Initial Recognition of a Cellodextrin Chain in the Cellulose-Binding Tunnel May Affect Cellobiohydrolase Directional Specificity

Pavan K. GhattyVenkataKrishna,[†‡] Emal M. Alekozai,[§¶] Gregg T. Beckham,[‖**] Roland Schulz,[¶‡‡] Michael F. Crowley,[‡††*] Edward C. Uberbacher,[†‡*] and Xiaolin Cheng[‡¶‡‡*]

[†]Computational Biology and Bioinformatics Group and [‡]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [§]Interdisciplinary Center for Scientific Computing, University of Heidelberg, Heidelberg Germany; [¶]UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [‖]National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colorado; [**]Department of Chemical Engineering, Colorado School of Mines, Golden, Colorado; [††]Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado; and [‡‡]Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, Tennessee

ABSTRACT　Cellobiohydrolases processively hydrolyze glycosidic linkages in individual polymer chains of cellulose microfibrils, and typically exhibit specificity for either the reducing or nonreducing end of cellulose. Here, we conduct molecular dynamics simulations and free energy calculations to examine the initial binding of a cellulose chain into the catalytic tunnel of the reducing-end-specific Family 7 cellobiohydrolase (Cel7A) from *Hypocrea jecorina*. In unrestrained simulations, the cellulose diffuses into the tunnel from the −7 to the −5 positions, and the associated free energy profiles exhibit no barriers for initial processivity. The comparison of the free energy profiles for different cellulose chain orientations show a thermodynamic preference for the reducing end, suggesting that the preferential initial binding may affect the directional specificity of the enzyme by impeding nonproductive (nonreducing end) binding. Finally, the Trp-40 at the tunnel entrance is shown with free energy calculations to have a significant effect on initial chain complexation in Cel7A.

INTRODUCTION

Nature has evolved cocktails of glycoside hydrolase (GH) enzymes to degrade plant cell walls. The primary components of many GH cocktails for plant cell wall deconstruction are cellulases, which degrade cellulose by hydrolyzing β-1,4-glycosidic bonds through either inverting or retaining acid hydrolysis mechanisms (1,2). The primary components of cellulase cocktails are typically processive enzymes, or cellobiohydrolases (CBHs), which are able to hydrolyze successive units of a cellulose chain before dissociation from the cellulose microfibril or before encountering an obstacle on the cellulose surface (3–5). CBHs typically are thought to bind via two modes, either *exo*-mode wherein they thread a chain from one end of the polymer or *endo*-mode where a chain is likely acquired in most or all of the subsites via putative loop motions (6–8). It has been shown that CBHs from a given GH family typically exhibit directionality in their processive motion (3,4,9,10). The directional preference is primarily thought to arise from the structural arrangement of the enzyme active site (11), and many GH structures have been solved with ligands in the active site that confirm a given directional preference for a specific GH family (11–14). In addition to the active site, other factors may also contribute to the directional specificity of a given cellulase enzyme, such as a preferential binding of the carbohydrate-binding module to cellulose and initial recognition of a specific cellulose chain end via the CD of the enzyme.

From crystal structures, most CBHs exhibit aromatic residues at the entrance of the active site tunnel (15–17). Previous mutagenesis and enzyme kinetic experiments demonstrated that mutation of tryptophan residues at the tunnel entrance of the Family 6 and Family 7 CBHs from *Hypocrea jecorina*, Cel6A and Cel7A, respectively, reduces the enzyme activity on crystalline cellulose, but not on amorphous substrates (18,19). Recently, high-speed atomic force microscopy (HS-AFM) experiments by Igarashi et al. (3) provided direct evidence that no obvious movement was observed for the W40A mutant on crystalline cellulose, suggesting the initial binding of a cellulose chain into the active site tunnel is important for enzymatic function. By monitoring the rate of initial-cut product generation, another recent experimental study suggested that the rate of *Trichoderma longibrachiatum* cellobiohydrolase I (*Tl*-Cel7A) catalyzed hydrolysis of bacterial microcrystalline cellulose is limited by the rate of enzyme complexation with cellulose chains, which implies that the initial binding step could be key in determining the enzymatic turnover rate (20). Even so, it remains unclear whether the initial binding of the substrate via *exo*-initiation to the active site tunnel occurs in a specific direction, thereby creating a directional preference for the enzyme's processive motion on crystalline cellulose.

The GH Family 7 CBH from *H. jecorina*, Cel7A, is of particular importance because it is the most abundant protein in the *H. jecorina* secretome by mass, and thus also typically the largest enzyme component of industrial

*Correspondence: michael.crowley@nrel.gov or uberbacherec@ornl.gov or chengx@ornl.gov

Editor: Michael Feig.

