

Probing Carbohydrate Product Expulsion from a Processive Cellulase with Multiple Absolute Binding Free Energy Methods*[§]

Received for publication, December 13, 2010, and in revised form, March 4, 2011. Published, JBC Papers in Press, March 24, 2011, DOI 10.1074/jbc.M110.212076

Lintao Bu^{†1}, Gregg T. Beckham^{‡§}, Michael R. Shirts^{||}, Mark R. Nimlos[‡], William S. Adney^{**}, Michael E. Himmel^{**}, and Michael F. Crowley^{**2}

From the [†]National Bioenergy Center and the ^{**}Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado 80401, the [§]Department of Chemical Engineering, Colorado School of Mines, Golden, Colorado 80401, and the ^{||}Department of Chemical Engineering, University of Virginia, Charlottesville, Virginia 22904

Understanding the enzymatic mechanism that cellulases employ to degrade cellulose is critical to efforts to efficiently utilize plant biomass as a sustainable energy resource. A key component of cellulase action on cellulose is product inhibition from monosaccharide and disaccharides in the product site of cellulase tunnel. The absolute binding free energy of cellobiose and glucose to the product site of the catalytic tunnel of the Family 7 cellobiohydrolase (Cel7A) of *Trichoderma reesei* (*Hypocrea jecorina*) was calculated using two different approaches: steered molecular dynamics (SMD) simulations and alchemical free energy perturbation molecular dynamics (FEP/MD) simulations. For the SMD approach, three methods based on Jarzynski's equality were used to construct the potential of mean force from multiple pulling trajectories. The calculated binding free energies, -14.4 kcal/mol using SMD and -11.2 kcal/mol using FEP/MD, are in good qualitative agreement. Analysis of the SMD pulling trajectories suggests that several protein residues (Arg-251, Asp-259, Asp-262, Trp-376, and Tyr-381) play key roles in cellobiose and glucose binding to the catalytic tunnel. Five mutations (R251A, D259A, D262A, W376A, and Y381A) were made computationally to measure the changes in free energy during the product expulsion process. The absolute binding free energies of cellobiose to the catalytic tunnel of these five mutants are -13.1 , -6.0 , -11.5 , -7.5 , and -8.8 kcal/mol, respectively. The results demonstrated that all of the mutants tested can lower the binding free energy of cellobiose, which provides potential applications in engineering the enzyme to accelerate the product expulsion process and improve the efficiency of biomass conversion.

Lignocellulosic biomass is considered to be a sustainable and renewable energy resource that can provide liquid transportation fuels in the near term. The current biochemical conversion process consists of sequential steps of thermo-chemical pretreatment, enzymatic saccharification, and fermentation (1). However, the selective deconstruction of biomass to simple sugars remains costly because of biomass recalcitrance. Because the high cost of cellulose-degrading enzymes is one of the major cost factors in this process, deeper understanding of the mechanisms of enzymatic degradation of cellulose is critical for commercialization of the technology (2–4).

Among the many families of enzymes contributing to cellulose deconstruction, the Family 7 cellobiohydrolase (Cel7A) from the filamentous fungus *Trichoderma reesei* (*Hypocrea jecorina*) provides the majority of the hydrolytic potential for cellulose conversion in *T. reesei* enzyme cocktails (5, 6). Cel7A contains a catalytic domain and a carbohydrate-binding module separated by a glycosylated linker peptide (7, 8). The carbohydrate-binding module of Cel7A binds to the cellulose surface, thus increasing the surface concentration of enzymatic active sites for catalysis. Despite intensive research efforts during the last two decades, the molecular level, mechanistic details of the *T. reesei* Cel7A action on crystalline cellulose remain unclear. Recently, our group has suggested several new functions of the carbohydrate-binding module (9–11) and the linker (12) of Cel7A using computational approaches, which in general can offer insights into cellulase action (13, 14).

Cel7A is hypothesized to acquire a single cellulose surface chain from the reducing end, where the enzyme decrystallizes the cellodextrin chain from the cellulose surface and guides the chain into the active site tunnel located in the catalytic domain where the cellulose chain is hydrolyzed to cellobiose (13, 15, 16). Once the cellulose chain is in position for reaction, the processive mechanism follows a three-step process to continue releasing cellobiose units from the bound cellulose chain: 1) the leading cellobiose unit is cleaved by hydrolysis, 2) the cellobiose unit leaves the tunnel (product expulsion), and 3) the cellulose chain advances into the tunnel by a cellobiose unit to the reaction position.

Understanding the processive cycle of Cel7A is essential for conducting a complete characterization of cellulase function. In this work, we aim to understand the thermodynamic properties of the product expulsion step of the three-step process,

* This work was supported by the Department of Energy Office of the Biomass Program and in part by the National Science Foundation through TeraGrid resources provided by Ranger at Texas Advanced Computing Center and Athena at the National Institute for Computational Sciences under Grant TG-MCB090159, as well as by the National Renewable Energy Laboratory Computational Science Center, which is supported by the Office of Energy Efficiency and Renewable Energy of the United States Department of Energy under Contract DE-AC36-08GO28308.

⌘ Author's Choice—Final version full access.

§ The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1–S6.

[†] To whom correspondence may be addressed. Fax: 303-384-6363; E-mail: Lintao.Bu@nrel.gov.

² To whom correspondence may be addressed. Fax: 303-384-7752; E-mail: Michael.Crowley@nrel.gov.

Product Expulsion Free Energy of Cel7A

because it significantly affects the efficiency of the conversion of lignocellulosic biomass to fermentable sugars (17, 18) and constitutes a major obstacle for achieving high glucose yield (19, 20). Experimental studies have been conducted to measure the product inhibition constants for cellulase hydrolysis of soluble substrates (21–23) as well as crystalline cellulose (24–29). However, the reported inhibition constants range over several orders of magnitude depending on the experimental conditions, e.g. pH, temperature, and substrate types, thus making it difficult to determine a reliable value. It has been reported that cellobiose inhibits the cellulase activity strongly, with the inhibition constant (K_i) ranging between 0.01 and 6 g/liter, whereas glucose inhibits the cellulase activity (other than β -glucosidase activity) with K_i varying from 0.1 to 70 g/liter (19). To quantify and clarify the product inhibition effect, computational methods can be used as an alternative method to estimate the absolute binding free energy of cellobiose and glucose to the Cel7A catalytic tunnel.

Calculating the binding free energy of the protein-ligand complex is a fundamentally important problem in computational biology and chemistry (30–33). The binding free energy can be calculated using either equilibrium methods such as “alchemical” approaches (34) or by nonequilibrium methods such as steered molecular dynamics (SMD).³ The method of SMD is based on applying a guiding force to a system with the intention of moving the system over barriers or through slowly moving phase space in a time frame that is normally inaccessible, i.e. too long, without the extra force (35, 36). Usually, the force is applied unidirectionally to atoms or centers of mass. The purpose of the biasing force is to quickly explore possible paths between states with the intent of eliminating unlikely paths and exploring promising paths more precisely later. The energetics of the ligands unbinding from the proteins can be investigated from a large number of unbinding SMD trajectories via Jarzynski’s equality (37–39). This nonequilibrium method based on Jarzynski’s equality has been used to investigate the free energy changes in some simple systems, as well as complex biological systems (40–43). The results converge to the true value of the free energy in the limit of a large number of samples, although no gains in efficiency have been found when compared with the traditional methods, such as free energy perturbation (37, 44, 45).

