Article

# Characterization of the Dielectric Constant in the *Trichoderma reesei* Cel7B Active Site

Xiangfei Song,[†,§] Yefei Wang,[†,§] Shujun Zhang,[†] Shihai Yan,[‡] Tong Li,[†] and Lishan Yao*,[†]

[†]Laboratory of Biofuels, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong, 266061, China

[‡]College of Chemistry and Pharmaceutical Sciences, Qingdao Agricultural University, Qingdao, Shandong, 266109, China

S *Supporting Information*

ABSTRACT: An attempt is made to evaluate the dielectric constant of the *Trichoderma reesei* Cel7B active site. Through kinetic measurements, the pK_a value of the catalytic acid E201 is determined. Mutations (away from E201) with net charge changes are introduced to perturb the E201 pK_a. It is shown that the mutation with a +1 charge change (including G225R, G230R, and A335R) decreases the pK_a of E201, whereas the mutation with a −1 charge change (including Q149E, A222D, G225D, and G230D) increases the pK_a. This effect is consistent with the electrostatic interaction between the changed charge and the E201 side chain. The fitting of the experimental data yields an apparent dielectric constant of 25−80. Molecular dynamics simulations with explicit water molecules indicate that the high solvent accessibility of the active site contributes largely to the high dielectric constant. ONIOM calculations show that high dielectric constant benefits the catalysis through decreasing the energy of the transition state relative to that of the enzyme substrate complex.
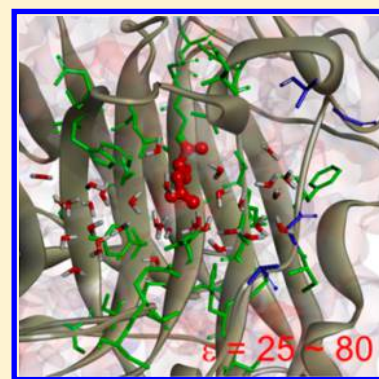
INTRODUCTION

The importance of cellulose hydrolysis in the conversion of biomass to renewable energy sources is well recognized. Cellulases are attractive catalysts for cellulose hydrolysis because they are more environmentally friendly than strong acid catalysts. The filamentous fungus *Trichoderma reesei* (*Hypocrea jecorina*) is of particular interest due to its remarkable ability to secrete large amounts of cellulases which include exoglucanases cleaving cellobiose from cellulose strand ends, endoglucanases cleaving strands randomly, and β-glucosidases converting soluble cellodextrins and cellobiose to glucose. Research is actively pursued to understand the cellulose hydrolysis mechanism (e.g. refs 1−3) by cellulases and improve the enzyme efficiency by optimizing the synergistic effect of different enzyme components[4] or enhancing their thermostability.[5,6] However, the efficiency of cellulases is still a key limiting factor for the biorefinary application.[7] A better understanding of cellulases structure functional relationship is desired to further improve their activity.

T. reesei (Tr.) Cel7B is one of the major endoglucanases and accounts for 5−10% of the total cellulases secreted by the fungus.[8] The enzyme is composed of two domains, the cellulose binding domain (CBD) and the catalytic domain (CD). The X-ray structure of apo Tr. Cel7B CD (371 amino acids) has been determined.[9] It adopts a β-sandwich fold formed by two large antiparallel β-sheets. The active site is located in a cleft which identifies and binds the cellulose chain for hydrolysis. *Tr.* Cel7B is a retaining glucosidase.[10−12] It hydrolyzes glucan chains through a two-step mechanism,

including the glycosylation step and the deglycosylation step. Our previous computational study[13] shows that the glycosylation step involves a proton transfer from the general acid (E201) to the leaving group O_4 and a nucleophile (E196) attack to the anomeric carbon atom C_1 at −1 binding site, and the deglycosylation step involves a proton transfer from a water molecule to E201 and the hydroxide nucleophilic attack to C_1 (Scheme 1). In both steps, a concerted mechanism is employed with an oxocarbenium ion-like transition state.[14] The same mechanism is also found in the cellulose hydrolysis[15,16] by *Tr.* Cel7A, the major exoglucanase which has active site architecture and overall folding[17,18] similar to Cel7B.

As discussed above, in the *Tr.* Cel7B catalyzed glycosidic bond cleavage process, a oxocarbenium ion-like transition state is formed where a sp2 hybridization is adopted by the anomeric C_1 atom and positive charges are accumulated as compared to the substrate which has a sp3 hybridized C_1 (Scheme 1). The electron density distribution differences between the two states indicate that electrostatics could have a large effect on the transition state stabilization. The amplitude of the electrostatic interaction depends on the dielectric constant. Considering that the *Tr.* Cel7B active site is located on top of the substrate binding cleft which is water accessible, the solvent may have a large contribution on the electrostatic screening and thus the charge screening heterogeneity is expected in the region.

Electrostatic interactions play a critical role in enzyme catalysis.[19] Fluorescence methods have been used to probe the