GH cocktails for biomass conversion (21). Cel7A catalyzes the hydrolysis of cellulose into cellobiose from the reducing end of a cellulose chain (16). *H. jecorina* Cel7A consists of a Family 1 carbohydrate binding module (CBM) (22) and GH Family 7 CD connected by a flexible, glycosylated linker peptide (14,16,23,24). The CD contains a ~50 Å long catalytic tunnel into which a single cellulose chain is threaded, cleaving one cellobiose unit per catalytic event (12,16). There are at least 10 binding sites in Cel7A from the −7 site at the tunnel entrance to the +3 site at the tunnel exit (12). Family 7 GHs generally employ a two-step retaining hydrolysis reaction to cleave glycosidic bonds between the −1 and +1 subsites via a catalytic triad of residues (15,25,26). In the case of *H. jecorina* Cel7A, Glu-217 is the catalytic acid and base and Glu-212 is the nucleophile that forms the enzyme-glycosyl intermediate. Asp-214 forms a stabilizing interaction with the nucleophile via a side-chain hydrogen bond. Once adsorbed to the cellulose surface, Cel7A is hypothesized to undergo a series of elementary steps including (3,19,20,27,28): 1), recognition of the reducing end of a cellulose chain; 2), initial threading of the cellulose chain into the catalytic tunnel; 3), formation of the catalytically active complex; 4), cleavage of the glycosidic bond via a two-step retaining mechanism; and 5), product release and threading of another cellobiose unit. The process repeats until the enzyme dissociates, gets stuck, or reaches the end of the cellulose chain (3,4,6,28).

Computer modeling is an important tool for understanding cellulase action due to the difficulty in experimentally studying individual steps in isolation (29). Computational approaches have been widely applied to study various aspects of cellulose hydrolysis, such as the binding of CBM onto the cellulose surface (30–33), hydrolysis of the β-1,4-glycosidic bond (34,35), modularity of the cellulase (36,37), and product expulsion (38,39). However, the initial threading of a single cellulose chain into the catalytic tunnel has not been studied computationally to date. In this work, we performed molecular dynamics (MD) simulations to characterize the dynamics of initial binding of a single cellulose chain into the catalytic tunnel of Cel7A from surrounding solutions. Specifically, here we address how the free cellulose chain end is recognized at the entrance of the catalytic tunnel and whether there is a directional preference for the initial binding of the cellulose to the catalytic tunnel.

To investigate these questions, we performed multiple unrestrained MD simulations of the Cel7A CD and a single cellodextrin chain of nine glucose molecules. In these simulations, the cellodextrin chain spontaneously diffuses into the catalytic tunnel by a cellobiose unit (~10 Å), independent of the initial orientation of the cellulose chain. We subsequently performed umbrella sampling simulations to determine the free energy profiles for the binding of the cellulose chain in four orientations at the entrance of

the catalytic tunnel, which covers initial 20 Å distance of the threading pathway, with 10 Å outside and the other 10 Å inside the tunnel. The free energy results suggest that the free energy landscape is flat for the initial processivity from positions −7 to −5, and that the chain is thermodynamically more stable at the −5 position. The comparison of the free energy profiles for the wild-type and the W40A mutant indicates that W40 stabilizes the initial complexation by ~3 kcal/mol, and thus may play a role in facilitating the processive action of Cel7A. Interestingly, the free energy profiles also show a clear difference for the binding at the −5 position between the reducing end and the nonreducing end, suggesting that the catalytic tunnel is able to preferentially bind the reducing end of a cellulose chain.

MATERIALS AND METHODS

The first set of equilibrium MD simulations were performed on *H. jecorina* Cel7A in complex with a cellodextrin nanomer chain (reducing end-Glc-1-Glc-2-Glc-3-Glc-4-Glc-5-Glc-6-Glc-7-Glc-8-Glc-9-nonreducing end) placed at five different positions—namely with Glc-1 in the −7, −5, −3, −1, +2 binding sites of the cellulase. The crystal structure of Cel7A in complex with a modeled cellulose oligomer (PDB code: 8CEL) was used as the starting structure (12), which is referred to as the +2 position, and the complex structures in the −1, −3, −5, and −7 positions were constructed by sequentially translating the cellulose chain out of the protein tunnel by two glucose units. The protonation states of the titratable residues were determined by a combined pKa calculation using the Karlsberg webserver (http://agknapp.chemie.fu-berlin.de/karlsberg/) and manually checking for local hydrogen bonding residues. Two independent simulations were performed for each of the previous five systems. The second set of unrestrained simulations were performed on the CD of Cel7A with the cellulose chain all started from the −7 position (the chain end glucose unit stacks against Trp-40) but in four different orientations, namely the original A orientation, the B orientation: rotated 180° from A orientation around the tunnel axis and with the α face of the Glc-1 ring stacked against Trp-40, the C orientation: with the nonreducing end facing the tunnel entrance and the β face of Glc-9 stacked against Trp-40, and the D orientation: rotated 180° from C orientation around the tunnel axis and with the α face of Glc-9 stacked against Trp-40. Ten independent simulations were conducted for each of the previous four orientations.

