THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Experimental and Modeling Studies of an Unusual Water-Filled Pore Structure with Possible Mechanistic Implications in Family 48 Cellulases

Mo Chen,[†] Maxim Kostylev,[‡] Yannick J. Bomble,[§] Michael F. Crowley,[§] Michael E. Himmel,[§] David B. Wilson,*[,‡] and John W. Brady*[,†]

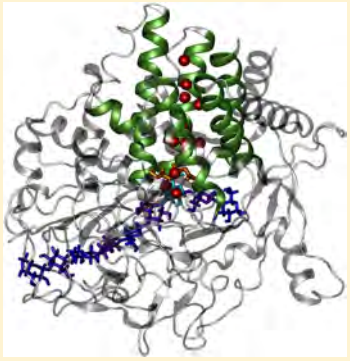
[†]Department of Food Science, Cornell University, Ithaca, New York 14853, United States

[‡]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, United States

[§]Biosciences Center, National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, Colorado 80401-3393, United States

Supporting Information

ABSTRACT: Molecular dynamics simulations were used to study the possible catalytic role of an unusual conserved water-filled pore structure in the family 48 cellulase enzyme Cel48A from *Thermobifida fusca*. It was hypothesized that this pore serves as the pathway for the water molecules consumed in the hydrolysis catalyzed by the enzyme to reach the active site in a continuous stream to participate in the processive reactions. Theoretical mutants of this enzyme were created in which all of the residues lining the pore were made hydrophobic, which had the effect in molecular dynamics simulations of emptying the pore of water molecules and preventing any from passing through the pore on the simulation time scale. Mutants with smaller numbers of substitutions of this nature, which could be created experimentally by site-directed mutagenesis, were also identified from simulations, and these proteins were subsequently produced in *Escherichia coli*, expressed and purified, but were found to not fold in a manner similar to the wild type protein, preventing the determination of the importance of the water pore for activity. It is possible that the presence of a small vacuum in the pore was responsible for the instability of the mutants. In addition, alternate pathways were observed in the simulations that would allow water molecules to reach the active site of the enzyme, suggesting that the hypothesis that the pore has functional significance might be incorrect.



INTRODUCTION

Cellulose is notoriously difficult to deconstruct, and nature has evolved a number of complex molecular machines for accomplishing the tasks of fibril decrystallization and glycosidic hydrolysis, perhaps best exemplified by the elaborate multicomponent cellulosome complexes.[1] Many of the processive exocellulases are individually interesting examples of complex molecular machines, able to attract single chains, bind them into their active site tunnels, and successively hydrolyze the glycosidic linkages. In general, the rate-limiting factor in the enzymatic deconstruction of cellulose is the insolubility of the substrate.[2] However, in the processive exocellulases, once a chain has been separated from the crystal, it is threaded into a long tunnel, with the active site asymmetrically located at one end, and successive glycosidic linkages are hydrolyzed, with the consumption of a steady stream of water molecules, as the chain is moved along the tunnel in assembly line fashion.

One set of particularly interesting examples of such processive cellulases is the family 48 cellobiohydrolases. In these enzymes, primarily found in bacteria, only every-other glycosidic linkage of the substrate is hydrolyzed, working from the reducing end of the cellulose chains, and the product is cellobiose rather than glucose. Enzymes in this family can function either independently, such as Cel48A from *Thermobifida fusca* (*T. fusca*), or complexed as catalytic components in cellulosomal assemblies, as in the case of CelS from *Clostridium thermocellum* (*C. thermocellum*) and CelF from *Clostridium cellulolyticum* (*C. cellulolyticum*). In the case of cellulosomes, family 48 cellulases are of key importance in crystalline cellulose hydrolysis,[3] and certain examples such as the *T. fusca* Cel48A are considered promising candidates for "designer cellulosomes".[4] The processivity of family 48 cellulases ranges from ∼4 to ∼7 glycosidic bond cleavages per enzyme/substrate chain binding event, depending on the cellulase and the substrate.[5,6]

More than twenty family 48 cellulases have been identified, and the structures of several have been solved by X-ray crystallography. Family 48 cellulases have unique structural features. Most noticeably, they contain an active site tunnel that can accommodate nine substrate monomer units in specific sites (referred to as subsites −7, −6, −5, −4, −3, −2, −1, +1, and +2), providing a substrate pathway for processive action with the scissile bond being that between the residues in subsites −1 and +1.[7,8] After hydrolysis, the product cellobiose is

Received: September 2, 2013
Revised: January 16, 2014
Published: January 29, 2014

   dx.doi.org/10.1021/jp408767j | *J. Phys. Chem. B* 2014, 118, 2306−2315

expelled from the end of the tunnel, and the chain advances to again fill subsites +1 and +2 for the next cleavage event. These family 48 cellulases act through an inverting mechanism (Figure 1). The active site employs a pair of amino acids, a glutamic

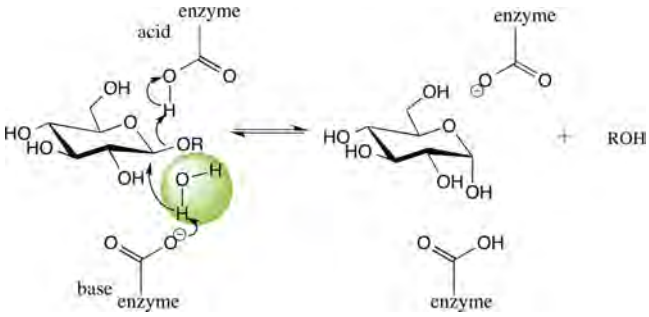


Figure 1. Inverting mechanism of family 48 cellulases.

acid acting as the catalytic acid and an aspartic acid acting as the catalytic base, to effect the catalysis. At the local pH in the active site region, the aspartic acid carboxyl group is unprotonated and acts as a nucleophile to form a covalent intermediate with a water molecule. This intermediate attacks the anomeric carbon of the substrate unit in the −1 subsite to trigger the hydrolysis. On the other side, the glutamic acid residue is protonated and attracts the glycosidic oxygen to gradually form a covalent bond, ending up breaking the glycosidic bond between two β-glucosyl units and resulting in an inversion of the stereochemistry at the product cleavage site.[9−11] Of course, water is essential for the hydrolysis, and each catalytic cycle consumes a water molecule.