Scheme 1. Cel7B Catalyzed PNPL Glycosidic Bond Cleavage Reaction Mechanism
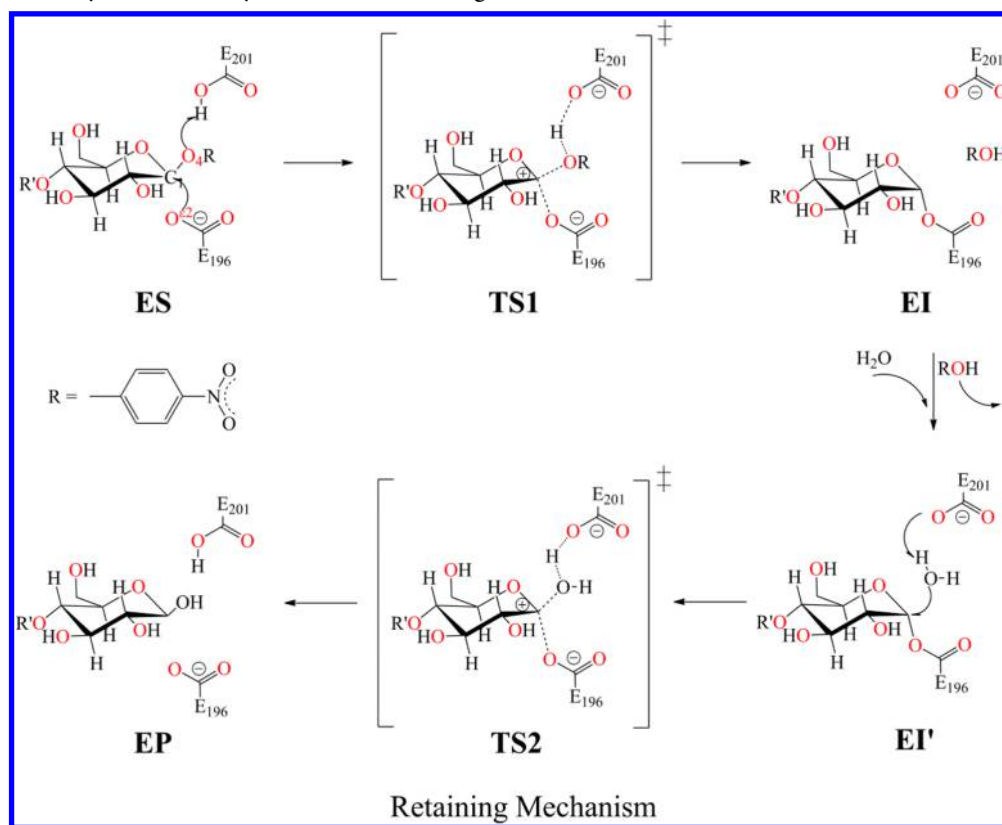
protein dielectric constant and values between 6 and ~50 were obtained.[20] An effective ε value of 8−30 was obtained by pK_a measurements in the Staphylococcal nuclease.[21,22] Computational studies using Kirkwood−Fröhlich formalism[23] reported ε values of 2−25 where the protein interior tends to have much smaller values than the outer part.[24−26] The large variation of ε is due to local composition and structural differences of a protein which is a mixture of polar, nonpolar, and charged amino acids surrounded by water and ions.

In this work, the dielectric constant in the active site of *Tr.* Cel7B is characterized through the pK_a measurement of the catalytic acid E201. Several distal charged mutations (~10 Å from E201) are introduced to perturb the pK_a of E201, through which an effective dielectric constant is derived. It shows that the dielectric constant in the active site is quite large. ONIOM calculations[27] demonstrate that the high dielectric constant medium facilitates the catalysis by lowering the energies of the transition state and various intermediates relative to that of the enzyme substrate complex.

■ METHODS AND MATERIALS

MD Simulations. The apo form *Tr.* Cel7B crystal structure (1EG1) reported by Jones[9] was used as the starting structure with the protonation state determined by the pK_a analysis where the residues with predicted pK_a values less than (larger than or equal to) 5.0 were deprotonated (protonated).[28−31] Specifically, D113, E117, E141, D198, E201, D220, D279, D288, D331, and all histidine/lysine/arginine residues were protonated, with a protein net charge of −2. No substrate was included in the system. Two Na^+ cations were added to neutralize the system. The simulations were carried out with the GROMACS[32] molecular dynamics package. The Amber

force field of FF99SB[33] was employed to model the enzyme. The enzyme was immersed in a cubic box of explicit TIP3P[34] water molecules, with a 12.5 Å distance between the solvent box wall and the nearest solute atoms. There were a total of ~50, 000 atoms in the system. The mutants Q149E, A222D, G225D, G225R, G230R, G230D, and A335R were built in the same way. All simulations were run in NPT ensemble (1 atm, 300 K). The pressure was regulated using the extended ensemble Parrinello−Rahman approach,[35,36] and the temperature was controlled by a modified Berendson thermostat.[37] The Particle-Mesh-Ewald Method with the tinfoil boundary condition[38,39] was used to evaluate the contributions of the long-range electrostatic interactions. A nonbonded pair list cutoff of 12.0 Å was used, and the nonbonded pair list was updated every 5 steps. All bonds to hydrogen atoms in the protein were constrained by using the LINC algorithm,[40] whereas bonds and angles of water molecules were constrained by the SETTLE algorithm,[41] and a time step of 2 fs was used. For the WT and mutants, a 10 ns MD simulation was performed with the first 2 ns simulation treated as the equilibration. The last 8 ns simulation was used for the dielectric constant or the postprocessing electrostatics calculation. The snapshots were saved every 20 ps.

Calculations of the Dielectric Constant. The dielectric constant was calculated using the following equation[26,42]

$$\varepsilon = 1 + \frac{4\pi \langle \mathbf{M}^2 \rangle}{3V k_B T} \tag{1}$$

where $\langle \mathbf{M}^2 \rangle$ is the mean fluctuation of the dipole moment $\mathbf{M}$, V is the volume, k_B is the Boltzmann constant, and T is the absolute temperature. Since we were interested in the dielectric constant of the active site, only the residues and solvent within

10 Å from the $O_{\varepsilon 1}$ atom (the side chain oxygen with a proton attached) of E201 were considered in the calculation. The residue selection was based on the X-ray structure, so that the number of residues in the dielectric constant calculation remained unchanged though during MD simulations the residue coordinates fluctuated. The water molecules were mobile and could move in and out of the defined sphere, and only those within the sphere were considered in the calculation. The system included 35 residues and approximately 40 water molecules (Na^+ ions were away from the active site during the simulation and thus not included in the system). The MD snapshots were fitted using backbone C_α atoms to remove the overall rotation and translation, and then $\langle \mathbf{M}^2 \rangle$ was calculated with an origin at the $O_{\varepsilon 1}$ atom of E201 and averaged over the last 8 ns. The running average of $\langle \mathbf{M}^2 \rangle$ is shown in Figure S1, suggesting that the average dipole moment fluctuation converges after ~5 ns. The volume of the system was calculated for 8 snapshots (corresponding to the structures at the time point of 3, 4, 5, 6, 7, 8, 9, 10 ns) by using the program CHIMERA[43] and averaged to give a mean volume. The average volume of the selected protein residues or the total system was 3956 ± 12 or 5123 ± 130 Å^3. The dielectric constant was calculated for protein and water separately and both together.