Alchemical free energy perturbation techniques combined with molecular dynamics (FEP/MD) simulations are another set of powerful and commonly used approaches to estimate protein-ligand binding free energies (34, 46–49). In the alchemical double-decoupling method (50), the interactions of the ligand with its surroundings (bulk or binding site) are progressively turned off, during which various restraining potentials (51–55) applied to the ligand translational, rotational, and conformational degrees of the freedom may be switched on and off to prevent large excursions of the conformations of the system. These restraints help to reduce the size of the conformational space that needs to be sampled, in most cases reducing

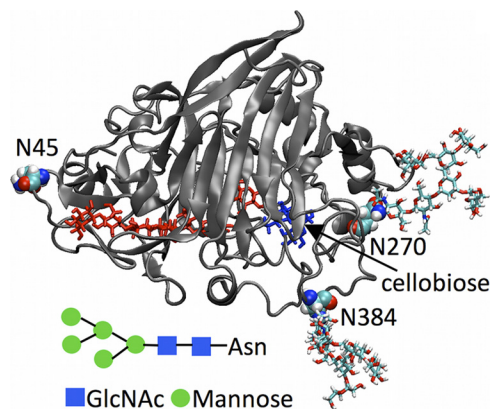


FIGURE 1. Structure of the catalytic domain of Cel7A with a cellobiose chain (containing seven glucose units, shown in red stick model) bound in the tunnel. The cellobiose unit is shown in blue stick model. The nine glucose units are labeled from 1 to 9 starting at the exit of the tunnel. The *N*-linked glycans with the pattern of $\text{Man}_5\text{GlcNAc}_2$ attached at two sites (Asn-270 (N270) and Asn-384 (N384)) are also shown.

the correlation time of samples collected from the system and thus increasing the statistical efficiency of the calculation (56–58).

In this work, we use both SMD and FEP/MD approaches to calculate the absolute binding free energy of cellobiose to the catalytic domain of Cel7A. To our best knowledge, we are the first to characterize the binding free energy of cellobiose and glucose to the catalytic domain of Cel7A. This characterization is crucial to understanding the mechanism of the *T. reesei* Cel7A action on crystalline cellulose at the molecular level. The approaches used in this study can also be applied to other cellulases to investigate the cellulose and glucose binding affinity to further our understanding of the product expulsion process in general.

EXPERIMENTAL PROCEDURES

Structure Preparation—The atomistic model of the catalytic domain of Cel7A was constructed using the reported crystal structure (7cel) (59, 60). Three *N*-linked sites (Asn-45, Asn-270, and Asn-384) of *T. reesei* Cel7A have been identified experimentally (61, 62). However, the glycosylation patterns vary considerably depending on the strains and their growth conditions (63). The $(\text{Man}_{1-2}\text{P})_{0-2}\text{Man}_{5-7}\text{GlcNAc}_2$ moieties have been observed as the predominant *N*-linked glycans for the wild type Cel7A strains and other mutant strains using a combination of chromatographic, electrophoretic, and mass spectrometric methods (64). In this work, a typical branched α -linked high mannose type structure with the pattern of $\text{Man}_5\text{GlcNAc}_2$ was chosen to represent the glycosylating oligosaccharides of the catalytic domain at Asn-270 and Asn-384 sites, as shown in Fig. 1. The Asn-45 site was not glycosylated in this study, because it is located at the tunnel entrance, which is ~ 50 Å away from the tunnel exit.

The CHARMM22 parameters (65, 66) with the CMAP (67–69) correction were used for the protein, and the C35 carbohydrate force field parameters (70, 71) were used for the cellobiose chain and glycosylation to generate the system via the CHARMM-GUI web server (72). The protein was placed in an equilibrated truncated octahedral box of TIP3P water mole-

³ The abbreviations used are: SMD, steered molecular dynamics; FEP, free energy perturbation; MD, molecular dynamics; PMF, potential of mean force; RMS, root mean square.

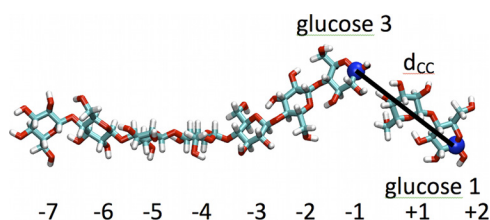


FIGURE 2. The structure of the celldextrin chain at the binding site. The two C1 atoms from glucose 1 and glucose 3 are shown as blue spheres.

cules with initial dimension of $103.0 \text{ \AA} \times 103.0 \text{ \AA} \times 103.0 \text{ \AA}$. Water molecules that overlapped with the cellulose heavy atoms were removed. To produce a neutral system, 19 Na^+ ions were introduced by transforming the water molecules into the ions randomly. Particle mesh Ewald summation (73) was used for electrostatics with a sixth order b-spline interpolation and a Gaussian distribution with a width of 0.34 \AA . The number of fast Fourier transform grid points for the charge mesh was 120 for each direction, and the nonbonded interaction cutoff was 12 \AA . The covalent bonds to hydrogen atoms were fixed using the SHAKE algorithm (74). The system was relaxed in the *NPT* ensemble at 300 K with a Nosé-Hoover thermostat and 1 atm for 100 ps with a step size of 1 fs, resulting in final box dimensions of $100.7 \text{ \AA} \times 100.7 \text{ \AA} \times 100.7 \text{ \AA}$.

SMD—CHAMBER (75) was used to convert the CHARMM protein structure file, coordinate file, and associated force field files to an AMBER topology file and coordinate file. The PMEMD engine in AMBER (76) was used to conduct molecular dynamics (MD) simulation to improve both the serial and parallel efficiency of the MD simulations. We note that all of the calculations conducted in this work used the CHARMM force field. Using PMEMD, the system was simulated in the *NVT* ensemble for 10 ns at 300 K with a step size of 2 fs. Subsequently, 40 starting structures were selected from the last 8 ns of the 10-ns simulation as the initial structures of 40 SMD simulations. During the SMD simulations, the distances between the C1 atoms of glucose 1 and glucose 3, shown in Fig. 2, were gradually increased with a speed of 2 \AA/ns , resulting in the total pulling distance of 14 \AA in 7 ns for the cellobiose unit in each starting structure. We chose the most appropriate pulling velocity by progressively reducing it from 1 \AA/ps to 1 \AA/ns until a flat PMF shape was observed after the cellobiose is out of the tunnel. We choose the reaction coordinate to be $\zeta = d_{\text{CC}} - d_{\text{CC}}(0)$, where d_{CC} is the distance between C1 of glucose 1 and C1 of glucose 3, and $d_{\text{CC}}(0)$ is the average distance over the 8 ns at the equilibrium state.