After the protein-cellulose complex structures were built, they were solvated with TIP3P water molecules with a minimum of 15 Å water on each side of a cubic box. Charge neutralization was accomplished with the addition of Na$^+$ and Cl$^-$ ions, resulting in a 0.1 M solution. This resulted in simulations ranging from 58,000 atoms (when the chain is fully threaded; Glc-1 at position +2) to 84,000 atoms (when the chain is completely outside the tunnel; Glc-1 at position −7). The solvated system underwent four equilibration steps: i), 2,000 steps of minimization with a fixed protein backbone, ii), five cycles of a 500-step minimization with decreasing positional restraints on the protein Cα atoms, iii), gradual temperature increase from 50 to 300 K in 10,000 steps of constant-volume MD simulation with harmonic restraints (with a force constant of 3 kcal mol^{-1}Å^{-2}) on the protein Cα atoms, and iv), 2 ns equilibration with decreasing positional restraints on the Cα atoms. All the MD simulations were performed with the NAMD 2.7 program (40) and the CHARMM27 force field (41) (with the backbone CMAP correction (42)) for the protein and the C35 carbohydrate force field (43) for the cellulose chain. A short-range cutoff of 9 Å was used for nonbonded interactions, and long-range electrostatic interactions were treated with the particle mesh Ewald method (44) with a grid spacing of 1.0 Å. Langevin dynamics and a Langevin piston algorithm were used to maintain the temperature at 300 K and a pressure of

1 atm. The r-RESPA multiple-time-step method was employed, with time steps of 2 fs for bonded, 2 fs for short-range nonbonded, and 4 fs for long-range electrostatic forces (45). The bonds between hydrogen and heavy atoms were constrained with the SHAKE algorithm (46).

Umbrella sampling simulations were performed to compute the potential of mean force (PMF) for the cellulose chain to enter the catalytic tunnel. Before the free energy simulations, the protein tunnel axis, which is defined as the line joining the C4 atoms of the two end glucose rings in the 8CEL structure (12), was aligned with the x axis. The displacement of the center of mass of the first reducing or nonreducing end pyranose ring (Glc-1 in A and B orientations, Glc-9 in C and D orientations) along the tunnel axis was chosen as the reaction coordinate. The simulations were conducted with 21 windows with a uniform spacing of 1 Å, which covers 10 Å outside and the other 10 Å inside the tunnel, namely from 10 Å outside the tunnel to −5 position. The 10 Å outside the tunnel position was created by translating the cellulose chain out by 10 Å from the −7 position along the tunnel axis. The initial configurations were generated by interpolating the x positions of the cellulose chain to the corresponding target window positions based on the −7 and −5 position structures. In each simulation, the x-coordinate of the center of mass of the heavy atoms of the first glucose ring next to the tunnel entrance (Glc-1 in A and B orientations, Glc-9 in C and D orientations) was subject to a harmonic positional restraint with a spring constant of 5 kcal·mol^{-1}Å^{-2}. A flat-bottomed restraint $u(y, z)$ with $R = 8$ Å was applied to the center of mass of the first glucose ring to prevent the cellulose chain from drifting laterally from the tunnel axis. The flat-bottomed restraint does not affect the cellulose inside the tunnel, and it was shown to only affect the overall offset of the resultant free energy profile in the tunnel region, but not its shape. In all the simulations, harmonic restraints (with force constant of 3 kcal·mol^{-1}·Å^{-2}) were applied to five selected protein alpha carbon atoms (Leu-14, Ile-203, Ala-224, Val-393, and Phe-423) to prevent the translational and rotational motions of the protein. These restraints are not coupled to the reaction coordinate through common atoms so we assume that they have no direct influence on the PMF calculation. For each window, after 2 ns of equilibration, 4 ns of simulation data were collected for analysis with the weighted histogram analysis method (47) to generate the PMF. Overall, four sets of umbrella sampling simulations were performed on the cellulase-cellulose complex with the cellulose chain in four different A, B, C, and D orientations. The fifth set of umbrella sampling simulations was performed on a W40A mutant with the cellulose chain in the A orientation. To test for convergence of the umbrella simulations, the data were split into four blocks, and then for each block a PMF was computed. Comparison of the PMFs from different blocks yields an average standard deviation <0.8 kcal mol^{-1}, suggesting convergence of the simulations. The standard deviation of the four PMFs was used as an estimate of the statistical errors of the computed PMFs.