Poisson−Boltzmann Electrostatics Calculations. The MM/PBSA module in the Amber 11 package[44] was employed to calculate the total energy of seven mutants Q149E, A222D, G225D, G225R, G230R, G230D, and A335R. The structures were from 400 snapshots of the last 8 ns MD simulation with the explicit solvent. In the MM/PBSA calculations, the water molecules and ions were removed. All atomic charges and radii were taken from the FF99SB.[33] The molecular surface was constructed using a probe sphere with a radius of 1.4 Å. The dielectric constant was varied from 5 to 80 (with an interval of 5) for the enzyme interior and the solvent, with an ionic strength of 4 mM (same as in the experiment). The nonpolar desolvation free energy was determined by the solvent accessible surface area (SASA) with a slope of 0.00542 kcal/mol·Å^2 and an intercept of 0.92 kcal/mol.[45]

ONIOM Calculations. A model system, including 36 residues in the Cel7B active site extracted from the X-ray structure, was treated by a two-layer ONIOM (B3PW91/6-31+G(d,p):AM1)[46−49] method. The high level system included the substrate, the catalytic acid E201, and the nucleophile E196 and E198. The details of the system were described previously.[13] ONIOM calculations were performed to study the dielectric constant effect on the catalysis by re-evaluating the energies of various reaction states determined previously without reoptimization.[13] The solvent effect was described with the self-consistent reaction field (SCRF) method employing the polarizable continuum model (PCM).[50] All the calculations were performed using Gaussian 09.[51] To probe the solvation effect on the catalysis, different dielectric constants were used, including $\varepsilon = 1, 2, 4, 9, 17, 25, 36$, and 78.5.

Kinetic Measurements. The Cel7B (CD only) wild type and mutants including Q149E, A222D, G225D, G225R, G230R, G230D, and A335R were expressed and purified as described previously.[13] The enzyme catalyzed p-nitrophenyl lactoside (PNPL) kinetic rates were measured as follows. A 100 μL mixture containing 1.7 mM PNPL and 2 μM WT Cel7B were incubated at 40 Celsius at various pH buffer conditions. The buffer contained an equal amount (1.33 mM) mixture of glycine, citric acid, and Na_2HPO_4 and had a pH range of 2.4−6.6. Aliquots of 10 μL were taken at time points of 0, 8, 16, 24,

and 32 min and transferred immediately to a microplate containing 190 μL 1-M Na_2CO_3 to stop the reaction. The amount of p-nitrophenol liberated was determined by measuring the absorbance at 405 nm using a microplate reader (Ultrospec Visible plate reader II 96, GE Healthcare Bio-Science). The slope of the product release versus time is the reaction rate v. Since the K_m value is ~14 mM,[13] much larger than the substrate concentration in the reaction mixture, this reaction rate can be converted to k_{cat}/K_m by $v = k_{cat}/K_m[E][S]$, where $[E]$ and $[S]$ are the enzyme and substrate concentrations, respectively. The k_{cat}/K_m values at different pHs were fitted to a modified equation[52]

$$\left(\frac{k_{cat}}{K_m} \right)_{pH} = \left(\frac{k_{cat}}{K_m} \right)_{max} \left(\frac{1}{1 + \left(\frac{10^{-pH}}{10^{-pK_{a1}}} \right)^n + \left(\frac{10^{-pK_{a2}}}{10^{-pH}} \right)^n} \right) \quad (2)$$

where n is the Hill coefficient. The fitted pK_{a1} and pK_{a2} are the apparent pK_a values of the general acid E201 and nucleophile E196 in the apo form, respectively.[52] The equation was fitted by using an in-house script.

■ RESULTS AND DISCUSSION

Experimental pK_a Shifts. pK_a has been widely used to study electrostatic interactions in proteins. However, extracting electrostatics from pK_a is not straightforward. The pK_a of a residue is determined by two factors: the solvation effect and the long-range charge−charge interactions.[53,54] The solvation effect corresponds to the solvation free energy difference of the charged residue in the protein (without electrostatic interactions with other residues) and water, which gives the intrinsic pK_a. To exclude the solvation effect, pK_a shift, which is defined as the pK_a change of the probe residue due to a charge state change of a residue away from the probe, can be used.[55,56] Experimentally, NMR is routinely used to accurately determine the protein residue pK_a values. But for Cel7B, it is extremely difficult to do such measurements due to the precipitation of the sample (at >10 μM) and its low expression (~1 mg enzyme per liter Lysogeny broth medium). Fortunately, apparent pK_a values of the general acid and nucleophile can also be measured by monitoring the k_{cat}/K_m ratio at different pHs, and they correlate quite well with the numbers extracted from the NMR chemical shift titration curve, as demonstrated in *Bacillus circulans* endo-β-(1,4)-xylanase which catalyzes the hydrolysis of xylose through the double-displacement mechanism,[52] similar as in *Tr.* Cel7B. The fitting of the k_{cat}/K_m vs pH for the wild type *Tr.* Cel7B yields the pK_a of 2.77 ± 0.02 for E196 (pK_{a1}) and 5.58 ± 0.02 for E201 (pK_{a2}) which correspond to the pK_as in the apo form (Figure 1, Figure S2, and Table 1).