Construction of PMF—Three different methods based on Jarzynski's equality were used to construct the PMF. The Jarzynski's equality states that the free energy difference between two states A and B (differing in their values of the reaction coordinate) can be calculated as $\Delta G = -kT \ln(\langle \exp(-W/kT) \rangle)$, where W is the nonequilibrium work performed to drive the system from state A to state B. Thus by computing the work to drive the system between the two states and averaging the work exponentially, equilibrium free energies can be extracted from nonequilibrium calculations.

It is well known that the free energy calculated using Jarzynski's estimator is dominated by the small work values (77). To

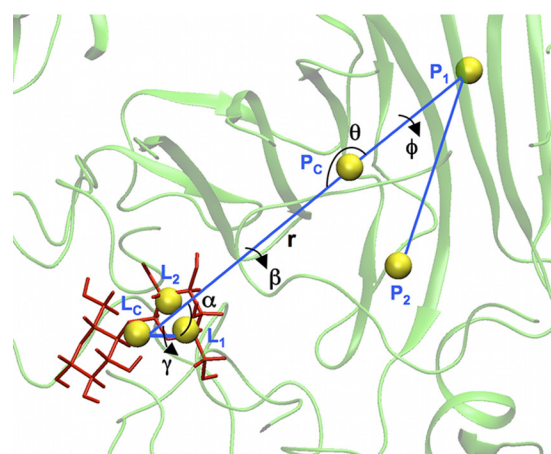


FIGURE 3. Translational and rotational restraints on cellobiose. Three positions in protein (P_c , P_1 , and P_2) and three positions in cellobiose (L_c , L_1 , and L_2) were chosen to set up the restraints. The positions in protein are the center-of-mass of the protein (P_c) and the centers of mass of residue Thr-417 (P_1) and residue Trp-367 (P_2). The positions in ligand are the center-of-mass of the cellobiose (L_c), atoms C4, C5, and O5 of the second glucose residue (L_1) and atoms C1, C2, and C3 of the second glucose residue (L_2). These six positions are shown as yellow spheres with solid blue lines connecting them. For the purpose of clarity, positions L_1 and L_2 are not connected.

solve this problem, two extrapolation methods, the linear extrapolation method and the cumulative integral extrapolation method, were derived to improve the statistics of work distribution from limited sampling (78). These two extrapolation methods were also used in the analyses.

FEP/MD—We also calculated the absolute binding free energy of cellobiose using FEP/MD simulations. Following the alchemical double-decoupling method of Roux and co-workers (56, 58, 79), the binding process of cellobiose to the Cel7A catalytic domain is decomposed into eight steps: 1) a conformational restraint is applied to the cellobiose in the bulk solution; 2) the interaction of the cellobiose with the bulk solution is turned off; 3) a translational restraint is applied to the cellobiose to maintain its relative position at the binding site; 4) a rotational restraint is applied to the cellobiose to maintain its orientation at the binding site; 5) the interactions of the cellobiose with the binding site are turned on; 6) the translational restraint is released; 7) the rotational restraint is released; and 8) the conformational restraint is released. The absolute binding free energy is given by the following,

$$\Delta G_{\text{bind}}^0 = \Delta G_{\text{int}}^{\text{bulk}} - \Delta G_{\text{int}}^{\text{site}} + \Delta \Delta G_t + \Delta \Delta G_r + \Delta \Delta G_c \quad (\text{Eq. 1})$$

where $\Delta G_{\text{int}}^{\text{bulk}}$ and $\Delta G_{\text{int}}^{\text{site}}$ correspond to the free energy cost arising from turning off the interactions of the cellobiose with its surrounding bulk and the binding site (steps 2 and 5), respectively, and $\Delta \Delta G_t$, $\Delta \Delta G_r$, and $\Delta \Delta G_c$ correspond to the translational (steps 3 and 6), rotational (steps 4 and 7), and conformational (steps 1 and 8) restrictions of the ligand upon binding, respectively.

Restraining Potentials—Fig. 3 illustrates the restraining potential definitions on the cellobiose ligand used in the FEP/MD approach. Six point positions were chosen to define the position and orientation of the ligand relative to the protein (58): three in the protein (P_c , P_1 , and P_2) and three in the ligand (L_c ,

Product Expulsion Free Energy of Cel7A

L_1 , and L_2). The translational restraining potential is defined as follows,

$$u_t = \frac{1}{2}[k_t(r_L - r_0)^2 + k_a(\theta_L - \theta_0)^2 + k_a(\phi_L - \phi_0)^2] \quad (\text{Eq. 2})$$

where r_L is the distance between the center-of-mass of protein (P_c) and the center-of-mass of ligand (L_c); θ_L is the angle $P_1-P_c-L_c$; ϕ_L is the dihedral angle $P_2-P_1-P_c-L_c$; k_t and k_a are the force constants; r_0 , θ_0 , and ϕ_0 are the average values of the fully interacting ligand in the binding site from the 8 ns of equilibrium run.

Similarly, the rotational restraining potential is defined as follows,

$$u_r = \frac{1}{2}[k_a(\alpha_L - \alpha_0)^2 + k_a(\beta_L - \beta_0)^2 + k_a(\gamma_L - \gamma_0)^2] \quad (\text{Eq. 3})$$

where the angle α_L , the torsional angle β_L , and the torsional angle γ_L are defined as $P_c-L_c-L_1$, $P_1-P_c-L_c-L_1$, and $P_c-L_c-L_1-L_2$, respectively; k_a is the force constant; α_0 , β_0 , and γ_0 are the average values of the fully interacting ligand in the binding site. The conformational restraining potential is defined as follows,

$$u_c = \frac{1}{2}[k_a(\Phi_L - \Phi_0)^2 + k_a(\Psi_L - \Psi_0)^2] \quad (\text{Eq. 4})$$

where the torsional angles Φ_L ($H_1-C_1-O-C_4'$) and Ψ_L ($H_4'-C_4'-O-C_1$) are along the glycosidic bond of the cellobiose ligand, k_a is the force constant, and Φ_0 and Ψ_0 are the average values. The reference values of the eight internal coordinates (r_0 , θ_0 , ϕ_0 , α_0 , β_0 , γ_0 , Φ_0 , and Ψ_0) were calculated from an unbiased simulation of the Cel7A-cellobiose complex for 10 ns. The last 8-ns trajectory of the unbiased simulation was used to calculate the averages and the root mean square (RMS) fluctuations of these eight internal coordinates. Subsequently, the corresponding force constants in each equation were estimated as $k_x \approx k_B T / \langle \Delta x^2 \rangle$ (58), where Δx is the RMS fluctuation of the associated internal coordinate.