The family 48 cellulases possess another very interesting structural feature, an extended pore structure running approximately 26 Å from the molecule surface to the active site in the core of the protein and formed by the six inner α-helices of a characteristic $(\alpha/\alpha)_6$ structure (see Figure 2a). The catalytic acid and catalytic base residues are both part of the helices that form this structure and are located at the base of the pore. This pore structure is large enough in diameter to accommodate a number of water molecules, and nine water molecules are present within this pore region of the CelF crystal structure[12] and twelve water molecules within that of the Cel48A crystal structure (Roman Brunecky, National Renewable Energy Laboratory (NREL), personal communication; Figure 2a), suggesting that the pore might function as a channel for water transport. Interestingly, most of the amino acid residues that line the sides of this pore are either conserved or partially conserved across family 48 cellulases; alignment of the water pore amino acid sequences for CelF, CelS, and Cel48A showed that 15 out of the 31 water pore residues were conserved and 12 were partially conserved (Table S1 of the Supporting Information), suggesting an evolutionary pressure to maintain this assemblage. Many studies have described how water molecules in such confined pores form chains, "wires", or clusters in nonpolar confinement, which are important in protein function.[13]

It is possible that the purpose of this conserved $(\alpha/\alpha)_6$ structure is simply to function as a rigid scaffold upon which to construct the functional portions of the enzyme, particularly since it serves to precisely position the catalytic residues with respect to one another and the substrate. However, the degree of conservation all along the pore and the fact that it leads directly and precisely to the active site, as well as the additional fact that even in the crystal structure it is filled with water molecules, all suggest that it is at least plausible that this structure serves the mechanistic role of providing an unobstructed pathway for the water of hydrolysis to enter the active site, allowing the last water molecule in the pore to be successively consumed by hydrolysis as each new cellobiose unit is cleaved from the substrate chain. In order to test this hypothesis, we have used a combination of computer simulations studies designed to identify mutants of the

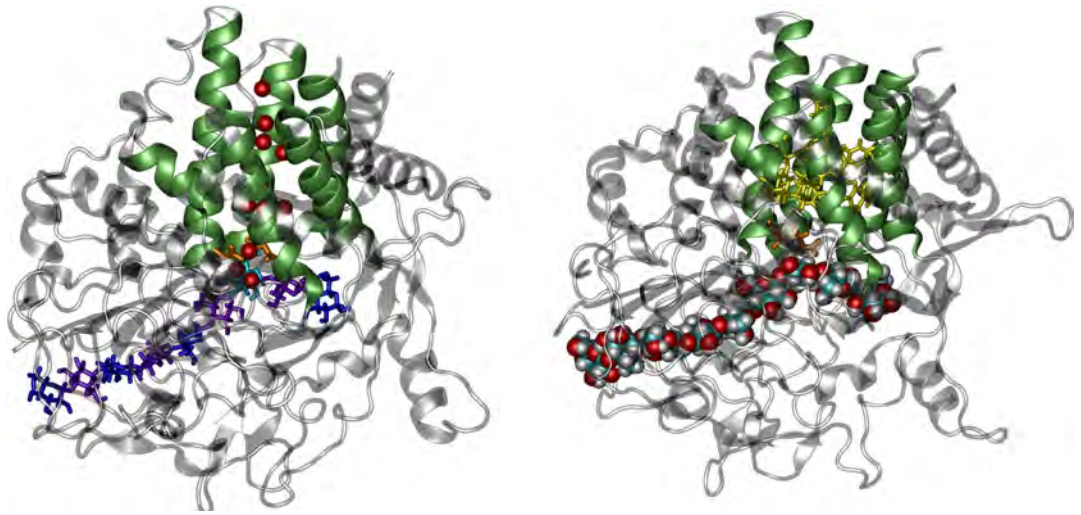


Figure 2. Left: Cartoon representation of the modeled structure of Cel48A. The substrate is illustrated in the so-called "lower path" conformation (see text) with successive glucose residues alternately colored dark blue and purple. In the crystal structure, the residue in the −1 binding site is missing, and in the illustration has been built with energy minimization using CHARMM, and is indicated in light blue. The catalytic groups are shown in orange. The helices constituting the hypothesized water pore are shown in green, and the crystallographic water molecules that fill this pore are shown as red van der Waals (vdW) spheres. Right: Mutant (see text) of this enzyme, indicating the location of the phenylalanine substitutions (in yellow) in this multiple mutant. The catalytic acid and base are shown in orange. This figure as well as all others displaying molecular images was prepared using the VMD graphics package.[14]

dx.doi.org/10.1021/jp408767j | J. Phys. Chem. B 2014, 118, 2306−2315

Table 1. Mutant Cel48A Cellulases Studied by MD Simulation[a]

location	residue	no. 1	no. 2	no. 3	no. 4	no. 5	no. 6	no. 7	no. 8	no. 9	Mut C	Mut T
helix 1	Gln427	X	X			X				X	X	X
	Glu423	X	X		X	X				X	X	X
	Arg424	X	X		X	X				X	X	X
	Trp420											X
helix 2	Lys507	X	X	X	X	X	X				X	X
helix 3	Asn328	X		X	X	X			X		X	X
helix 4	Trp235											X
	Gln231	X	X				X	X	X		X	X
	Arg228	X	X					X			X	X
helix 5	Trp58											X
	Tyr55			X	X	X	X	X	X	X	X	X
helix 6	Glu625	X	X	X	X	X	X	X	X	X	X	X

[a]The unsuccessful mutants are numbered from 1 through 9. In addition, a tenth mutant, mutant C (Mut C), which was found to be a promising candidate, and a control, mutant T (Mut T), in which all of these residues were mutated, were also simulated. Note: "X" indicates mutation sites in each mutant, and all mutation sites were changed into Phe. Each of the intermediate unsuccessful mutants (numbers 1−9) was simulated for 24 ns of data collection following the same protocol as that of the WT and Mut C simulations (see text). All of the mutants except Mut C allowed water molecules to pass through the pore on this time scale. Mutants 3−9, and Mut C, were all produced experimentally, but none were found to be stable during isolation.

Cel48A protein which blocked passage of water molecules along this pore, along with site-directed mutagenesis experiments to actually produce such mutants in the laboratory and assay their activity, to attempt to determine if obstructing this water channel affects the turnover rate of the enzyme.

METHODS

In order to test the hypothesis that this water pore serves the mechanistic role of providing a path to the active site for the water molecules consumed in the hydrolysis, it is necessary to identify mutants that can block the passage of water molecules through the pore without otherwise disrupting the native protein fold. This was attempted by using molecular dynamics (MD) simulations to determine which mutants prevented the passage of water molecules to the active site. Due to the slow turnover rate for these cellulases, simply slowing the passage of water molecules was deemed unlikely to produce a detectable effect, so multiple mutants were sought that could completely block the movement of water molecules along the channel. Of course, mutants that prevent the protein from folding properly would not be suitable, but MD simulations cannot determine overall stability beyond determining whether there is a short time tendency for the folded structure to change significantly, perhaps as a result of steric clashes, for example. Once an acceptable candidate Cel48A mutant was identified from the computational work, an attempt was made to produce and isolate this protein in the laboratory in order to assay its activity.