It is known that the charge−charge interaction is proportional to the inverse of the distance. The mutation site closer to the catalytic acid and the nucleophile is expected to have a larger effect on the pK_a shifts of the two catalytic residues, but the close proximity very likely perturbs the active site geometry and thus complicates the data interpretation.[13] To minimize the mutational effect on catalysis, we aligned nine glycoside hydrolases (GHs) sequences from GH family 7 which have a similar structural fold and active site geometry, including Cel7A from *Trichoderma harzianum*,[57] *T. reesei*,[58] *Rasamsonia emersonii*,[59] and *Heterobasidion annosum*,[60] Cel7B from *Melanocarpus albomyces*,[61] *T. reesei*,[9] *Fusarium oxysporum*,[62] and *Humicola insolens*,[63] and Cel7D from *Phanerochaete*
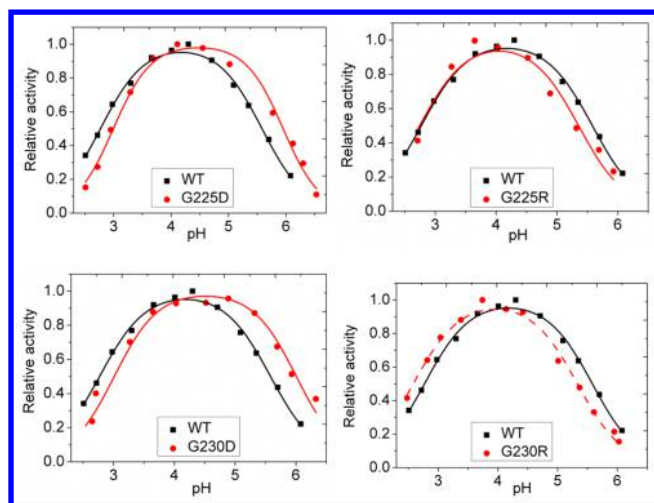
Figure 1. pH dependence of k_{cat}/K_m for *T. reesei* Cel7B WT and mutants G225D, G225R, G230D, and G230R at 40 °C toward the substrate PNPL. The curves were fitted to eq 2, yielding pK_{a1} and pK_{a2} values (Table 1). The difference of the pK_a values between the WT and the mutant is due to the electrostatic interactions between D/R225 (D/R230) and the general acid E201 or the nucleophile E196.
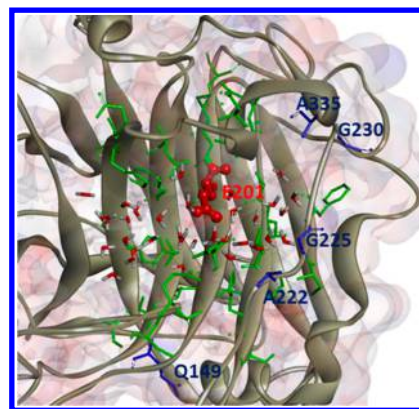

Figure 2. T. *reesei* Cel7B active site. E201 is the general acid (in red). The residues, drawn in green, and water molecules within 10 Å from the E201 $O_{\varepsilon2}$ atom (protonated side chain carboxylate oxygen) are used in the dielectric constant calculation which yields ε value of 4.0 for the protein, 43.8 for the water molecules, and 9.5 for the overall system. Five mutated residues Q149, A222, G225, G230, and A335 are displayed and labeled in blue.

chrysosporium.[64] On the basis of the sequence homology, five conservative charged mutations, including Q149E, A222D, G225D, G230R, and A335R (Figure 2), were selected to investigate the charge–charge interaction. The distances between E201 and the mutated sites are about 10 Å or larger. These mutations, which occur to at least one homologous structure listed above, are expected to have minor effects on the active site structure. Thus, the pK_a shift corresponds to the electrostatic interaction rather than the structural perturbation. The pK_a values listed in Table 1 correspond to those of E196 (pK_{a1}) and E201 (pK_{a2}), respectively (Figure 1, Figure S2). The mutational effect on the maximum k_{cat}/K_m is small except for G225D, suggesting the perturbation on the active site structure is small for most mutations. It appears that the pK_a shift effect on pK_{a2} (from E201, the general acid) is larger than that on pK_{a1}. By investigating the active site hydrogen bonding in the MD snapshots, we observe that the carboxyl of D198 (protonated) forms a strong hydrogen bond (as a donor) with the side chain of E196 (O–O distance of 2.6 ± 0.1 Å and angle ∠O–H–O of 168.2 ± 6.2°). We speculate that the strong hydrogen bond may compensate the pK_a change of E196 (pK_{a1}) through minor adjustment. Meanwhile, at the pH near pK_{a1} (2.5–3.0), most D and E in *Tr.* Cel7B should be protonated and the effect on pK_{a1} can be complex. Thus, we

focus on pK_{a2} which is about 4.8–6.2. ΔpK_a, defined by pK_{a2} (mutant) − pK_{a2} (WT), is also listed in Table 1. As can be seen, mutations to D and E generally have a positive ΔpK_a (increase), whereas mutations to R have a negative ΔpK_a (decrease). This effect is consistent with the direct electrostatic interaction model, meaning that the mutation to D/E shifts the E201 protonation equilibrium toward the protonated state, whereas the mutation to R shifts the equilibrium to the deprotonated state. To further confirm this effect, two mutations G225R and G230D were also made and both showed a pK_a shift opposite to that of G225D and G230R (Table 1, Figure 1). In the following, we attempt to quantify the pK_{a2} shift with the electrostatic model.

Predicting the Apparent Dielectric Constant from the Experimental pK_a Shifts. The pK_a shift was predicted as follows. First, 400 MD snapshots with solvent and counterions excluded were extracted from the last 8 ns simulation of a mutant (taking G230R as an example). Second, for each snapshot, four energy calculations (Scheme 2) were performed using the MM/PBSA method, yielding the following: E_1, the energy of the system (sum of the MM energy and the PBSA solvation free energy); E_2, the energy of the same system but with the charge distribution of E201 changed to that of a deprotonated glutamic acid; E_3, the energy of the same system with the charge distribution of R230 changed to that of a glycine (The side chain atoms of R230 have a charge of zero and van der Waals parameters unchanged.); E_4, the energy of

Table 1. Apparent pK_a Values of the Catalytic Acid and Nucleophile of the WT and Mutants