Protocol for FEP/MD— $\Delta G_{\text{int}}^{\text{bulk}}$ and $\Delta G_{\text{int}}^{\text{site}}$ were calculated using FEP/MD simulations. The Lennard-Jones potential was separated into repulsive and dispersive free energy using the Weeks-Chandler-Andersen method (80, 81). A linear coupling parameter λ was used to control the electrostatic interactions with 11 windows ($\lambda = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$, and 1.0) and the dispersive interactions with five windows ($\lambda = 0, 0.25, 0.5, 0.75$, and 1.0). The repulsive interactions were calculated using 18 windows for cellobiose at the binding site and 26 windows for cellobiose at the bulk solution. In the context of a typical decoupling process, the electrostatic interactions were turned off first, then the dispersive interactions were turned off, and finally the repulsive interactions were turned off. The system was simulated using Langevin dynamics for 560 ps for the cellobiose in the binding site, with the last 500-ps simulation being used to calculate the free energy for each window. For cellobiose in bulk solution, 1 ns of Langevin dynamics was generated, and the last 800-ps simulation was used to calculate the free energy.

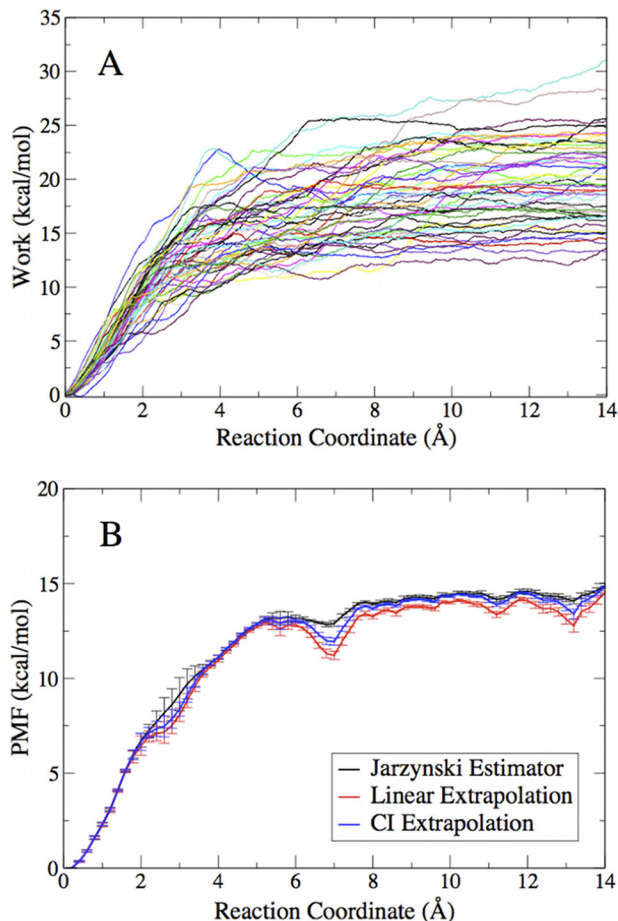


FIGURE 4. A, collection of 40 work profiles along the pulling distance for cellobiose unbinding from Cel7A catalytic tunnel. B, PMF profiles calculated by using Jarzynski estimator and linear and cumulative integral (CI) extrapolation estimators.

The free energy changes corresponding to the translational (step 7), rotational (step 6), and conformational (steps 1 and 8) restraints with the ligand in the binding site (steps 6–8) or in the surrounding bulk solutions (step 1) were calculated using FEP/MD simulations. The restraining potentials were gradually turned off via the linear coupling parameters with 10 windows. The free energy changes corresponding to the translational (step 3) and rotational (step 4) restraints with the ligand in the gas phase (when decoupled from its surrounding bulk solution) were calculated analytically (58).

RESULTS AND DISCUSSION

PMF for Cellobiose Expulsion from Cel7A—Fig. 4A illustrates the value of accumulated work as a function of the reaction coordinate for the 40 SMD simulations of cellobiose expulsion from the catalytic tunnel of Cel7A. The averaged work is 20.3 ± 4.5 kcal/mol. In general, direct use of Jarzynski's equality requires the variance in the work distribution within a few kT . However, this usually cannot be satisfied when studying large biomolecular systems.

Some studies have claimed that certain extrapolation methods provide more reliable free energy estimates than the direct application of Jarzynski's equality (78). Fig. 4B shows the PMF profiles calculated from 40 pulling trajectories using three esti-

TABLE 1
Force constants for the restraining potentials

	Distance	Angle	
		Translation and rotation	Conformation
Δ_x	0.3 Å	1–8°	15°
k_x	5 kcal/mol/Å ²	200 kcal/mol/rad ²	100 kcal/mol/rad ²

TABLE 2
Free energy components for cellobiose binding

The data are shown as kcal/mol.

$-\Delta G_{\text{bulk}}$	ΔG_{site}	$\Delta\Delta G_t$	$\Delta\Delta G_r$	$\Delta\Delta G_c$	$\Delta G_{\text{binding}}$
40.3 ± 0.5	-58.8 ± 0.6	3.2 ± 0.2	3.1 ± 0.2	1.0 ± 0.3	-11.2 ± 0.6

mators. The calculated free energy values are 14.4 ± 0.1 , 14.1 ± 0.2 , and 14.4 ± 0.1 kcal/mol using Jarzynski's estimator, the linear extrapolation estimator, and the cumulative integral extrapolation estimator, respectively. The uncertainty of each estimator was measured using the bootstrap method (82). Our results show that the free energy values derived from different methods are in excellent agreement with each other. However, because the extrapolation methods usually produce a more rugged PMF compared with the Jarzynski's estimator (77), the Jarzynski's estimator was used to construct the PMFs of Cel7A mutants in this work. We note that the SMD simulations were conducted for the unbinding process. Therefore, the binding free energy should be -14.4 kcal/mol, suggesting that cellobiose is more stable in the Cel7A tunnel than in the bulk solution, which is consistent with the experimental observations that cellobiose can inhibit the Cel7A hydrolysis rate (19).

PMF-based approaches for calculating binding free energy usually become less practical if the binding site is deeply buried inside a protein, and a simple path for ligand association cannot be clearly defined *a priori*. However, because the direction of the external force in this study is adjusted instantaneously based on the current conformation of the protein and the cellobiose, our approach is more effective compared with other studies applying unidirectional pulling forces.

Restraining Potentials—The fluctuations of the eight relative coordinates (r , θ , ϕ , α , β , γ , Φ , and Ψ) were obtained from the 10-ns unbiased equilibrium simulation of the fully interacting ligand in the binding site (shown in supplemental Fig. S1). These relative coordinates were used to define the restraining potentials in the FEP/MD approach. The RMS fluctuation in the distance r was ~ 0.3 Å, whereas the RMS fluctuations of the angles (θ , ϕ , α , β , and γ) were ~ 1 – 8° for the translational and rotational restraining potentials and $\sim 15^\circ$ (Φ and Ψ) for the conformational restraining potentials. Therefore, the force constants for the distance and angle restraint were chosen as 5 kcal/mol/Å², 200 kcal/mol/rad², and 100 kcal/mol/rad², respectively, as shown in Table 1.