MOLECULAR MODELING AND SIMULATIONS

Although mutants of the protein Cel48A can be produced in the laboratory in *Escherichia coli* (*E. coli*), at the beginning of this study the crystal structure of this protein was not yet available. However, in general, cellulases from the same family share similar three-dimensional structures.[9] Since the Protein Data Bank contains structures for two family 48 cellulases that could be used to build Cel48A homologous models, it was decided to use homology modeling to construct a plausible starting structure for Cel48A for the simulations. These two proteins are CelF variants from *C. cellulolyticum*, which has the highest sequence identity (54%), and CelS from *C. thermocellum*, which has a sequence identity of 52% with Cel48A. Based on the amino acid sequence identity, CelF was used as the template to build the homology model, even though a high-resolution structure of the wild type (WT) CelF itself with substrate occupying all subsites was not available. Several X-ray crystal structures of mutants of CelF with various ligands occupying the substrate and product site in the protein have been solved and reported by Parsiegla and co-workers,[12,15,16] The most recently reported CelF structures correspond to two single mutants, namely, CelF_E55Q and CelF_E44Q, with each containing a substrate analog that fills the active site tunnel.[12] However, in these two structures, the substrate chains do not occupy identical paths in the active site tunnel. The former structure contains a thio-oligosaccharide nonamer occupying the binding subsites from −7 to +2 of a "lower path" in the active site tunnel of the enzyme (Figure 2 and Figure S1a of the Supporting Information), while the latter contains a thio-oligosaccharide octamer occupying the subsites from −7 to +1 of an "upper path" in the tunnel (see Figure S1b of the Supporting Information).

The Standard Protein BLAST engine[17] with the Position-Specific Iterated BLAST algorithm was employed to perform Cel48A sequence alignment to all of the available protein crystal structures in the Protein Data Bank. As a result of this search, the CelF_E55Q structure was selected as the optimal template for building the Cel48A homologous model (Table S2 of the Supporting Information), which was achieved with Swiss-PdbViewer using the "magic fit" alignment algorithm.[18]

In the final stages of this study, workers at the National Renewable Energy Laboratory solved the structure of a Cel48A single mutant (D224A) using X-ray crystallography (Vladimir V. Lunin, personal communication). The comparison of the residue-based root-mean-square (rms) difference between the homologous model and the Cel48A wild type crystal structure that was obtained from the D224A crystal structure is shown in Figure S2 of the Supporting Information. As can be seen from that figure, the homologous model was acceptably close to the crystal structure for the present purposes.

To analyze the role of the pore in water transport, molecular dynamics simulations were used to study the behavior of the cellulases and their mutants in aqueous solution. The molecular

dx.doi.org/10.1021/jp408767j | J. Phys. Chem. B 2014, 118, 2306−2315

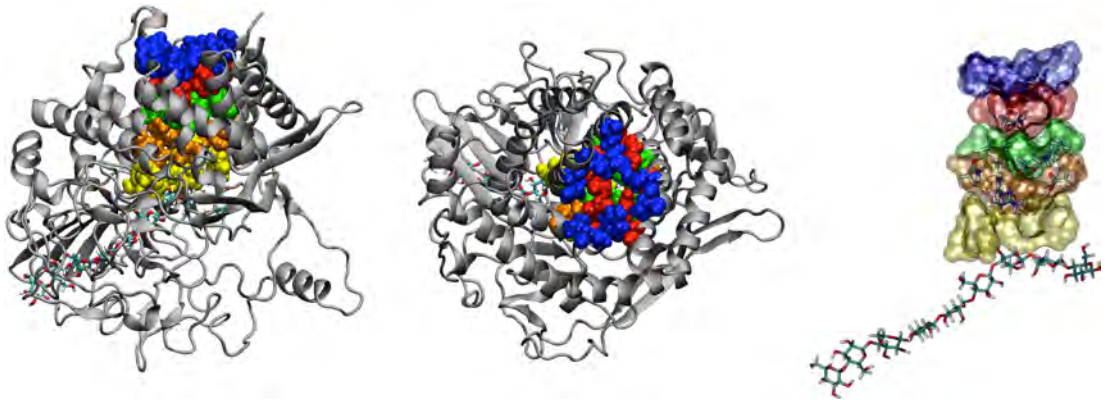


Figure 3. Side view (left) and top view (middle) of the water pore structure in Cel48A, illustrating the division of the pore into rings 1, 2, 3, 4, and 5, colored blue, red, green, orange, and yellow. On the right, only the five rings are illustrated as colored van der Waals surfaces, along with the substrate chain.

mechanics program CHARMM[19,20] was used to build the systems, while the program AMBER[21] was used to calculate the trajectories. In each simulation, the cellulase molecule with a cellooligomer ligand (degree of polymerization = 9) was built from the crystal structure or the homologous model. The protein and ligand were then placed in a water box, and the overlapping water molecules were removed, resulting in about 20,000 water molecules in each system. Counter ions were added to the solution to neutralize the charges of the systems. The CHARMM22 force field with the CMAP correction[22] was used to describe the protein, while the new CHARMM carbohydrate force field[23] was used to describe the cellooligomer, and water molecules were represented by the TIP3P water model.[24,25] The lengths of the covalent bonds involving hydrogen atoms were kept fixed using the SHAKE algorithm.[26,27] The CHARMM structure files, coordinate files, and the associated force fields were converted into the AMBER format using CHAMBER.[28] The PMEMD engine of AMBER was employed to carry out the molecular dynamics simulations to exploit its greater computational efficiency.

The system preparation of Cel48A and its mutants followed four stages, which were as follows: solvent minimization with 1000 steepest descent steps and 1000 conjugate gradient steps; system minimization with the same minimization strategy; equilibration of the solvent at constant volume from 0 to 300 K for 20 ps; and equilibration of the system in the *NPT* ensemble at 300 K with a Langevin thermostat at 1 atm with a "weak coupling" constant pressure algorithm for 250 ps with a step size of 2 fs. The production runs were carried out in the *NPT* ensemble with the same strategy as the system equilibration and were initially carried out for 24 ns. The Cel48A simulations performed are listed in Table 1. During the preliminary search for mutants that might be effective in preventing the passage of water molecules through the tunnel, this short 24 ns simulation time was used since once a mutant was found to fail on this short time scale (in the sense that at least one water molecule was able to transit through the pore), it was no longer necessary to simulate that mutant further. However, for the Cel48A wild type control and for mutant C, which was found to be promising, the production runs were carried out three times, with the water box positioned at different orientations each time in setting up the systems, and each production run was carried out for 100 ns. For the all-Ile CelF mutant, the system preparation was the same as for Cel48A except that the equilibration of the system was conducted for 100 ps. Subsequently, 400 ps of *NPT* production run data were collected for two simulations, one in which the pore was initially filled with water (called the "dewetting" simulation) and one in which the pore contained no water molecules (called the "dry" simulation). Because the dewetting in this all-Ile mutant occurred quickly, this simulation was run for a total of just 500 ps as a proof of concept only.