	$pK_{a1}{}^a$	$pK_{a2}{}^a$	n^a	$(k_{cat}/K_m)_{max}$	ΔpK_a (exp[b])	ΔpK_a (cal[c])	ε
WT	2.77 ± 0.02	5.58 ± 0.02	1.13 ± 0.05	91 ± 4	0	n/a	n/a
Q149E	2.93 ± 0.02	6.14 ± 0.03	1.09 ± 0.05	112 ± 7	0.56	0.57 ± 0.01[d]	25
A222D	2.75 ± 0.03	5.99 ± 0.05	0.94 ± 0.08	81 ± 2	0.41	0.40 ± 0.01	50
G225R	2.74 ± 0.10	5.36 ± 0.07	1.12 ± 0.15	8.2 ± 0.5	−0.22	−0.42 ± 0.03	80
G225D	2.99 ± 0.03	5.95 ± 0.03	1.33 ± 0.09	13 ± 1	0.37	0.32 ± 0.04	80
G230R	2.58 ± 0.03	5.33 ± 0.04	1.11 ± 0.07	110 ± 8	−0.25	−0.28 ± 0.03	80
G230D	3.01 ± 0.06	6.01 ± 0.06	1.21 ± 0.14	60 ± 4	0.43	0.44 ± 0.01	40
A335R	2.70 ± 0.05	4.89 ± 0.05	1.22 ± 0.10	77 ± 4	−0.69	−0.71 ± 0.06	30

[a]The errors were estimated based on duplicate measurements. [b]ΔpK_a(exp) = pK_{a2}(mut) − pK_{a2}(WT). [c]ΔpK_a(cal) was from the data fitting. [d]The error was from the standard deviation of the MD block (2 ns each) averages.

Scheme 2. Calculation of $\Delta\Delta G = \Delta G_2 - \Delta G_1 = -1.36\Delta pK_a{}^a$
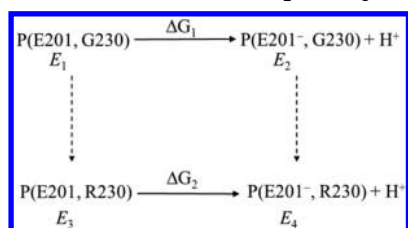
$^a\Delta pK_a$ is pK_a difference of E201 between the mutant R230 and the WT, $pK_a(R230) - pK_a(WT)$. $\Delta\Delta G$ is approximated by $(E_1 - E_2) - (E_3 - E_4)$.

the same system with the charge distribution of E201 and R230 changed to that of a deprotonated glutamic acid and a glycine, respectively. Third, the $\Delta\Delta G \approx (E_1 - E_2) - (E_3 - E_4)$ was calculated and averaged over 400 MD snapshots (Table S1). The averaged $\Delta\Delta G$ is equal to $-1.36\Delta pK_a$.[55] We assume a uniform apparent dielectric constant for the protein and the solvent, and by adjusting such a parameter (from 5 to 80 with an interval of 5) the best match between the experimental and predicted ΔpK_a can be achieved (We did not choose the two dielectric constant models (e.g., $\varepsilon = 80$ for the solvent and a smaller value for the enzyme) because the calculated ΔpK_a was sensitive to the boundary between the enzyme and the water causing ambiguity in the fitted dielectric constant.). The apparent dielectric constant is listed for different mutations (Table 1). As can be seen, overall the dielectric constant is quite large, and it varies from 25 (Q149E) to 80 (G225R, G225D, and G230R). The large dielectric constant value is likely due to the active site solvent accessibility (see the Discussion section). No correlation is observed between the position of the mutational site and the value of the dielectric constant. It is worth noting that the MM/PBSA calculation in this work provides a very rough estimation of the electrostatics. The solvent distribution change and the enzyme structure reorganization (and polarization) due to the mutation (e.g., G230R) are not treated explicitly. Instead, the dielectric constant is used to account for such effects (as in a continuum model).

Active Site Dielectric Constant from the Kirkwood–Fröhlich Formalism. 35 residues and about 40 water molecules within 10 Å from E201 $O_{\varepsilon 1}$, which donates its proton to O4 of the substrate in the first step of catalysis (Scheme 1), are included in the dielectric constant calculation. The overall dielectric constant is 9.5 (Table 2), close to that of

Table 2. Dielectric Constant ε of the Cel7B Active Site in the Apo Form Calculated from MD Simulations

	active site^a (protein)	water^b	together^c
apo form	4.0	43.8	9.5

^{a}The active site includes 35 residues surrounding the general acid E201. ^{b}Water molecules within 10 Å from E201 $O_{\varepsilon 2}$ are included. ^{c}All atoms in ^{a} and ^{b} are combined together.

trypsin active site which is also located on a cleft,[65] but smaller than that fitted from the mutation. The predicted dielectric constant is 4.0 for the active site residues and 43.8 for the water in the cleft (Table 2). The dielectric constant of the cleft water molecules is lower than that of the TIP3P bulk water ($\varepsilon = 94$),[66] suggesting that amino acid residues encompassing the water slightly decrease its mobility. Compared to water, the

protein has a much smaller ε, so that the charge–charge interaction between a remote fragment (e.g., a charged residue) and the active site atoms is screened less (more) if it goes through the protein (water). The higher dielectric constant of the cleft water molecules indicates that their total dipole moment has a much larger fluctuation $\langle \mathbf{M}^2 \rangle$ per unit volume than that of the active site residues. In fact, inclusion of water in the $\langle \mathbf{M}^2 \rangle$ calculation increases its value for the system by 2-fold (Figure S1). This is very different from a previous work by Pitera et al.[24] where including bound water only increases $\langle \mathbf{M}^2 \rangle$ slightly for the apo rat fatty acid binding protein (FABP). The difference is most likely due to the higher water mobility in the Cel7B active site than that in the cavity of FABP. It is expected that if the radius of the active site (10 Å sphere around $O_{\varepsilon 1}$ of E201) becomes larger, the predicted dielectric constant will be larger as well due to the inclusion of more water molecules. Compared to the values in Table 1, the predicted dielectric constant of the active site ($\varepsilon = 9.5$) from the Kirkwood–Fröhlich formalism is 3–9 times smaller. As discussed above, the dielectric constant fitted from ΔpK_as corresponds to the screening of the charge–charge interaction (e.g., R230-E201 in the G230R mutant) which has contributions from the solvation change, the reorganization of the enzyme residues, and the polarization effect due to the mutation (G230 to R230). If all these effects are computed explicitly, a dielectric constant of 1 should be used to predict ΔpK_as. In contrast, the dielectric constant from the Kirkwood–Fröhlich formalism is derived from the statistical mechanics and has a unique value when the system is defined. In this sense, the dielectric constants from the two different calculations are not strictly comparable.