To examine the structure of the hydration water, the proximal distribution function $g_{prox}(r)$ is given by (48):

$$g_{prox} = \frac{\langle n \rangle}{A(r)\Delta r \rho_{bulk}},$$

where $\langle n \rangle$ is the average number of water oxygen atoms found at a distance $[r, r + \Delta r]$ from a nonhydrogen atom on the surface of the cellulose ($\Delta r = 0.1$ Å), A(r) is the solvent-accessible surface area of the cellulose calculated with a probe radius of r, and $\rho_{bulk} = 0.0327$ Å^{-3} is the bulk number densities of the TIP3P water model determined from pure water simulations at 300 K. The hydration shell water is defined by a 4.3 Å distance cutoff between water oxygen atoms and any nonhydrogen cellulose atoms of the first three glucose residues, where this cutoff distance was determined as the minimum in the proximal distribution function.

The entropies of hydration waters around the cellulose were calculated with the two-phase thermodynamic (2PT) model (49,50), which partitions the translational and rotational density of states of water molecules $g(\omega)$ into gas-like, $g_g(\omega)$, and solid-like, $g_s(\omega)$ components:

$$g(\omega) = g^g(\omega) + g^s(\omega) = \frac{2}{k_B T} \lim_{\tau \to \infty} \int_{-\tau}^{\tau} C(t)e^{-iwt}dt,$$

where $g(\omega)$ is the Fourier transform of the velocity-autocorrelation function (VACF) $C(t)$. $C(t)$ can be split into the mass weighted VACF of the center of mass velocities and the moment of inertia weighted angular VACF. The translational (S_T) and rotational (S_R) entropies of water can be calculated by assigning the appropriate weight, λ, to the gas- and solid-like components:

$$S = \frac{1}{2\pi} \int_0^\infty g^g(\omega)\lambda^g(\omega) + \frac{1}{2\pi} \int_0^\infty g^s(\omega)\,\lambda^s(\omega).$$

The decomposition of $g(\omega)$ and the derivation of the weighting functions λ^g and λ^s have followed the exact procedure described previously (51). $C(t)$ for the hydration-shell water molecules was determined from five sets of 20 ps cellulose-Cel7A simulations, with five simulations for each of the A, B, C, and D and W40A configurations, from which velocities and coordinates were saved every 1 fs. Error bars in entropy values were estimated as standard deviations with the five 20 ps trajectories of each configuration.

RESULTS AND DISCUSSION

We divided the entire threading process into five stages, and simulated a single cellodextrin chain at five different locations, named −7, −5, −3, −1, and +2 as defined by the position of the first reducing end glucosyl moiety (Glc-1) relative to the catalytic residues. For each of the five binding positions, two independent MD simulations are started and run for a time duration of 20-30 ns. For clarity, the nine glucose rings are identified as Glc-1 to Glc-9 with Glc-1 at the reducing end and Glc-9 at the nonreducing end of the cellulose chain. The displacements of the center of Glc-1 are plotted in Fig. 1 for the simulations at the five binding positions. The cellulose chain undergoes significant movement in the position −7 simulation as indicated by the broadly distributed positions of Glc-1 (*green dots*) during the simulation, whereas no processive motion is observed in the simulations at any other four positions (also in Fig. S2 in the Supporting Material). The large-scale displacement at position −7 corresponds to a spontaneous movement from the −7 site to the −5 site. Ten additional MD trajectories (each ~20 ns long) were run, which resulted in five spontaneous diffusion events wherein the reducing end of the cellulose chain moves into the tunnel by an average distance of 7.1 Å. Further diffusion inside the tunnel was not observed in any trajectories.