Binding Free Energy—The results of the calculations using FEP/MD simulations are shown in Table 2. The error was estimated by using different fractions of the overall time series data collected during the FEP/MD simulations. The binding free energy was shown to be -11.2 ± 0.6 kcal/mol, which is in reasonably good agreement with the SMD result, -14.4 ± 0.1 kcal/mol. The difference between these two values is likely because the final unbound state in the SMD simulations does not rigor-

ously correspond to the standard state. The cellobiose is still restrained by the steering potential, which prevents it from occupying the volume associated with the standard concentration. Supplemental Fig. S2 shows the decomposition of the binding free energy. The free energy for decoupling the non-bonded terms in the potential for the cellobiose in the binding site was 18.5 ± 0.6 kcal/mol more favorable than that of the cellobiose in the bulk solution. By component, the electrostatic free energy contributed -8.3 ± 0.3 kcal/mol, and the van der Waals free energy contributed -10.2 ± 0.6 kcal/mol. The net contributions from the translational, rotational, and conformational free energy were 3.2, 3.1, and 1.0 kcal/mol, respectively.

Effect of Glycosylation—Protein glycosylation is the covalent linkage of carbohydrates to the protein asparagine (*N*-linked) or serine/threonine (*O*-linked) residues. Glycosylation is an important post-translational modification of many proteins, which is thought to serve multiple functions for secretion, signaling, and stability (83, 84). *N*-Linked glycans are typically bonded to the GlcNAc residue, which in turn is bonded to the protein asparagine residue. Studies have shown that expression host and growth conditions change glycosylation patterns, which can affect enzyme activity significantly (61, 63, 64, 85, 87, 88). We concluded in our previous studies that *N*-linked glycans in peptide loops that form part of the Cel7A tunnel have the greatest impact on both thermal stability and enzymatic activity on crystalline cellulose for both the *T. reesei* and *Penicillium funiculosum* Cel7A enzymes (89).

The question of whether the *N*-linked glycans can affect the product expulsion process can be directly addressed by simulation. Because the cellobiose is leaving the catalytic tunnel, it is possible that it can associate with the surrounding *N*-linked glycans located near the tunnel exit. As demonstrated in Fig. 5, there was a significant overlap between the sampled volumes of the glycan linked to the protein residue Asn-384 and the cellobiose during the 40 SMD simulations. At first glance, this seems to suggest that cellobiose could interact directly with the glycan linked to residue Asn-384 when it is expelled from the catalytic tunnel.

However, further detailed analysis of the interaction energy between the glycan linked to residue Asn-384 and the cellobiose unit demonstrated that the direct interaction is small. Interaction between the cellobiose and the Asn-384 glycan was observed only in 7 of the 40 SMD trajectories (supplemental Fig. S3). In most of the pulling simulations, *N*-linked glycans did not interact directly with the cellobiose, suggesting they do not actually block the tunnel and therefore the product expulsion pathway. We note that there was no correlation between the SMD trajectories with lower work values and the ones that interacted with the glycans. For the seven SMD trajectories that interacted with the glycans, four of them showed work values above 20.3 kcal/mol (the average work value of 40 SMD trajectories), whereas three of them showed work values below the average work value.

The observation that the *N*-linked glycans do not directly obstruct the expulsion site does not rule out the possibility that these glycans can affect the binding mode of cellobiose in the Cel7A tunnel by modifying the flexibility of the loop to which it is attached. This second hypothesis is supported by comparison

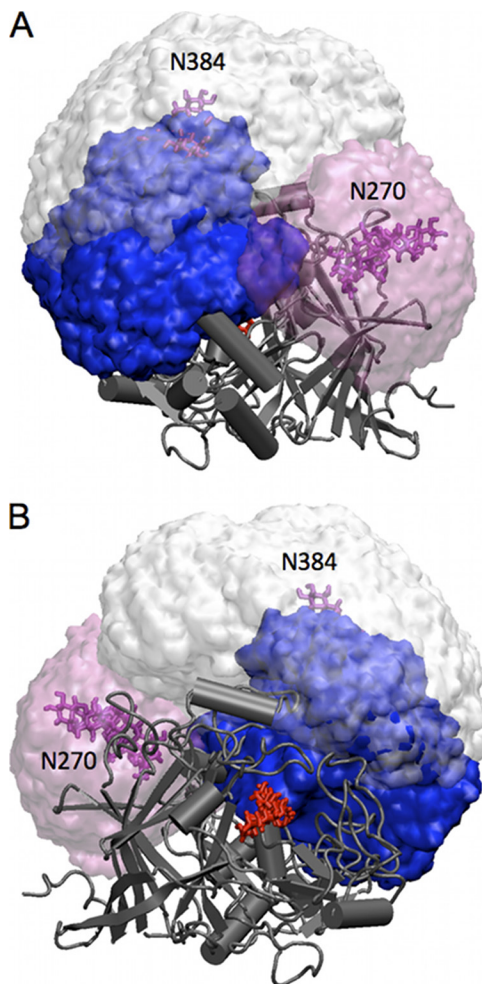


FIGURE 5. The accessible volume of cellobiose (solid blue) and two N-linked glycans (transparent mauve or white) attached on Asn-270 (N270) and Asn-384 (N384). A, top view from tunnel exit down to the entrance. B, top view from tunnel entrance down to the exit. The cellobiose chain bound in the tunnel is shown as a red stick model, and the two N-linked glycans are shown as a purple stick model.

of the RMS fluctuation of the Cel7A residues with and without the glycans linked to the protein. The two RMS fluctuation profiles were almost identical except for four loops. The loop region (from residues 381 to 394) at the exit end of the tunnel showed a smaller fluctuation when the glycans are linked (supplemental Fig. S4), indicating that the glycans linked to Asn-384 appear to stabilize the exit tunnel structure. Specifically, the RMS fluctuation for residue Asn-384 was ~ 1 Å when the glycans were attached and ~ 2 Å when the glycans were not attached. The other three loops that did not form the exit end of the tunnel showed moderately larger fluctuation when the glycans are attached. Because these loops are quite flexible and disordered, we assumed that there would be no allosteric effects caused by long distance fluctuations, which will be examined further in a future study.

The results of the 10-ns equilibrium simulation of the Cel7A catalytic domain without the N-linked glycans attached at Asn-384 show a larger fluctuation of the relative position of cellobiose to the exit tunnel, which makes the initial binding state difficult to characterize. Thus it is difficult to conduct SMD simulations to directly measure the free energy change associ-

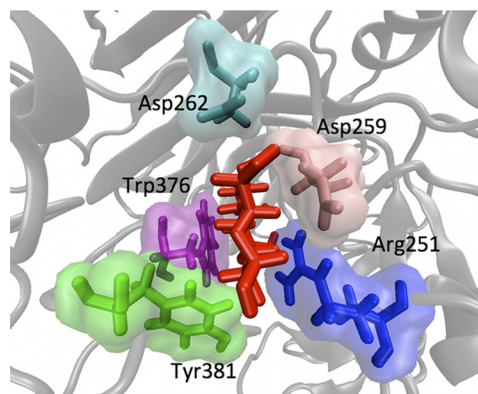


FIGURE 6. The structure of the Cel7A catalytic tunnel at the exit site (top view from tunnel exit down to the entrance). The five key binding residues are highlighted and labeled. The cellobiose unit is shown as a red stick model.

ated without glycosylation. This problem will be addressed in a future study.