Dielectric Constant Effect on Catalysis. As discussed above, the electron density distribution is quite different between the substrate and transition state, implying that the dielectric constant can impact the catalytic barrier. To investigate this effect, ONIOM calculations, with different solvents corresponding to the dielectric constant ranged from 1 to 78.5, were performed for all the states, from the enzyme substrate (ES) to the enzyme product (EP) complex. The energies of all the states (except EP) relative to the ES decrease as the dielectric constant of the solvent increases. The largest change is from EI which decreases from 0 kcal/mol in the vacuum to -6.8 kcal/mol in water. The reaction barrier, corresponding to the energy difference between the transition state (TS1) and the ES, deceases by 3.1 kcal/mol in water compared to that in vacuum, with the largest change occurring as the dielectric constant increases from 1 to 10 (Figure 3). ONIOM calculations demonstrate that the high dielectric constant environment facilitates the catalysis. This also suggests that potentially one can improve the catalysis by tuning the dielectric constant (e.g., through changing the ionic strength), but considering that the dielectric constant of the enzyme active site is already quite high ($\varepsilon = 25-80$, Table 1), adding more salts to the reaction buffer might only have a very small effect on the catalysis based on the ONIOM prediction.

■ CONCLUSION

The dielectric constant of the Cel7B active site is studied through the pK_a of E201. Distal mutations are introduced to alter the charge state of the mutated site. The mutation increases (decreases) the pK_a of E201 if the charge of the mutated site decreases (increases). This observation suggests that the pK_a changes are due to the electrostatic interaction between E201 and the mutated side chain charge change. An
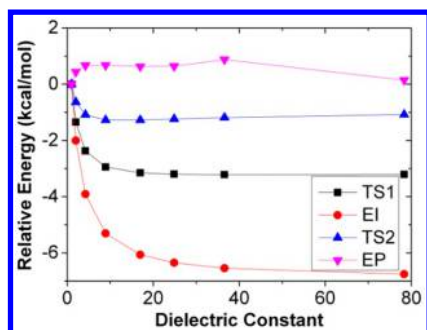
Figure 3. Relative energy changes of various states as a function of the dielectric constant of the solvent. In the gas phase, the relative energies (with the ES energy set at 0 kcal/mol) are 19.0 for TS1, 0.0 for EI, 10.5 for TS2, and −13.9 for EP. It appears that higher dielectric constant reduces the relative energy of EI, TS1, and TS2 but slightly increases that of EP.

effective dielectric constant of 25−80 is obtained from the fitting. The high dielectric constant of the active site facilitates the catalysis through reducing the catalytic barrier.

ASSOCIATED CONTENT

Ⓢ Supporting Information

One figure of $\langle \mathbf{M}^2 \rangle$ versus number of MD snapshots (Figure S1); one figure of relative activity versus pH values (Figure S2); one table of different energy terms of the mutants (Table S1). The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00155.

AUTHOR INFORMATION

Corresponding Author

*Phone 86 532 80662792. Fax: 86 532 80662778. E-mail: yaols@qibebt.ac.cn. Corresponding author address: Room 413 QIBEBT, Qingdao, 266061, China.

Author Contributions

§These two authors contributed equally to the work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Igarashi, K.; Uchihashi, T.; Koivula, A.; Wada, M.; Kimura, S.; Okamoto, T.; Penttila, M.; Ando, T.; Samejima, M. Traffic Jams Reduce Hydrolytic Efficiency of Cellulase on Cellulose Surface. *Science* **2011**, *333*, 1279−1282.

(2) Cruys-Bagger, N.; Elmerdahl, J.; Praestgaard, E.; Tatsumi, H.; Spodsberg, N.; Borch, K.; Westh, P. Pre-Steady-State Kinetics for Hydrolysis of Insoluble Cellulose by Cellobiohydrolase Cel7A. *J. Biol. Chem.* **2012**, *287*, 18451−18458.

(3) Payne, C. M.; Bomble, Y.; Taylor, C. B.; McCabe, C.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. Multiple Functions of Aromatic-Carbohydrate Interactions in a Processive Cellulase Examined with Molecular Simulation. *J. Biol. Chem.* **2011**, *286*, 41028−41035.

(4) Lantz, S. E.; Goedegebuur, F.; Hommes, R.; Kaper, T.; Kelemen, B. R.; Mitchinson, C.; Wallace, L.; Stahlberg, J.; Larenas, E. A. *Hypocrea jecorina* CEL6A Protein Engineering. *Biotechnol. Biofuels* **2010**, *3*, 20.

(5) Heinzelman, P.; Komor, R.; Kanaan, A.; Romero, P.; Yu, X. L.; Mohler, S.; Snow, C.; Arnold, F. Efficient Screening of Fungal Cellobiohydrolase Class I Enzymes for Thermostabilizing Sequence Blocks by SCHEMA Structure-Guided Recombination. *Protein Eng. Des. Sel.* **2010**, *23*, 871−880.

(6) Heinzelman, P.; Snow, C. D.; Wu, I.; Nguyen, C.; Villalobos, A.; Govindarajan, S.; Minshull, J.; Arnold, F. H. A Family of Thermostable Fungal Cellulases Created by Structure-Guided Recombination. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 5610−5615.

(7) Klein-Marcuschamer, D.; Oleskowicz-Popiel, P.; Simmons, B. A.; Blanch, H. W. The Challenge of Enzyme Cost in the Production of Lignocellulosic Biofuels. *Biotechnol. Bioeng.* **2012**, *109*, 1083−1087.