In two of the 12 simulations from position −7, the cellulose chain binds to the −5 position. The first three glucose monomers Glc-1-Glc-3 overlap with those at position −5 (and equivalently the Glc-7-Glc-9 in the crystal structure 8CEL) (12) with a root mean-square deviation of 0.6 Å (Fig. 2 *a*) and 1.7 Å in the two simulations (Fig. S3). The protein shows no large-scale motions during the −7 to −5 transitions, the average backbone root mean-square
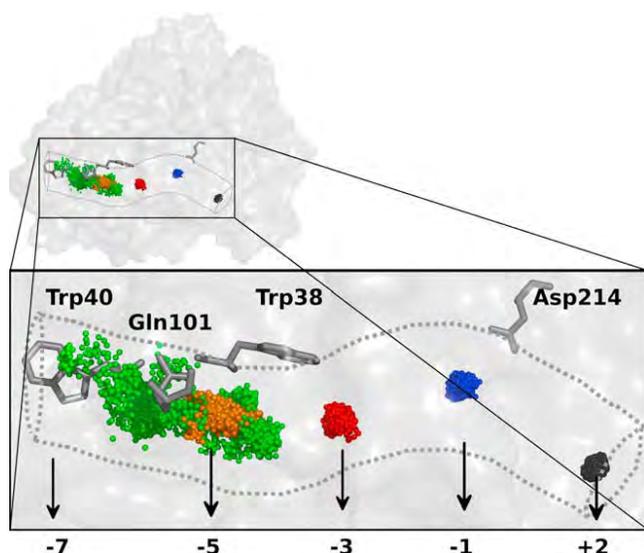
FIGURE 1 The movement of the cellulose chain inside the cellulase tunnel. Trp-40 marks the entrance of the tunnel and Asp-214 is one of the residues in the catalytic triad. Dotted lines roughly denote the contours of the catalytic tunnel. The distributions of the center of mass of the first glucose monomer (Glc-1) are plotted for the five sets of simulations started from positions −7 (*green*), −5 (*orange*), −3 (*red*), −1 (*blue*), and +2 (*black*), respectively.

deviation is 1.2 ± 0.4 Å. Among the residues lining the tunnel, hydrophobic residues are less flexible than polar residues (Fig. S4). Trp-40, which may play a critical role in orienting the cellulose substrate in the tunnel and the active site, exhibits quite small fluctuations while the chain enters the tunnel. Concerted formation and breaking of hydrogen bonds are evident between the hydroxyl groups of cellulose and protein side chains. As shown in Figs. 2, *b* and *c*, Asn-49 successively forms hydrogen bonds with Glc-1, Glc-2, and Glc-3, whereas Glc-1 hydrogen bonds with Gln-7, Gln-101, and Lys-181 as the cellulose chain moves into the tunnel during the first 6 ns. The concerted hydrogen bond formation and breaking may facilitate the spontaneous chain diffusion to the −5 position.

The cellulose chain possesses a directional asymmetry from the reducing end to the nonreducing end. Moreover, the two sides of the pyranose ring, referred to as face α (the opposite side from the CH_2OH group at C5) and face β (opposite from α face) may also affect the initial binding landscape due to different aromatic-carbohydrate interactions. To investigate if there is any directional and/ or face preference for the initial binding, we conducted 30 additional 20 ns long simulations from position −7 with the cellulose chain in three other orientations: B, C, and D, 10 simulations each. The initial orientation based on the 8CEL structure is referred to as the A orientation, which has the reducing end leading to the tunnel and the β glucose face stacking against Trp-40. The B orientation has the reducing end leading and the α face stacking to Trp-40. The C and D orientations have the nonreducing end leading,
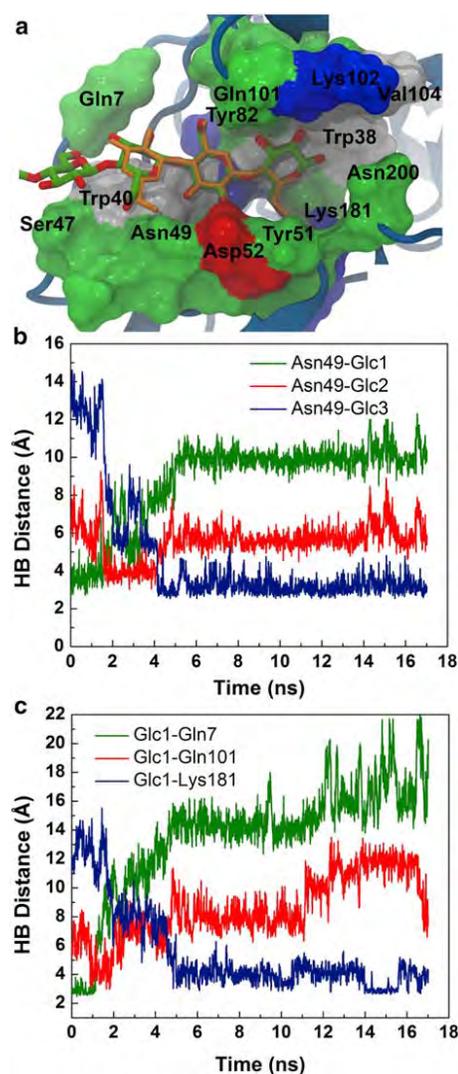

FIGURE 2 (*a*) The final snapshot from the simulation showing that the cellulose chain (*licorice* in *red* and *green*) moves by two glucose units and overlays with that in the −5 position binding structure (*licorice* in *orange*); other nearby residues are labeled and shown in surface representation; (*b*) Hydrogen bond distances between Asn-49 and the first (Glc-1), second (Glc-2), and third (Glc-3) glucosyl moieties of the cellulose chain as a function of time during the simulation; (*c*) Hydrogen bond distances between Glc-1 and Gln-7, Gln-101, and Lys-181 as a function of time during the simulation.

and the β and α faces stacking to Trp-40, respectively. The simulation results summarized in Fig. S5 show that the cellulose chain in the three other orientations can also spontaneously progress from the −7 to −5 positions, suggesting no significant barrier for the initial binding process.