To analyze water flow within the water pore structure, the structure was divided into five sections, labeled as rings 1−5 (see Figure 3). These ring-shaped sections stacked together, forming a columnlike shape connecting the hydrophobic core at the active site to the protein surface. The components of each ring in CelF and Cel48A were listed in Table S3 of the Supporting Information.

The free energy penalty for creating a small cavity inside the water pore region of the Cel48A WT was calculated using a thermodynamic integration (TI) method and the Widom particle-insertion approach.[29,30] In this calculation the inner pore region was defined to be a spherical region centered at the geometric center of the β carbon atoms of the five residues making up the wall of ring 3 (with protein residue identities (ids) of 628, 507, 58, 424, and 231), and with a radius R_{vac} (6.4 Å), within which no water molecules exist in the corresponding well-equilibrated mutant C (Mut C) structure. The initial stage corresponded to no cavity in the pore, and the end stage corresponded to a cavity defined by an empty sphere of radius R_{vac} in the pore, where a user-defined energy (a rescaled truncated-shifted Lennard-Jones (LJ) potential) was applied at the center of the spherical cavity:

$$U_{e0} = \begin{cases} 100\left(\left(\frac{R_{vac}}{r}\right)^{12} - 2\left(\frac{R_{vac}}{r}\right)^{6} + 1\right) & \text{if } r \le R_{vac} \\ 0 & \text{if } r > R_{vac} \end{cases} \quad (1)$$

This repulsive energy was only calculated with the water atoms; the protein atoms did not experience this force.

The free energy change was calculated for both directions: (1) starting from the initial stage and gradually creating a spherical cavity in the pore until it reached the end stage; (2) starting from the end stage and gradually compressing the cavity size until it disappeared. Each process was achieved by introducing a scaling factor λ for the cavity radius. That is, at each intermediate stage

dx.doi.org/10.1021/jp408767j | J. Phys. Chem. B 2014, 118, 2306−2315

$$U_e = \begin{cases} 100\left(\left(\frac{\lambda R_{vac}}{r}\right)^{12} - 2\left(\frac{\lambda R_{vac}}{r}\right)^{6} + 1\right) & \text{if } r \le \lambda R_{vac} \\ 0 & \text{if } r > \lambda R_{vac} \end{cases} \quad (2)$$

Using the TI method, the free energy change corresponding to each process (under *NPT* conditions) was calculated as

$$\Delta G = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda \quad (3)$$

where U is the total energy of the system. In general, U does not vary too much due to small volume changes; for example, here it was caused by the creation of the small cavity, and as only the user-energy component (U_e) is explicitly λ-dependent, $\langle \partial U/\partial \lambda \rangle_\lambda$ was estimated as

$$\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda = \left\langle \frac{\partial U_e}{\partial \lambda} \right\rangle_\lambda = \begin{cases} 100 \times 12 \times \left(\left(\frac{R_{vac}}{r}\right)^{12} \lambda^{11} - \left(\frac{R_{vac}}{r}\right)^{6} \lambda^{5}\right) & \text{if } r \le \lambda R_{vac} \\ 0 & \text{if } r > \lambda R_{vac} \end{cases} \quad (4)$$

The MD simulations were carried out in CHARMM using Langevin dynamics under constant pressure (1 bar) and constant temperature (300 K) conditions. A series of simulations were carried out sequentially (corresponding to increasing or decreasing λ values), and each simulation lasted for 100 ps with a time step of 1 fs. The $\langle \partial U_e/\partial \lambda \rangle_\lambda$ term was calculated from each simulation. The discrete integration was calculated using the trapezoidal rule.[31]

EXPERIMENTAL CHARACTERIZATIONS OF CEL48A AND ITS MUTANTS

1. Enzyme Expression and Purification. *T. fusca* Cel48A (TfCel48A) contains a 6-His tag on both the amino and carboxyl termini and no signal peptide. WT and mutant TfCel48A were expressed inside *E. coli* BL21−CodonPlus (*DE3*)−*RIPL* cells (Agilent Technologies) as follows. A starter culture was grown in Luria broth (LB) medium overnight at 37 °C. The starter culture was diluted 33 times in fresh LB, and the cells were grown at 37 °C until the optical density at 600 nm reached ∼0.8 (about 2.5 h). The culture was transferred to 25 °C and isopropyl β-D-1-thiogalactopyranoside was added to 0.8 mM. The culture was incubated with shaking for 16 h. Cells were harvested by centrifugation and resuspended in 1/20 original culture volume of 20 mM sodium phosphate buffer, pH 8.0, 0.01 M imidazole (Sigma-Aldrich), 0.5 M sodium chloride (solution A). Cells were lysed with a French pressure cell press (at cell pressure of 20,000 PSI). Lysed cells were placed in a 50 °C water bath for 30 min (to precipitate out *E. coli* proteins) and centrifuged to remove the precipitate. The supernatant was diluted in five additional volumes of solution A and loaded on a nickel Sepharose 6 fast flow (GE Healthcare) column by gravity. The protein was eluted by a gradient of 0.01−0.5 M imidazole in 20 mM sodium phosphate buffer, pH 8.0, and 0.5 M sodium chloride. Fractions containing TfCel48A, as determined by denaturing polyacrylamide gel electrophoresis (SDS-PAGE), were combined and diluted 10−15 times with DI water (to lower the conductivity of the solution to less than 0.5 mA). The protein was then loaded on a Q Sepharose (Sigma-Aldrich) column equilibrated with 0.1 M sodium chloride in 5 mM Bis-Tris buffer, pH 5.8, and 10% glycerol (v/v). The protein was eluted by a gradient of 0.1−0.5 M sodium chloride in 5 mM Bis-Tris buffer, pH 5.8, and 10% glycerol (v/v). Fractions containing TfCel48A, as determined by SDS-PAGE, were combined and concentrated using Millipore centrifugal filter units with a 30 kDa cutoff membrane. Buffer exchange was carried out in the same filter units by washing the concentrated protein three times with 5 mM sodium acetate buffer, pH 5.5, and 10% glycerol (v/v). Protein was filtered through a 0.22 μm filter using a syringe and stored at −20 °C.

