–10024.
- (21) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.
- (22) Bruice, T. C. *Chem. Rev.* **2006**, *106*, 3119–3139.
- (23) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3188–3209.
- (24) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci.* **2005**, *102*, 6679–6685.
- (25) Wu, R.; Xie, X.; Cao, Z.; Mo, Y. *J. Am. Chem. Soc.* **2008**, *130*, 7022–7031.
- (26) Hu, P.; Zhang, Y. *J. Am. Chem. Soc.* **2006**, *128*, 1272–1278.
- (27) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (28) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (29) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Ciosowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision D.01*; Gaussian, Inc.: Wallingford, CT, 2004.
- (30) Lee, C.; Yang, W.; Parr, G. *Phys. Rev.* **1988**, *37*, 785–794.
- (31) Fukui, K. *J. Phys. Chem.* **1970**, *74*, 4161–4163.
- (32) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363–368.
- (33) MacKerell, A. D.; Bashford, D.; Bellott, D.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (34) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580–10594.
- (35) de Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigo, M.; Kramer, G. *J. Phys. Chem. B* **1999**, *103*, 6133–6141.
- (36) de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsson, E.; Sjovoll, M.; Fahmi, A.; Schafer, A. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 1–28.
- (37) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
- (38) Christiansen, O.; Koch, H.; Jorgensen, P. *Chem. Phys. Lett.* **1995**, *243*, 409–418.
- (39) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143–152.
- (40) Smith, W.; Forester, T. R. *J. Mol. Graphics* **1996**, *14*, 136–141.
- (41) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177–2186.
- (42) Knowles, J. K. C.; Lehtovaara, P.; Murray, M.; Sinnott, M. L. *J. Chem. Soc., Chem. Commun.* **1988**, 1401–1402.
- (43) Claeysse, M.; Tomme, P.; Brewer, C. F.; Hehre, E. J. *FEBS Lett.* **1990**, *263*, 89–92.
- (44) Koshland, D. E. *Biol. Rev. Cambridge Philos. Soc.* **1953**, *28*, 416–436.
- (45) Zechel, D. L.; Withers, S. G. *Acc. Chem. Res.* **2000**, *33*, 11.
- (46) McCarter, J. D.; Withers, S. G. *Curr. Opin. Struct. Biol.* **1994**, *4*, 885–892.
- (47) Strynadka, N. C. J.; James, M. N. G. *J. Mol. Biol.* **1991**, *220*, 401–424.
- (48) Kuroki, R.; Weaver, L. H.; Matthews, B. W. *Science* **1993**, *262*, 2030–2033.
- (49) Bottoni, A.; Miscione, G. P.; Vivo, M. D. *Proteins: Struct. Funct. Bioinf.* **2005**, *59*, 118–130.
- (50) Bras, N.; Moura-Tamames, S. A.; Fernandes, P. A.; Ramos, M. J. *J. Comput. Chem.* **2008**, *29*, 2565–2574.