We conducted umbrella sampling simulations (52) to determine free energy (PMF) profiles for the −7 to −5 binding. These simulations were performed on the wild-type Cel7A enzyme with the cellulose chain in the A, B, C, and D orientations, and the W40A Cel7A mutant with the chain in the A orientation. The starting configuration at reaction coordinate value −10 Å had the cellulose chain

10 Å outside the tunnel, the chain entering the tunnel at 0 Å in the −7 position, and at 10 Å in the −5 position. The PMF for the initial binding of the cellulose into the cellulase tunnel in orientation A is −5 kcal/mol downhill with no significant free energy barriers. The free energy profile is relatively flat outside the tunnel (Fig. 3 a). A stabilization of ~2 kcal/mol is observed at position −7, indicating the formation of a marginally stable initial enzyme-substrate complex. More importantly, the cellulose chain is stabilized by 5 kcal/mol at the −5 position. As the cellulose moves into the tunnel, hydrogen bonds are formed between the cellulose and the protein, offsetting the loss of hydrogen bonds between the cellodextrin chain and water. At position −5, the first three glucose residues Glc-1-Glc-3 make extensive interactions with charged and polar residues, including Gln-7, Asn-49, Asp-52, Gln-101, Asn-103, and Lys-181. When the cellulose binds in position −5, Glc-3 also appears to stack better with Trp-40 than the corresponding Glc-1 when the cellulose binding at position −7, as evidenced by more

favorable Glc-3/1-Trp-40 vdW interaction for the binding at position −5 than at the −7 position (−5.7 ± 2.2 kcal/mol vs. −3.8 ± 1.6 kcal/mol), indicating a stronger carbohydrate-aromatic enthalpic interaction for the −7 to −5 binding. Note that the entropic contribution to the −7 to −5 transition is not directly considered in the previous analysis, whereas the entropic contribution of hydration waters to the binding mechanisms of different cellulose chain orientations will be further discussed below.

The PMF for orientation B (*blue line* in Fig. 3 a) exhibits a similar profile as that for orientation A, except that all the energy barriers and wells in the former PMF are elevated by ~1 kcal/mol. By computing the average solvent excluded surface areas for the Trp-40 and Glc-3 pair (45.7 ± 1.1 Å^2 for A and 42.1 ± 1.4 Å^2 for B), we observe slightly improved aromatic-carbohydrate stacking for the β face in orientation A than the α face in orientation B. We note that the PMF difference between the A and B binding faces is only meaningful for the initial stages of the threading process. Given the twofold screw-axis symmetry of a single chain in crystalline cellulose I, the PMF for orientation A or B will converge as the chain in one orientation moves forward or backward by one glucose unit with respect to the other. Therefore, the binding of a cellulose chain in either A or B orientation can be productive. The difference is that the A binding will lead to an initial cut of cellobiose, while the B binding will produce glucose or cellotriose.