(8) Kolbe, J.; Kubicek, C. P. Quantification and Identification of the Main Components of the *Trichoderma* Cellulase Complex with Monoclonal Antibodies Using an Enzyme-Linked Immunosorbent Assay (ELISA). *Appl. Microbiol. Biotechnol.* **1990**, *34*, 26−30.

(9) Kleywegt, G. J.; Zou, J. Y.; Divne, C.; Davies, G. J.; Sinning, I.; Stahlberg, J.; Reinikainen, T.; Srisodsuk, M.; Teeri, T. T.; Jones, T. A. The Crystal Structure of the Catalytic Core Domain of Endoglucanase I from *Trichoderma reesei* at 3.6 Å Resolution, and a Comparison with Related Enzymes. *J. Mol. Biol.* **1997**, *272*, 383−397.

(10) Claeyssens, M.; Tomme, P.; Brewer, C. F.; Hehre, E. J. Stereochemical Course of Hydrolysis and Hydration Reactions Catalysed by Cellobiohydrolases I and II from *Trichoderma reesei*. *FEBS Lett.* **1990**, *263*, 89−92.

(11) Knowles, J. K. C.; Lentovaara, P.; Murray, M.; Sinnott, M. L. Stereochemical Course of the Action of the Cellobioside Hydrolases I and II of *Trichoderma reesei*. *J. Chem. Soc., Chem. Commun.* **1988**, 1401−1402.

(12) Claeyssens, M.; Henrissat, B. Specificity Mapping of Cellulolytic Enzymes: Classification into Families of Structurally Related Proteins Confirmed by Biochemical Analysis. *Protein Sci.* **1992**, *1*, 1293−1297.

(13) Zhang, Y.; Yan, S. H.; Yao, L. S. A Mechanistic Study of *Trichoderma reesei* Cel7B Catalyzed Glycosidic Bond Cleavage. *J. Phys. Chem. B* **2013**, *117*, 8714−8722.

(14) Vocadlo, D. J.; Davies, G. J. Mechanistic Insights into Glycosidase Chemistry. *Curr. Opin. Chem. Biol.* **2008**, *12*, 539−555.

(15) Li, J. H.; Du, L. K.; Wang, L. S. Glycosidic-Bond Hydrolysis Mechanism Catalyzed by Cellulase Cel7A from *Trichoderma reesei*: A Comprehensive Theoretical Study by Performing MD, QM, and QM/MM Calculations. *J. Phys. Chem. B* **2010**, *114*, 15261−15268.

(16) Yan, S. H.; Li, T.; Yao, L. S. Mutational Effects on the Catalytic Mechanism of Cellobiohydrolase I from *Trichoderma reesei*. *J. Phys. Chem. B* **2011**, *115*, 4982−4989.

(17) Divne, C.; Stahlberg, J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J. K. C.; Teeri, T. T.; Jones, T. A. The Three-Dimensional Crystal Structure of the Catalytic Core of Cellobiohydrolase I from *Trichoderma reesei*. *Science* **1994**, *265*, 524−528.

(18) Divne, C.; Stahlberg, J.; Teeri, T. T.; Jones, T. A. High-resolution Crystal Structures Reveal How a Cellulose Chain is Bound in the 50 Å Long Tunnel of Cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* **1998**, *275*, 309−325.

(19) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106*, 3210−3235.

(20) Cohen, B. E.; McAnaney, T. B.; Park, E. S.; Jan, Y. N.; Boxer, S. G.; Jan, L. Y. Probing Protein Electrostatics with a Synthetic Fluorescent Amino Acid. *Science* **2002**, *296*, 1700−1703.

(21) Isom, D. G.; Castaneda, C. A.; Cannon, B. R.; Garcia-Moreno, B. E. Large Shifts in pKa Values of Lysine Residues Buried Inside a Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 5260−5265.

(22) Isom, D. G.; Castaneda, C. A.; Velu, P. D.; Garcia-Moreno, B. Charges in the Hydrophobic Interior of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16096−16100.

(23) Fröhlich, H. *Theory of Dielectrics*; Oxford University Press: London, U.K., 1958.

(24) Pitera, J. W.; Falta, M.; van Gunsteren, W. F. Dielectric Properties of Proteins from Simulation: The Effects of Solvent, Ligands, pH, and Temperature. *Biophys. J.* 2001, 80, 2546−2555.

(25) Simonson, T.; Brooks, C. L. Charge Screening and the Dielectric Constant of Proteins: Insights from Molecular Dynamics. *J. Am. Chem. Soc.* 1996, 118, 8452−8458.

(26) Patargias, G. N.; Harris, S. A.; Harding, J. H. A Demonstration of the Inhomogeneity of the Local Dielectric Response of Proteins by Molecular Dynamics Simulations. *J. Chem. Phys.* 2010, 132, 235103.

(27) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. The ONIOM Method: Its Foundation and Applications to Metalloenzymes and Photobiology. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2012, 2, 327−350.

(28) Li, H.; Robertson, A. D.; Jensen, J. H. Very Fast Empirical Prediction and Rationalization of Protein pK_a Values. *Proteins* 2005, 61, 704−721.

(29) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very Fast Prediction and Rationalization of pK_a Values for Protein−Ligand Complexes. *Proteins* 2008, 73, 765−783.

(30) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *J. Chem. Theory Comput.* 2011, 7, 525−537.

(31) Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK_a Values. *J. Chem. Theory Comput.* 2011, 7, 2284−2295.

(32) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* 2008, 4, 435−447.

(33) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* 2006, 65, 712−725.

(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 1983, 79, 926.

(35) Nose, S.; Klein, M. L. Constant Pressure Molecular-Dynamics for Molecular-Systems. *Mol. Phys.* 1983, 50, 1055−1076.

(36) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals - A New Molecular-Dynamics Method. *J. Appl. Phys.* 1981, 52, 7182−7190.

(37) Berendsen, H. J. C. Transport-Properties Computed by Linear Response through Weak-Coupling to a Bath. *NATO ASI Ser., Ser. E* 1991, 205, 139−155.

(38) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* 1993, 98, 10089.

(39) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* 1995, 103, 8577.