PMF for Cel7A Mutants—Identifying which residues interact strongly with the cellobiose in the catalytic tunnel is essential to our understanding of product inhibition. Knowledge of these sites can provide potential opportunities to engineer the protein to lower the free energy barrier for product expulsion, thus speeding the product expulsion process. The interaction energy between each residue along the tunnel and the cellobiose as a function of the pulling distance was calculated to identify several candidates for further analysis (supplemental Fig. S5). The hypothesis is that if the interaction energy between the cellobiose and a protein residue is significant (defined as at least ~ 5 kcal/mol when cellobiose is inside the tunnel) along a pulling trajectory, this residue may contribute substantially to the binding free energy. In other words, we chose the residues that show the strongest interactions with the cellobiose during the pulling simulations as the candidates. Based on the criteria, five residues, Arg-251, Asp-259, Asp-262, Trp-376, and Tyr-381, are chosen as the candidates for computational point mutation studies. The positions of these five residues are shown in Fig. 6.

SMD simulations were conducted on the five Cel7A mutants (R251A, D259A, D262A, W376A, and Y381A) to characterize the absolute binding free energy of cellobiose. Fig. 7 shows the PMFs of cellobiose unbinding from the catalytic tunnel of WT Cel7A and five mutants. The original 40 work profiles for each mutant are shown in supplemental Fig. S6. Using Jarzynski's estimator, the binding free energy of cellobiose to the catalytic tunnel of five Cel7A mutants is shown in Table 3. The uncertainty ranges from 0.2 to 0.6 kcal/mol. The results suggest that all of these mutants can decrease the binding free energy of cellobiose. These results are useful in guiding experimental design of Cel7A mutants to expedite the product expulsion process. Whether the five mutants identified in this study can speed up the product expulsion process is currently under experimental investigation.

PMF for Glucose Expulsion—Experimental studies suggested that both cellobiose and glucose can decrease the cellulose hydrolysis rate and the product yields (17, 18). To investigate the binding affinity of glucose to the Cel7A catalytic tunnel relative to cellobiose, the absolute binding free energy of glucose to the Cel7A tunnel was calculated via SMD simulations.

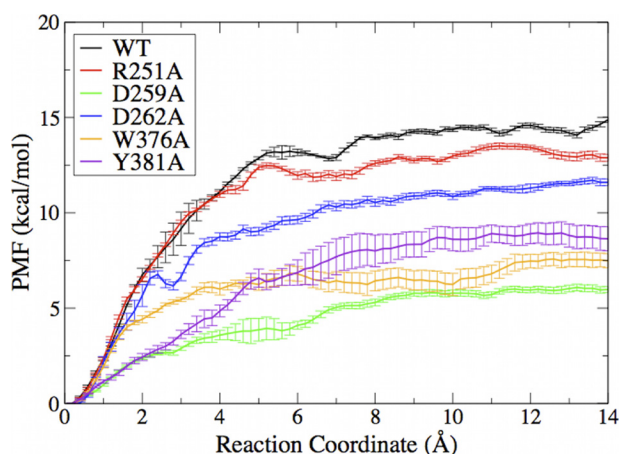


FIGURE 7. PMF profiles for cellobiose unbinding from the catalytic tunnel of WT Cel7A and five mutants.

TABLE 3
Binding free energy for Cel7A WT and mutants

The data are shown as kcal/mol.

WT	R251A	D259A	D262A	W376A	Y381A
-14.4 ± 0.1	-13.1 ± 0.2	-6.0 ± 0.2	-11.5 ± 0.2	-7.5 ± 0.4	-8.8 ± 0.6

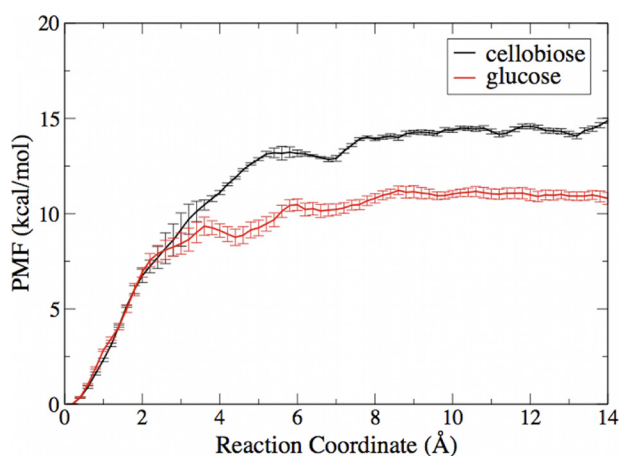


FIGURE 8. PMF profiles for cellobiose and glucose unbinding from the catalytic tunnel of Cel7A.

Glucose was initially positioned at the +1 site, as shown in Fig. 1, in the tunnel. The original 40 work profiles are shown in supplemental Fig. S6. As shown in Fig. 8, the absolute binding free energy of glucose unit was -10.9 kcal/mol using Jarzynski's estimator, indicating that the glucose is 3.5 kcal/mol less stable in the catalytic tunnel of Cel7A than cellobiose. This is consistent with the experimental observation that cellobiose strongly inhibits enzyme activity, whereas glucose inhibition is relatively marginal (26).

The absolute binding free energy of cellobiose and glucose characterized in this study furthers our understanding of the experimental studies on product inhibition and will aid in developing kinetic models of cellulase action on cellulose (86). We note that only the binding mode with the +1 site occupied by the glucose unit is investigated in this study. It is possible that other binding modes, *i.e.* binding at the +2 site (the position of the first glucose in cellobiose), may also contribute to the glucose binding to the Cel7A tunnel, but the binding of glucose to

the +1 site is expected to be the worst case scenario for glucose inhibiting the enzyme, because the glucose needs to diffuse across both the +1 and the +2 binding sites to be removed from the binding site.

Conclusions—We have characterized the absolute binding free energy of cellobiose to the catalytic tunnel of Cel7A using SMD simulations and FEP/MD simulations. The binding free energies derived from these two approaches are qualitatively consistent, suggesting that either the SMD or the FEP/MD method can be applied to investigate the binding affinities of cellobiose and glucose to other cellulases to further our understanding of product expulsion process in general.

From our simulations, we identified five targets for enzyme engineering of Cel7A that may decrease product inhibition: R251A, D259A, D262A, W376A, and Y381A. The PMF profiles of cellobiose unbinding from the catalytic tunnel of these five mutants have been constructed using SMD simulations. The calculated binding free energies of cellobiose to the Cel7A mutants are significantly less favorable than that of the WT Cel7A, suggesting that all of these mutants can accelerate product expulsion process, thus improving the efficiency of biomass conversion. Experimental studies are being conducted currently on the five mutants to examine our findings reported here.