Compared to the reducing end binding orientations A and B, the PMFs for the nonreducing end binding in orientations C (*red line*) and D (*green line*) exhibit significant differences (Fig. 3 a). First, the stabilization at position −7 disappears. The PMFs for both C and D orientations show a barrier of 2–4 kcal/mol toward the −5 binding site in contrast to a moderate stabilization of ~4.5 kcal/mol in the PMFs for the two reducing end binding orientations. To further understand this observation, we defined and calculated three energy components during the simulation of each orientation at position −5: 1), the aromatic-carbohydrate stacking energy, which includes the nonbonded interactions between the three leading glucosyl rings (Glc-1–3 in A and B, Glc-7–9 in C and D) and the two tryptophan residues, Trp-40 and Trp-38; 2), the electrostatic energy of the first three glucose monomers and protein residues excluding Trp-40 and Trp-38; and 3), the vdW energy, involving all the dispersion and steric interactions of the first three glucose monomers with nearby residues except Trp-40 and Trp-38. As listed in Table 1, all three components favor the reducing end binding, but to different degrees. The aromatic-carbohydrate stacking makes only a small contribution, while both electrostatic and vdW interactions are 3–4 kcal/mol more favorable for the reducing end binding. This difference arises from a slightly altered binding mode at the −5 subsite for the reducing end and the nonreducing end conformations. The cellulose chain end in the reversed direction extends slightly off the sliding path of the reducing end
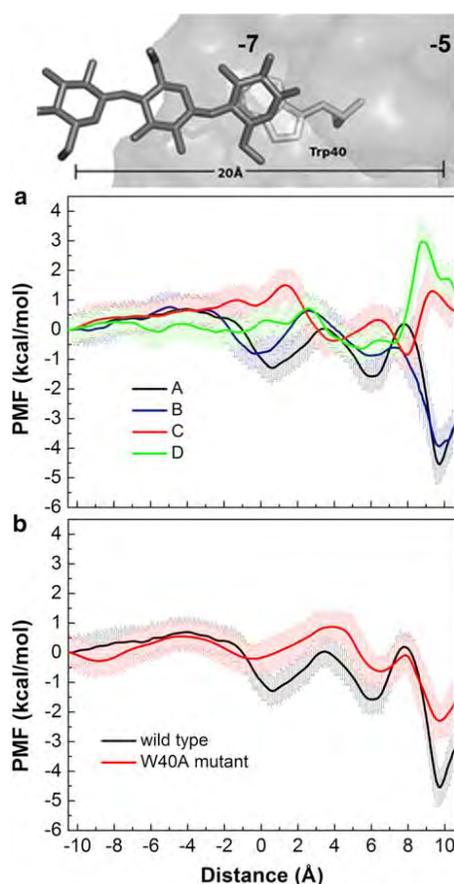

FIGURE 3   PMFs of the initial binding of the cellulose chain to the cellulase catalytic tunnel (*a*) for four different cellulose orientations: A (*black line*), B (*blue line*), C (*red line*), and D (*green line*), and (*b*) in a W40A mutant (*red line*) together with the wild-type cellulase (*black line*). The picture on the top shows the distance covered in the PMF calculations and the corresponding −5 and −7 positions.

TABLE 1   Binding energy decomposition (in kcal/mol) at position −5 for four different cellulose orientations

	A	B	C	D
Stacking	−9.3 ± 0.9	−8.3 ± 1.0	−8.1 ± 1.0	−7.3 ± 0.9
Electrostatic	−38.0 ± 4.2	−36.6 ± 3.7	−34.5 ± 5.3	−33.2 ± 6.3
van der Waals	−20.1 ± 1.8	−21.6 ± 1.9	−17.7 ± 1.9	−17.4 ± 2.1
Total	−67.4	−66.5	−60.3	−57.9

FIGURE 4   Proximal distribution functions of water oxygen atoms at a distance r from the surface of the first three glucose residues in four simulations of the cellulose in four different A, B, C, and D orientations and one simulation of the A orientation cellulose in the W40A mutant. The cellulose binds at subsite −5 in all simulations.

conformations (Fig. S6), thereby leading to less favorable interactions with nearby residues Tyr-51, Asn-103, Val-104, and Lys-181 on both sides of the sugar ring. This result suggests that the geometrical asymmetry of the cellulose chain affects both the geometrical fit and optimal hydrogen bonding, which results in a preferential reducing end binding at the −5 position. Although the spontaneous nonreducing end binding of a cellulose chain up to the −5 position is still possible and has also been observed in multiple unrestrained simulations, it is likely that the absence of stabilization at position −5 will not provide the necessary thermodynamic stabilization needed for decrystallizing crystalline cellulose (53).

Water molecules are often observed in protein-carbohydrate binding sites, mediating the interaction between the ligand and protein (54,55). Based on the crystal structure (8CEL), the catalytic tunnel of Cel7A is very well hydrated. We identified the crystal water molecules that are within 3 Å of the cellulose chain, shown as red spheres in Fig. S7. Those crystal waters that were displaced by the advancing reducing end or nonreducing end cellulose chain during the simulations are shown as green spheres. It is evident that there are some differences in the waters that were displaced when the cellulose entered first with the reducing or nonreducing end. However, it seems unlikely that this difference in water displacement can necessarily account for the differential binding of the nonreducing versus reducing end cellulose. This is because all these crystal waters are rather dynamic, and exchanged with bulk water on ps-ns timescales. Unlike in the case of concanavalin A (56), no conserved or strongly bound water was observed in the course of the simulations presented here.