TfCel48A mutants were generated using an Agilent Quickchange II XL site-directed mutagenesis kit following the manufacturer's instructions. All mutations and sequences were verified by Sanger DNA sequencing. Mutant proteins were expressed and purified in the same way as WT. All enzyme concentrations were determined by spectroscopy using a NanoDrop 1000 spectrophotometer.

2. Protein Stability Analysis by Native (Nondenaturing) PAGE. Mutant proteins were expressed in *E. coli* BL21 cells as described for WT. A 0.3−0.5 g amount of cells was harvested by centrifugation and resuspended in 100 μL of 9 mM Tris-Cl buffer, pH 8.0, and 1.9 mM ethylenediaminetretraacetic acid (EDTA), containing 1 mg/mL chicken egg white lysozyme (Sigma). Cells were incubated at 37 °C for 2 h and centrifuged at 4 °C at maximum speed for 30 min. Cell lysate supernatant was analyzed by native PAGE, using purified WT Cel48A for reference and WT Cel48A-expressing BL21 cell lysate for positive control. If an expected overexpressed band was present on SDS-PAGE, but not on a native gel, it was concluded that the mutant protein is unstable.

RESULTS AND DISCUSSION

In the initial effort to find mutants that would halt the movement of water through the pore, an attempt was made to mechanically block the pore by substitution with bulky side chains, specifically, with tryptophan and arginine, two of the largest of the amino acids. Two single-site mutants of the CelF enzyme were made: Q420W, placing a tryptophan residue in ring 3 of the pore, and W62R, replacing a wild type tryptophan in ring 3 with an arginine residue, which it was anticipated would make salt bridges across the pore with Glu 424 and Glu 615. Three double substitutions introducing bulky tryptophan side chains were also attempted: Q420W and V622W, Q420W and V618W, and Q420W and Q237W, where in each case the pair was chosen such that the selected residues were on opposite sides of the pore. Each of these mutants was simulated for no more than 2 ns, since none of these substitutions was found to inhibit the diffusion of water through the pore, as the bulky side chains simply rotated out of the way on a frequent basis. From these results it seemed unlikely that the pore could be closed off solely by mechanical obstruction, and a new approach was adopted that attempted to make the lining of the pore sufficiently hydrophobic such that water molecules could not pass though. This approach thus involved substituting the residues lining the interior surface of the pore with hydrophobic amino acids such as isoleucine or phenylalanine.

For example, a CelF mutant was constructed by substituting all of the water pore residues except the catalytic acid (Glu 55) and catalytic base (Asp 230) with isoleucines. Since it would be infeasible to experimentally produce a corresponding mutant protein with so many substitutions, several Cel48A mutants were designed with a reduced number of mutation sites to ease

dx.doi.org/10.1021/jp408767j | J. Phys. Chem. B 2014, 118, 2306−2315

Table 2. Mutation Sites of Cel48A Water Pore Mutants

location	mutant 1	mutant 2	mutant 3	mutant 4	mutant C	ring site
helix 1		Glu423	Gln427	Gln427	Gln427	ring 2
		Arg424	Glu423	Glu423	Glu423	ring 3
			Arg424	Arg424	Arg424	rings 3 and 4
helix 2	Lys507	Lys507	Lys507	Lys507	Lys507	ring 3
helix 3	Asn328	Asn328	Asn328		Asn328	ring 4
helix 4				Gln231	Gln231	ring 3
				Arg228	Arg228	ring 4
helix 5	Tyr55	Tyr55	Tyr55	Tyr55	Tyr55	ring 4
helix 6	Glu625	Glu625	Glu625	Glu625	Glu625	ring 4

the experimental efforts. The mutations sites were all located within the inner rings (rings 2, 3, and 4) of the water pore (Table 2), with all residues selected for mutation converted to Phe, with a side chain that is both bulky and hydrophobic.

1. CelF Mutant. Two simulations were conducted for the all-isoleucine CelF mutant: one termed the "dewetting" simulation that began with the water pore filled with water and one termed the "dry" simulation that began with the pore empty (on the assumption that water could potentially enter and wet the pore). The dewetting simulation was initiated by filling the pore with water in the simulation starting structure. As expected, the water molecules initially occupying the middle rings of the pore quickly emptied out, as can be seen from Figure 4a. In the second, dry simulation, which began from a starting structure without water molecules in this pore, no water molecules entered this region during this short simulation (Figure 4b). These results demonstrate the feasibility of using hydrophobicity to prevent the passage of water molecules from the exterior through the pore to the active site.

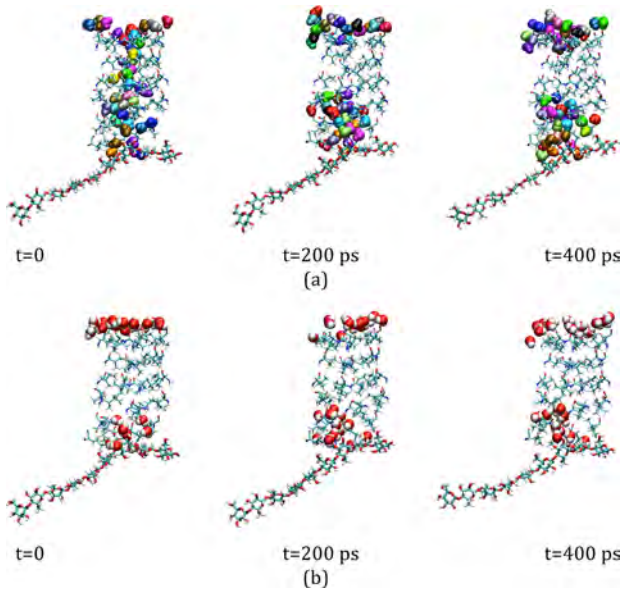


Figure 4. "Dewetting" (a) and "dry" (b) simulations of CelF water pore mutant, in which all water pore residues were converted into Iles except Glu55 and Asp230. The molecules represented using vdW spheres were water molecules; the water pore residues were presented using thin licorice, and the substrate residues in medium licorice.

2. Cel48A Mutants. While short-term MD simulations, even those that extend to tens of nanoseconds, cannot determine whether a protein conformation is stable, no tendency was observed for the mutant proteins studied in these simulations to unfold over the course of the trajectories. In each case, the rms deviation (rmsd) of the protein backbone atoms (C, CA, and N) from their positions in the equilibrated protein structure quickly reached stable values between 1.5 and 2.5 Å and subsequently showed no further tendency to drift away from the starting conformation, suggesting stability at least on this time scale (see Figure S3 of the Supporting Information).