The effect of the *N*-linked glycosylation on the product expulsion was also investigated. The *N*-linked glycans may affect the binding mode of cellobiose to the Cel7A tunnel by modifying the flexibility of the loop it is attached to rather than by direct interacting with the cellobiose. More detailed understanding of the effect of glycosylation on the function and activity of Cel7A at the molecular level requires better experimental characterization of the glycosylation pattern and structure.

Acknowledgments—We thank Larry Taylor and James Matthews for helpful discussions.

REFERENCES

1. Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J., Jr., Hallett, J. P., Leak, D. J., Liotta, C. L., Mielenz, J. R., Murphy, R., Templer, R., and Tschaplinski, T. (2006) *Science* **311**, 484–489
2. Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., and Foust, T. D. (2007) *Science* **315**, 804–807
3. Himmel, M. E., and Picataggio, S. K. (2008) *Biomass Recalcitrance: Deconstructing the Plant Cell Wall for Bioenergy*, pp. 1–6, Blackwell Publishing, London, UK
4. Himmel, M. E., Ruth, M. F., and Wyman, C. E. (1999) *Curr. Opin. Biotechnol.* **10**, 358–364
5. Barr, B. K., Hsieh, Y. L., Ganem, B., and Wilson, D. B. (1996) *Biochemistry* **35**, 586–592
6. Vršanská, M., and Biely, P. (1992) *Carbohydr. Res.* **227**, 19–27
7. Srisodsuk, M., Reinikainen, T., Penttilä, M., and Teeri, T. T. (1993) *J. Biol. Chem.* **268**, 20756–20761
8. Teeri, T. T., Lehtovaara, P., Kauppinen, S., Salovuori, I., and Knowles, J. (1987) *Gene* **51**, 43–52
9. Beckham, G. T., Matthews, J. F., Bomble, Y. J., Bu, L., Adney, W. S., Himmel, M. E., Nimlos, M. R., and Crowley, M. F. (2010) *J. Phys. Chem. B* **114**, 1447–1453
10. Bu, L., Beckham, G. T., Crowley, M. F., Chang, C. H., Matthews, J. F., Bomble, Y. J., Adney, W. S., Himmel, M. E., and Nimlos, M. R. (2009) *J. Phys. Chem. B* **113**, 10994–11002