(40) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* 1997, 18, 1463−1472.

(41) Miyamoto, S.; Kollman, P. A. Settle - An Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* 1992, 13, 952−962.

(42) Neumann, M. Dipole Moment Fluctuation Formulas in Computer Simulations of Polar Systems. *Mol. Phys.* 1983, 50, 841−858.

(43) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 2004, 25, 1605−1612.

(44) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.;
Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecker, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *Amber 11*; 2011.

(45) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Computational Alanine Scanning Mutagenesis - An Improved Methodological Approach. *J. Comput. Chem.* 2007, 28, 644−654.

(46) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. Atoms, Molecules, Solids, and Surfaces - Applications of the Generalized Gradient Approximation for Exchange and Correlation. *Phys. Rev. B* 1992, 46, 6671−6687.

(47) Perdew, J. P.; Wang, Y. Accurate and Simple Analytic Representation of the Electron-Gas Correlation-Energy. *Phys. Rev.B* 1992, 45, 13244−13249.

(48) Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.* 1993, 98, 1372.

(49) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* 1985, 107, 3902−3909.

(50) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* 2005, 105, 2999−3094.

(51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E., Jr.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian, Gaussian 09*, Revision B.01; Gaussian, Inc.: Wallingford, CT, 2010.

(52) Joshi, M. D.; Sidhu, G.; Nielsen, J. E.; Brayer, G. D.; Withers, S. G.; McIntosh, L. P. Dissecting the Electrostatic Interactions and pH-Dependent Activity of a Family 11 Glycosidase. *Biochemistry* 2001, 40, 10115−10139.

(53) Schutz, C. N.; Warshel, A. What Are the Dielectric "Constants" of Proteins and How to Validate Electrostatic Models? *Proteins* 2001, 44, 400−417.

(54) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Modeling Electrostatic Effects in Proteins. *Biochim. Biophys. Acta, Proteins Proteomics* 2006, 1764, 1647−1676.

(55) Lee, K. K.; Fitch, C. A.; Garcia-Moreno, B. Distance Dependence and Salt Sensitivity of Pairwise, Coulombic Interactions in a Protein. *Protein Sci.* 2002, 11, 1004−1016.

(56) Baran, K. L.; Chimenti, M. S.; Schlessman, J. L.; Fitch, C. A.; Herbst, K. J.; Garcia-Moreno, B. E. Electrostatic Effects in a Network of Polar and Ionizable Groups in Staphylococcal Nuclease. *J. Mol. Biol.* 2008, 379, 1045−1062.

(57) Textor, L. C.; Colussi, F.; Silveira, R. L.; Serpa, V.; de Mello, B. L.; Muniz, J. R. C.; Squina, F. M.; Pereira, N.; Skaf, M. S., Jr.; Polikarpov, I. Joint X-ray Crystallographic and Molecular Dynamics Study of Cellobiohydrolase I from *Trichoderma harzianum*: Deciphering the Structural Features of Cellobiohydrolase Catalytic Activity. *FEBS J.* 2013, 280, 56−69.

(58) Stahlberg, J.; Divne, C.; Koivula, A.; Piens, K.; Claeyssens, M.; Teeri, T. T.; Jones, T. A. Activity Studies and Crystal Structures of Catalytically Deficient Mutants of Cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* 1996, 264, 337−349.

(59) Grassick, A.; Murray, P. G.; Thompson, R.; Collins, C. M.; Byrnes, L.; Birrane, G.; Higgins, T. M.; Tuohy, M. G. Three-

Dimensional Structure of a Thermostable Native Cellobiohydrolase, CBH IB, and Molecular Characterization of the *cel*7 Gene from the Filamentous Fungus, *Talaromyces emersonii. Eur. J. Biochem.* **2004**, *271*, 4495−4506.

(60) Momeni, M. H.; Payne, C. M.; Hansson, H.; Mikkelsen, N. E.; Svedberg, J.; Engstrom, A.; Sandgren, M.; Beckham, G. T.; Stahlberg, J. Structural, Biochemical, and Computational Characterization of the Glycoside Hydrolase Family 7 Cellobiohydrolase of the Tree-Killing Fungus *Heterobasidion irregular. J. Biol. Chem.* **2013**, *288*, 5861−5872.

(61) Parkkinen, T.; Koivula, A.; Vehmaanpera, J.; Rouvinen, J. Crystal Structures of *Melanocarpus albomyces* Cellobiohydrolase Cel7B in Complex with Cello-Oligomers Show High Flexibility in the Substrate Binding. *Protein Sci.* **2008**, *17*, 1383−1394.

(62) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schulein, M.; Davies, G. J. Structure of the *Fusarium oxysporum* Endoglucanase I with a Nonhydrolyzable Substrate Analogue: Substrate Distortion Gives Rise to the Preferred Axial Orientation for the Leaving Group. *Biochemistry* **1996**, *35*, 15280−15287.

(63) Davies, G. J.; Ducros, V.; Lewis, R. J.; Borchert, T. V.; Schulein, M. Oligosaccharide Specificity of a Family 7 Endoglucanase: Insertion of Potential Sugar-Binding Subsites. *J. Biotechnol.* **1997**, *57*, 91−100.

(64) Munoz, I. G.; Ubhayasekera, W.; Henriksson, H.; Szabo, I.; Pettersson, G.; Johansson, G.; Mowbray, S. L.; Stahlberg, J. Family 7 Cellobiohydrolases from *Phanerochaete chrysosporium*: Crystal Structure of the Catalytic Module of Cel7D (CBH58) at 1.32 Å Resolution and Homology Models of the Isozymes. *J. Mol. Biol.* **2001**, *314*, 1097−1111.

(65) King, G.; Lee, F. S.; Warshel, A. Microscopic Simulations of Macroscopic Dielectric Constants of Solvated Proteins. *J. Chem. Phys.* **1991**, *95*, 4366.

(66) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L. What Ice can Teach Us about Water Interactions: A Critical Comparison of the Performance of Different Water Models. *Faraday Discuss.* **2009**, *141*, 251−276.