The movement of water molecules in the pore suggested that mutant C, with nine Phe substitutions, was the best candidate, among all those with limited substitutions studied, in preventing water transport through the pore. The temporal evolution of the number of water molecules within the inner rings (rings 2, 3, and 4) were characterized, and the results demonstrated that almost no water molecules were present in ring 3 of mutant C, and the total number of water molecules within the three inner rings in mutant C was much less than that in the wild type (Table 3). Trajectories for all of the water molecules that were once in the pore in each of the simulations were followed to determine their positions relative to the water pore over the simulations. The results demonstrated that water molecules could move from one side of the water pore to the other in the wild type, but could not traverse the pore in mutant C on the time scale simulated, suggesting that this more limited mutant might be as successful as the case where all 31 of the residues lining the pore were changed to isoleucine.

Table 3. Mean Number of Water Molecules within Rings 2, 3, and 4, and Ring 2 + Ring 3 + Ring 4, and Their Standard Deviations[a]

	WT	Mut C
ring 2	4.426 ± 0.877	6.027 ± 0.939
ring 3	8.937 ± 0.906	0.837 ± 0.432
ring 4	11.129 ± 1.220	5.223 ± 1.814
rings 2 + 3 + 4	15.999 ± 1.666	11.367 ± 2.212

[a]Here the data are the averages of three trajectories. WT refers to wild type, and Mut C refers to mutant C. Note: Water molecules may exist within the limits of multiple rings.

This multiple point mutant was successfully produced experimentally, but unfortunately, it did not prove possible to isolate the folded protein in order to assay its activity, as it appears to be unstable. The mutant is expressed at similar to WT levels, as judged by SDS-PAGE of cell lysates after protein expression. However, when analyzed by native PAGE, cell lysate containing mutant C has a very weak band corresponding to the overexpressed Cel48A protein, indicating that the protein is either denatured or is not folded similarly to WT. In addition, when attempts at purification were made, mutant C

dx.doi.org/10.1021/jp408767j | J. Phys. Chem. B 2014, 118, 2306−2315

did not interact with the anion exchange column in the same way that is observed for WT. Analysis of mutants 3−9 by SDS and native PAGE produced the same results as observed for mutant C, indicating that all mutants had similar stability issues. The reason that the mutant proteins are not stable is not clear. As already noted, there was no indication from the relatively short trajectories that the conformation of the multiple mutant was unstable since the rms drift away from the starting structure quickly reached a low and stable value with no further drift. It is possible that the instability of the protein is due to the very possibility that the interior of the pore is unoccupied in this mutant, since as the water molecules empty out of the pore, they leave behind a vacuum, and there is of course a free energy penalty to creating a vacuum against the exterior pressure of 1 atm. This free energy penalty can be estimated from standard treatments of cavity formation in liquid water.[32,33]

The free energy cost for generating small cavities in water can be approximated as[34]

$$\Delta G \approx k_B T \frac{\rho^2 V^2}{2\chi_V} + k_B T(\ln 2\pi\chi_V)/2 \tag{5}$$

with

$$\chi_V = \langle(\delta N)^2\rangle_N = \rho V + \rho^2 \int_V \mathrm{d}r \int_V \mathrm{d}r' \,[g(|r - r'|) - 1] \tag{6}$$

where ρ is the number density of the solvent (water), V is the cavity volume, and $g(r)$ is the radial distribution function of a pure water oxygen−oxygen pair. In our study, we defined the cavity volume as the vacuum space in the pore that could fit in water molecules. That is, the cavity volume was equal to the number of water molecules that could fit into the vacuum multiplied by the volume of an individual water molecule in solvent at the desired temperature. Approximating the total volume of the cavity to be a sphere, χ_V is given by

$$\chi_V = \rho V + \frac{16\pi^2}{3}\rho^2 r_0^3 \int_0^{r_0} \mathrm{d}n \,[n^2 g(n)] \tag{7}$$

where

$$r_0 = \left(\frac{3V}{4\pi}\right)^{1/3} \tag{8}$$

The results for this estimated free energy cost corresponding to N, the number of water molecules that could fit in the pore vacuum, are shown in Table 4. In the CelF mutant, five water molecules were lost from the pore, with an increase in the free energy of ∼2.0 kcal/mol, whereas in the Cel48A mutant C, the vacuum volume was equivalent to only one or two water molecules, corresponding to 1−1.5 kcal/mol of free energy. Since the ΔG for protein unfolding is typically on the order of 5 kcal/mol; these relatively small free energy penalties for vacuum formation in the pores might nevertheless be large enough to destabilize the proteins. Also, aromatic−aromatic side chain interactions can be on the order of −1.0 to −1.4 kcal/mol;[35] thus these mutated Phe residues might lead to protein misfolding as well.

Table 4. Free Energy Cost of Small Cavity Creation in the Water Pore[a]

N	ΔG (kcal/mol)
1	1.1
2	1.5
3	1.7
4	1.9
5	2.0

[a]Note: N is the number of water molecules displaced by the vacuum.

Conversely, the results of the TI free energy estimate based on the Widom particle-insertion approach to growing a vacuum in the specific environment of the pore interior found a lower value. In that simulation, approximately nine water molecules were displaced from the central pore of the wild type protein, but at a free energy cost of only around 0.23 kcal/mol. An energy change of this magnitude would be unlikely to fatally destabilize a protein toward folding at room temperature. It should be noted, however, that this value is for the wild type protein, not the mutant C produced in the experiments. It is operationally more difficult to design a comparable calculation for the mutant, since the water molecules immediately leave the pore spontaneously in the simulations of that system, and it would be necessary to force water molecules back into the pore in a TI simulation to get an estimate of the free energy difference, which would potentially give rise to significant artifacts. It is difficult to extrapolate from the wild type simulation, where the side chains are polar and presumably have an affinity for the water molecules that should make them harder to remove, to the mutant with its nonpolar side chains that expel the water molecules automatically.

3. Pathways of the Hydrolytic Water Molecules. While the simulations demonstrated that water molecules can easily move through the pore from the protein surface to the active site (Figure 5a), supporting the hypothesis that this channel has possible mechanistic significance, the Cel48A wild type simulation also offered evidence to suggest otherwise. In this simulation, water molecules were also observed to enter the product exit channel of the active site tunnel and diffuse "backwards" around the substrate/product carbohydrate residues until they reached the active site. Eighty-nine instances of this type were observed in the three replicates of the wild type simulations of Cel48A (see Table 5; note that no net flux is implied since an approximately equal number of water molecules diffused out of the active site region, but their trajectories were not followed in detail). More surprisingly, the loops that close to form the bottom wall of the active site tunnel undergo fluctuations during the trajectories that allow water molecules to work their way into the active site from underneath the protein surface, apparently not requiring the paths through either the active site tunnel or the pore. Figure 5b illustrates a typical example of the path followed by one such water molecule. Twenty-six water molecules were observed to follow this route to the active site in the three replicate simulations. Both of these routes were not only still available in the mutant C trajectories, but there were actually more water molecules which entered the active site by these routes in those simulations, while none entered through the hydrophobic pore of this mutant. Given the slow turnover rate for this enzyme, these alternate routes would seem to obviate the necessity of a more facile path for the waters of hydrolysis to follow, since they could supply water molecules at a rate faster than water molecules would be consumed by hydrolysis. Unfortunately, the instability of the actual mutant proteins produced experimentally made it impossible to test this hypothesis in the present studies.