11. Nimlos, M. R., Matthews, J. F., Crowley, M. F., Walker, R. C., Chukkapalli, G., Brady, J. W., Adney, W. S., Cleary, J. M., Zhong, L., and Himmel, M. E. (2007) *Protein Eng. Des. Sel.* **20**, 179–187
12. Beckham, G. T., Bomble, Y. J., Matthews, J. F., Taylor, C. B., Resch, M. G., Yarbrough, J. M., Decker, S. R., Bu, L., Zhao, X., McCabe, C., Wohler, J., Bergensträhle, M., Brady, J. W., Adney, W. S., Himmel, M. E., and Crowley, M. F. (2010) *Biophys. J.* **99**, 3773–3781
13. Beckham, G. T., Bomble, Y. J., Bayer, E. A., Himmel, M. E., and Crowley, M. F. (2011) *Curr. Opin. Biotechnol.*, 10.1016/j.copbio.2010.11.005
14. Chundawat, S., Beckham, G. T., Himmel, M. E., and Dale, B. E. (2011) *Annu. Rev. Chem. Biomol. Eng.*, 10.1146/annurev-chem-bioeng-061010-114205
15. Zhong, L., Matthews, J. F., Crowley, M. F., Rignall, T., Talon, C., Cleary, J. M., Walker, R. C., Chukkapalli, G., McCabe, C., Nimlos, M. R., Brooks, C. L., Himmel, M. E., and Brady, J. W. (2008) *Cellulose* **15**, 261–273
16. Zhong, L., Matthews, J. F., Hansen, P. I., Crowley, M. F., Cleary, J. M., Walker, R. C., Nimlos, M. R., Brooks, C. L., 3rd, Adney, W. S., Himmel, M. E., and Brady, J. W. (2009) *Carbohydr. Res.* **344**, 1984–1992
17. Gan, Q., Allen, S. J., and Taylor, G. (2003) *Process Biochem.* **38**, 1003–1018
18. Gusakov, A. V., Sinityn, A. P., and Klyosov, A. A. (1987) *Biotechnol. Bioeng.* **29**, 906–910
19. Andrić, P., Meyer, A. S., Jensen, P. A., and Dam-Johansen, K. (2010) *Biotechnol. Adv.* **28**, 308–324
20. Bansal, P., Hall, M., Realff, M. J., Lee, J. H., and Bommarius, A. S. (2009) *Biotechnol. Adv.* **27**, 833–848
21. Hsu, T. A., Gong, C. S., and Tsao, G. T. (1980) *Biotechnol. Bioeng.* **22**, 2305–2320
22. Nidetzky, B., Zachariae, W., Gercken, G., Hayn, M., and Steiner, W. (1994) *Enzyme Microb. Technol.* **16**, 43–52
23. von Ossowski, I., Ståhlberg, J., Koivula, A., Piens, K., Becker, D., Boer, H., Harle, R., Harris, M., Divne, C., Mahdi, S., Zhao, Y., Driguez, H., Claeysens, M., Sinnott, M. L., and Teeri, T. T. (2003) *J. Mol. Biol.* **333**, 817–829
24. Bezerra, R. M., and Dias, A. A. (2004) *Appl. Biochem. Biotechnol.* **112**, 173–184
25. Bezerra, R. M., and Dias, A. A. (2005) *Appl. Biochem. Biotechnol.* **126**, 49–59
26. Gruno, M., Våljamäe, P., Pettersson, G., and Johansson, G. (2004) *Biotechnol. Bioeng.* **86**, 503–511
27. Oh, K. K., Kim, S. W., Jeong, Y. S., and Hong, S. I. (2000) *Appl. Biochem. Biotechnol.* **89**, 15–30
28. Philippidis, G. P., Smith, T. K., and Wyman, C. E. (1993) *Biotechnol. Bioeng.* **41**, 846–853
29. Wald, S., Wilke, C. R., and Blanch, H. W. (1984) *Biotechnol. Bioeng.* **26**, 221–230
30. Beveridge, D. L., and DiCapua, F. M. (1989) *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431–492
31. Kollman, P. (1993) *Chem. Rev.* **93**, 2395–2417
32. Simonson, T., Archontis, G., and Karplus, M. (2002) *Acc. Chem. Res.* **35**, 430–437
33. Straatsma, T. P., and Mccammon, J. A. (1992) *Annu. Rev. Phys. Chem.* **43**, 407–435
34. Shirts, M. R., Mobley, D. L., and Chodera, J. D. (2007) *Annu. Rev. Comput. Chem.* **3**, 41–59
35. Israelowitz, B., Baudry, J., Gullingsrud, J., Kosztin, D., and Schulten, K. (2001) *J. Mol. Graphics Modell.* **19**, 13–25
36. Park, S., Khalili-Araghi, F., Tajkhorshid, E., and Schulten, K. (2003) *J. Chem. Phys.* **119**, 3559–3566
37. Hummer, G. (2001) *J. Chem. Phys.* **114**, 7330–7337
38. Hummer, G., and Szabo, A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3658–3661
39. Jarzynski, C. (1997) *Phys. Rev. Lett.* **78**, 2690–2693
40. Crespo, A., Martí, M. A., Estrin, D. A., and Roitberg, A. E. (2005) *J. Am. Chem. Soc.* **127**, 6940–6941
41. Cuendet, M. A., and Michielin, O. (2008) *Biophys. J.* **95**, 3575–3590
42. Jensen, M. Ø., Park, S., Tajkhorshid, E., and Schulten, K. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6731–6736
43. Martin, H. S., Jha, S., Howorka, S., and Coveney, P. V. (2009) *J. Chem. Theory Comput.* **5**, 2135–2148
44. Gore, J., Ritort, F., and Bustamante, C. (2003) *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12564–12569
45. Shirts, M. R., and Pande, V. S. (2005) *J. Chem. Phys.* **122**, 144107
46. Florian, J., Goodman, M. F., and Warshel, A. (2000) *J. Phys. Chem. B* **104**, 10092–10099
47. Jayachandran, G., Shirts, M. R., Park, S., and Pande, V. S. (2006) *J. Chem. Phys.* **125**, 084901
48. Lu, N. D., and Kofke, D. A. (2001) *J. Chem. Phys.* **115**, 6866–6875
49. Lu, N. D., and Kofke, D. A. (2001) *J. Chem. Phys.* **114**, 7303–7311
50. Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. (1997) *Biophys. J.* **72**, 1047–1069
51. Boresch, S., Tettinger, F., Leitgeb, M., and Karplus, M. (2003) *J. Phys. Chem. B* **107**, 9535–9551
52. Roux, B., Nina, M., Pomes, R., and Smith, J. C. (1996) *Biophys. J.* **71**, 670–681
53. Hermans, J., and Wang, L. (1997) *J. Am. Chem. Soc.* **119**, 2707–2714
54. Mobley, D. L., Chodera, J. D., and Dill, K. A. (2006) *J. Chem. Phys.* **25**, 084902
55. Woo, H. J., and Roux, B. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6825–6830
56. Deng, Y., and Roux, B. (2009) *J. Phys. Chem. B.* **113**, 2234–2246
57. Deng, Y. Q., and Roux, B. (2006) *J. Chem. Theory Comput.* **2**, 1255–1273
58. Wang, J., Deng, Y., and Roux, B. (2006) *Biophys. J.* **91**, 2798–2814
59. Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K., Teeri, T. T., and Jones, T. A. (1994) *Science* **265**, 524–528
60. Divne, C., Ståhlberg, J., Teeri, T. T., and Jones, T. A. (1998) *J. Mol. Biol.* **275**, 309–325
61. Harrison, M. J., Nouwens, A. S., Jardine, D. R., Zachara, N. E., Gooley, A. A., Nevalainen, H., and Packer, N. H. (1998) *Eur. J. Biochem.* **256**, 119–127
62. Klarskov, K., Piens, K., Ståhlberg, J., Høj, P. B., Beeumen, J. V., and Claeysens, M. (1997) *Carbohydr. Res.* **304**, 143–154
63. Stals, I., Sandra, K., Geysens, S., Contreras, R., Van Beeumen, J., and Claeysens, M. (2004) *Glycobiology* **14**, 713–724
64. Stals, I., Sandra, K., Devreese, B., Van Beeumen, J., and Claeysens, M. (2004) *Glycobiology* **14**, 725–737
65. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caffisch, 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., and Karplus, M. (2009) *J. Comput. Chem.* **30**, 1545–1614
66. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkerich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) *J. Phys. Chem. B* **102**, 3586–3616
67. Feig, M., MacKerell, A. D., and Brooks, C. L. (2003) *J. Phys. Chem. B* **107**, 2831–2836
68. Mackerell, A. D., Jr., Feig, M., and Brooks, C. L., 3rd (2004) *J. Comput. Chem.* **25**, 1400–1415
69. MacKerell, A. D., Jr., Feig, M., and Brooks, C. L., 3rd (2004) *J. Am. Chem. Soc.* **126**, 698–699
70. Guvench, O., Greene, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W., and Mackerell, A. D., Jr. (2008) *J. Comput. Chem.* **29**, 2543–2564
71. Guvench, O., Hatcher, E. R., Venable, R. M., Pastor, R. W., and Mackerell, A. D. (2009) *J. Chem. Theory Comput.* **5**, 2353–2370
72. Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008) *J. Comput. Chem.* **29**, 1859–1865
73. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) *J. Chem. Phys.* **103**, 8577–8593
74. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. (1977) *J. Comput. Phys.* **23**, 327–341
75. Crowley, M. F., Williamson, M. J., and Walker, R. C. (2009) *Int. J. Quan-*

- tum. *Chem.* **109**, 3767–3772
76. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) *J. Comput. Chem.* **26**, 1668–1688
77. Zhang, D., Gullingsrud, J., and McCammon, J. A. (2006) *J. Am. Chem. Soc.* **128**, 3019–3026
78. Ytreberg, F. M., and Zuckerman, D. M. (2004) *J. Comput. Chem.* **25**, 1749–1759
79. Deng, Y., and Roux, B. (2008) *J. Chem. Phys.* **128**, 115103
80. Deng, Y. Q., and Roux, B. (2004) *J. Phys. Chem. B* **108**, 16567–16576
81. Weeks, J. D., Chandler, D., and Andersen, H. C. (1971) *J. Chem. Phys.* **54**, 5237–5247
82. Efron, B., and Tibshirani, R. (1986) *Stat. Sci.* **1**, 54–77
83. Spiro, R. G. (2002) *Glycobiology* **12**, 43–56
84. Varki, A. (1993) *Glycobiology* **3**, 97–130
85. Godbole, S., Decker, S. R., Nieves, R. A., Adney, W. S., Vinzant, T. B., Baker, J. O., Thomas, S. R., and Himmel, M. E. (1999) *Biotechnol. Progress* **15**, 828–833
86. Levine, S. E., Fox, J. M., Blanch, H. W., and Clark, D. S. (2010) *Biotechnol. Bioeng.* **107**, 37–51
87. Hui, J. P., Lanthier, P., White, T. C., McHugh, S. G., Yaguchi, M., Roy, R., and Thibault, P. (2001) *J. Chromatogr. B* **752**, 349–368
88. Jeoh, T., Michener, W., Himmel, M. E., Decker, S. R., and Adney, W. S. (2008) *Biotechnol. Biofuels* **1**, 10
89. Adney, W. S., Jeoh, T., Beckham, G. T., Chou, Y. C., Baker, J. O., Michener, W., Brunecky, R., and Himmel, M. E. (2009) *Cellulose* **16**, 699–709