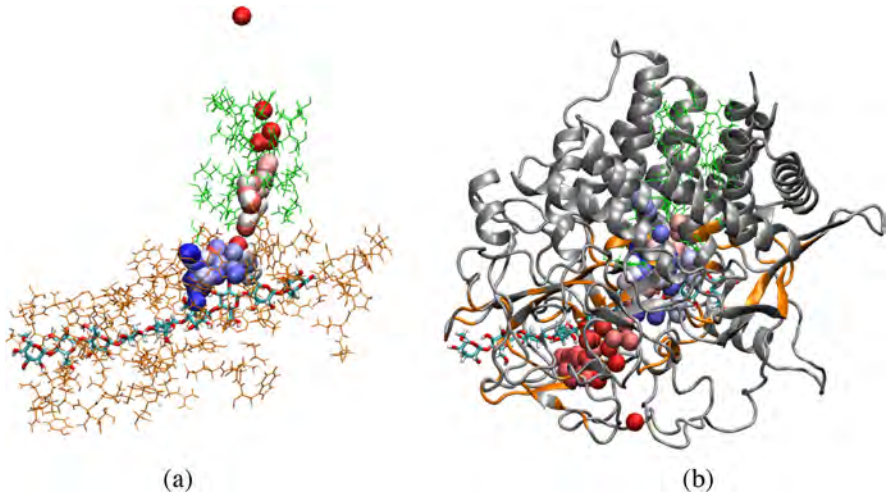


Figure 5. Trajectories for representative single water molecules penetrating to the active site by different routes: (a) a water molecule diffusing through the pore in the wild type Cel48A simulation, where only the residues lining the active site tunnel (in red) and water pore (in green) are shown, along with the cellooligomer substrate; (b) a water molecule penetrating to the active site by penetrating between the loops forming the "underside" of the active site tunnel. In these trajectories, only the water oxygen atoms are shown, color coded to represent the time evolution, with the color trend from red to white to blue equivalent to progression from beginning to end.

Table 5. Pathway of Water Molecules That Had Been Present at the Active Site

	no. of water molecules					
	WT_rep1	WT_rep2	WT_rep3	CMUT_rep1	CMUT_rep2	CMUT_rep3
N_{tot}[a]	57	53	49	127	134	250
path 1[b]	17	5	4	0	41	24
path 2[c]	26	32	31	122	86	217
path 3[d]	10	13	12	N/A	N/A	N/A
path 4[e]	4	3	2	5	7	9

[a]N_{tot}: the total number of the selected water molecules. [b]Path 1: bottom loops → active site region. [c]Path 2: tunnel exit → active site region. [d]Path 3: water pore → active site region. [e]Path 4: active site region → aqueous environment.

CONCLUSION

The water-filled pore observed in the crystallographically determined conformations of family 48 cellulases is a remarkable structure which inevitably raises the question of whether it has any functional significance. The hypothesis that it provides a pathway to the active site for the water molecules used in the processive hydrolysis of the cellulose substrate is reasonable and at least plausible on circumstantial grounds. The processive nature of the catalytic action of these exocellulases would require a steady stream of water molecules, and an unobstructed pathway such as this pore would appear to be the most efficient means of ensuring that the availability of water was not a rate-limiting factor in the reaction. The fact that this pore leads from the outer surface directly to the active site is also highly suggestive, as is the fact that it is filled with water molecules in the crystal structures. In addition, the very high degree of conservation of the residues that line this pore also suggests that there is significant selective pressure maintaining this structure. For all of these reasons, testing whether or not the disruption of the passage of water molecules through the pore affects activity or turnover rates is worth testing.

Although the simulations conducted in this study were quite extensive, they were not able to resolve the question of the functional importance of the water pore. The MD simulations of multiple point mutants in which the pore is made significantly hydrophobic demonstrate that, at least on the time scale of the simulations, water can be prevented from passing through the pore. Unfortunately, however, the difficulties of isolating such actual experimental mutants of Cel48A with folds similar to the native WT enzyme prevented the determination of whether these mutations would have affected activity if folded correctly. It is unclear why the mutant proteins should be so unstable, but it is possible that the presence of an empty, vacuum-filled pore represents a significant source of instability disfavoring nativelike folding. It is possible that future studies might identify a cofactor such as a short-chain fatty acid like butanoic acid or a linear alcohol that might be attracted to and bind in the pore, both blocking it physically and stabilizing the protein fold. It might even be possible to design a synthetic ligand with a specific affinity for the pore in the WT enzyme that would effectively block the pore without the necessity of making mutants. Such a system could then be used to assay the effects on the activity of blocking the pore.

However, the results of the present simulations also muddy the simple picture that originally motivated the study, since it was observed that water molecules in the simulation could diffuse back through the active site tunnel exit channel all the way to the active site. Even more unexpectedly, water molecules were also observed to bore their way into the active site tunnel from "below" by exploiting brief local fluctuations in the loops that close off the active site tunnel. Because the action of these cellulases is relatively slow, both of these paths would suggest that a special pathway for water to follow to the active site may not be necessary and would seem to disprove the hypothesis. At

dx.doi.org/10.1021/jp408767j | J. Phys. Chem. B 2014, 118, 2306−2315

this point it would seem that the present results have failed to resolve the question one way or the other and that further studies will be necessary to settle the issue.

■ ASSOCIATED CONTENT

Supporting Information

Figure S1 illustrating the "upper" and "lower" substrate paths in the CelF active site tunnels, Figure S2 showing a comparison of the crystal structure of Cel48A with the conformation built by homology modeling before the crystal structure was available, Figure S3 plotting the rms drift of the WT and MutC proteins away from their starting structures and demonstrating the stability of the mutants over 100 ns, Table S1 illustrating the conservation of the water pore residues across the family 48 cellulases, Table S2 listing the protein sequence alignment of Cel48A, and Table S3 comparing the water pore residue numbers between the CelF and Cel48A proteins and how they are used to define the rings in the pore. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*Tel.: +1 (607) 255-2897. E-mail: jwb7@cornell.edu.
*Tel.: +1 (607) 255-5706. E-mail: dbw3@cornell.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank U. Schnupf for helpful discussions. This project was supported by the U.S. Department of Energy (DOE) Office of Science, Office of Biological and Environmental Research through the BioEnergy Science Center (BESC), a DOE Bioenergy Research Center.

■ REFERENCES

(1) Bayer, E. A.; Chanzy, H.; Lamed, R.; Shoham, Y. Cellulose, Cellulases and Cellulosomes. *Curr. Opin. Struct. Biol.* **1998**, *8*, 548−557.
(2) Bergenstråhle, M.; Wohlert, J.; Himmel, M. E.; Brady, J. W. Simulation Studies of the Insolubility of Cellulose. *Carbohydr. Res.* **2010**, *345*, 2060−2066.
(3) Wilson, D. B. Demonstration of the importance for cellulose hydrolysis of CelS, the most abundant cellulosomal cellulase in *Chlostridium thermocellum*. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 17855−17856.
(4) Caspi, J.; Irwin, D.; Lamed, R.; Li, Y.; Fierobe, H.-P.; Wilson, D. B.; Bayer, E. A. Conversion of *Thermobifida fusca* free exoglucanases into cellulosomal components: Comparative impact on cellulose-degrading activity. *J. Biotechnol.* **2008**, *135*, 351−357.
(5) Vuong, T. V.; Wilson, D. B. Processivity, Synergism, and Substrate Specificity of Thermobifida fusca Cel6B. *Appl. Environ. Microbiol.* **2009**, *75*, 6655−6661.
(6) Zhang, X.-Z.; Zhang, Z.; Zhu, Z.; Sathitsuksanoh, N.; Yang, Y.; Zhang, Y.-H. P. The noncellulosomal family 48 cellobiohydrolase from Clostridium phytofermentans ISDg: Heterologous expression, characterization, and processivity. *Appl. Microbiol. Biotechnol.* **2010**, *86*, 525−533.
(7) Guimarães, B. G.; Souchon, H.; Lytle, B. L.; Wu, J. H. D.; Alzari, P. M. The Crystal Structure and Catalytic Mechanism of Cellobiohydrolase CelS, the Major Enzymatic Component of the Clostridium thermocellum Cellulosome. *J. Mol. Biol.* **2002**, *320*, 587−596.
(8) Parsiegla, G.; Reverbel, C.; Tardif, C.; Driguez, H.; Haser, R. Structures of mutants of cellulase Ce148F of *Clostridium cellulolyticum* in complex with long hemithiocellooligosaccharides give rise to a new view of the substrate pathway during processive action. *J. Mol. Biol.* **2008**, *375*, 499−510.
(9) Henrissat, B.; Bairoch, A. Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **1996**, *316*, 695−696.
(10) Sinnott, M. L. Catalytic Mechanisms of Enzymatic Glycosyl Transfer. *Chem. Rev.* **1990**, *90*, 1171−1202.
(11) Barr, B. K.; Hsieh, Y.-L.; Ganem, B.; Wilson, D. B. Identification of Two Functionally Different Classes of Exocellulases. *Biochemistry* **1996**, *35*, 586−592.
(12) Parsiegla, G.; Reverbel, C.; Tardif, C.; Driguez, H.; Haser, R. Structures of mutants of cellulase Cel48F of Clostridium cellulolyticum in complex with long hemithiocellooligosaccharides give rise to a new view of the substrate pathway during processive action. *J. Microbiol.* **2008**, *375*, 499−510.
(13) Rasaiah, J. C.; Garde, S.; Hummer, G. Water in nonpolar confinement: From nanotubes to proteins and beyond. *Annu. Rev. Phys. Chem.* **2008**, *59*, 713−740.
(14) Humphrey, W.; Dalke, A.; Schulten, K. VMD-Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(15) Parsiegla, G.; Juy, M.; Reverbel-Leroy, C.; Tardif, C.; Belaich, J. P.; Driguez, H.; Haser, R. The crystal structure of the processive endocellulase CelF of Clostridium cellulolyticum in complex with a thiooligosaccharide inhibitor at 2.0 Ångstrom resolution. *EMBO J.* **1998**, *17*, 5551−5562.
(16) Parsiegla, G.; Reverbel-Leroy, C.; Tardif, C.; Belaich, J. P.; Driguez, H.; Haser, R. Crystal structures of the cellulase Cel148F in complex with inhibitors and substrates give insights into its processive action. *Biochemistry* **2000**, *39*, 11238−11246.
(17) Attschul, S. F.; Wootton, J. C.; Gertz, E. M. Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* **2005**, *272*, 5101−5109.
(18) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714−2723.
(19) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(20) Brooks, B. R.; Brooks, C. L.; A.D. MacKerell, J.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(21) 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.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER*;University of California: San Francisco, CA, USA, 2012.
(22) MacKerell, A. D.; Feig, M.; Brooks, C. L. Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(23) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. Additive Empirical Force Field for Hexopyranose Monosaccharides. *J. Comput. Chem.* **2008**, *29*, 2543−2564.
(24) 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*, 926−935.
(25) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. Solvent-Induced Forces between Two Hydrophilic Groups. *J. Phys. Chem.* **1994**, *98*, 2198−2202.

(26) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithms for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(27) van Gunsteren, W. F.; Berendsen, H. J. C. Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Mol. Phys.* **1977**, *34*, 1311−1327.
(28) Crowley, M. F.; Williamson, M. J.; Walker, R. C. CHAMBER: Comprehensive support for CHARMM force fields within the AMBER software. *Int. J. Quantum Chem.* **2009**, *109*, 3767−3772.
(29) Widom, B. Some Topics in the Theory of Fluids. *J. Chem. Phys.* **1963**, *39*, 2808−2812.
(30) Widom, B. Potential-Distribution Theory and the Statistical Mechanics of Fluids. *J. Phys. Chem.* **1982**, *86*, 869−872.
(31) Bruckner, S.; Boresch, S. Efficiency of Alchemical Free Energy Simulations. II. Improvements for Thermodynamic Integration. *J. Comput. Chem.* **2011**, *32*, 1320−1333.
(32) Pratt, L. R.; Chandler, D. Theory of the Hydrophobic Effect. *J. Chem. Phys.* **1977**, *67*, 3683−3703.
(33) Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437*, 640−647.
(34) Garde, S.; Hummer, G.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. Origin of entropy convergence in hydrophobic hydration and protein folding. *Phys. Rev. Lett.* **1996**, *77*, 4966−4968.
(35) Hong, H.; Park, S.; Jiménez, R. H. F.; Rinehart, D.; Tamm, L. K. Role of Aromatic Side Chains in the Folding and Thermodynamic Stability of Integral Membrane Proteins. *J. Am. Chem. Soc.* **2007**, *129*, 8320−8327.