To quantify the hydration water structures inside the protein tunnel, we computed the proximal distribution functions $g_{\text{prox}}(r)$ for the waters close to the surface of the first three glucose residues. As shown in Fig. 4, no statistically significant differences in hydration are found when the cellulose binds at subsite −5 in four different A, B, C, and D orientations, indicating that the interaction patterns of the hydration waters with the reducing end or nonreducing end cellulose chain are similar. We further analyzed differences in dynamics of the hydration waters in the four simulations by computing translational and rotational velocity autocorrelation functions and the associated entropies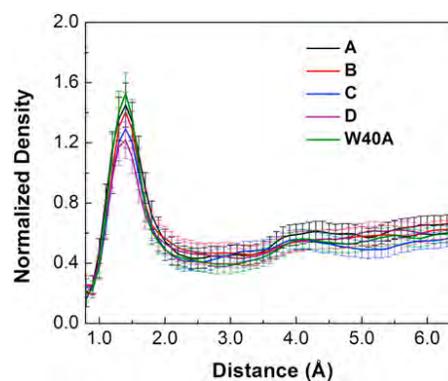 using a two-phase thermodynamic model (49,50). As listed in Table 2, our results show no statistically significant differences in translational or rotational entropies for the hydration waters when the cellulose enters first with the reducing or nonreducing end. Taken together, our calculations suggest that the initial binding of cellulose to the cellulase tunnel is not (water) entropically driven but enthalpically driven, akin to recent computational results on the concanavalin A-trimannoside complexes (56). Likewise, the origin of the preferential binding of the reducing end over the nonreducing end cellulose chains does not seem to arise from those waters at the entrance of the protein tunnel. Instead, we speculate that the −5 binding site is pre-organized in such an arrangement that is more favorable for the recognition of the reducing end. To accommodate the cellulose chain in the opposite direction, the structures of the protein and carbohydrate need to be slightly distorted (Fig. S6), which leads to less favorable interaction between the nonreducing end cellulose and the protein (Table 1).

A previous HS-AFM study suggested the importance of Trp-40 for the processive movement of Cel7A on crystalline cellulose (3,18,19). Our unrestrained simulations also indicate the importance of the Trp-40-glucose interaction in the initial binding. To quantitatively assess the role of Trp-40 in the initial binding process, we computed the PMF for the binding of the cellulose chain in orientation A to the W40A mutant (Fig. 3 b). Compared to the wild-type PMF, the mutant PMF exhibits broader peaks and shallower wells in the tunnel entrance region, with the first minimum at the −7 position increased by 1.5 kcal/mol, the second minimum at 6 Å by 1 kcal/mol and the third minimum at the −5 position by 3 kcal/mol, indicating an overall weaker binding for W40A. This provides potential insight into the thermodynamic origin underlying the hindered processive movement of W40A on the crystalline cellulose—the destabilization of the initial encounter complex for the mutant due to the absence of

TABLE 2   Comparison of entropies of hydration water in four simulations of cellulose at 300 K in four different orientations (A, B, C, and D) and one simulation of the orientation A cellulose in the W40A mutant (W40A)

	A	B	C	D	W40A
Entropy trans TS_T (kcal/mol)	3.41 ± 0.05	3.41 ± 0.09	3.37 ± 0.07	3.40 ± 0.08	3.51 ± 0.13
Entropy rot TS_R (kcal/mol)	0.94 ± 0.01	0.93 ± 0.02	0.91 ± 0.01	0.91 ± 0.00	0.94 ± 0.04
Total TS (kcal/mol)	4.35 ± 0.06	4.24 ± 0.10	4.29 ± 0.08	4.32 ± 0.08	4.45 ± 0.17

The cellulose binds at subsite −5 in all simulations.

favorable glucose-aromatic interaction. Previous free energy calculations have shown that the strength of the aromatic-carbohydrate interactions varies dramatically depending on the location of the aromatic residues, in the range of +1.3 to −9.4 kcal/mol (57). Our calculated binding energy difference on the order of 3 kcal/mol is comparable to the contribution of Trp-272 (3.8 kcal/mol) in *H. jecorina* Cel6A, which likely plays a similar role as Trp-40 in *H. jecorina* Cel7A (18).

CONCLUSIONS

In summary, we observe spontaneous diffusion of a cellononaose chain from the −7 position to the −5 position in multiple simulations, with the reducing end facing the enzyme's catalytic tunnel entrance (A and B orientations), which arises from an overall downhill free-energy landscape for the initial threading process. PMF calculations reveal a clear difference in the binding of a cellulose chain at position −5 between the reducing end and the reversed (C and D) directions, suggesting a potential mechanism for the recognition of a free cellulose reducing chain end by Cel7A. The main differences in the PMFs are near the −5 position. The spatial arrangement, or stereochemistry, of the different cellulose chain ends results in different interaction modes with the tunnel entrance. The binding of the nonreducing end appears to experience a slight energetic frustration to initial processivity. Our PMF calculations also substantiate the importance of Trp-40 in the initial binding of cellulose to Cel7A. Compared to the wild-type Cel7A, the W40A mutant exhibits weaker binding at both −5 and −7 positions, which correlate with the observation of decreased processivity toward crystalline substrates from recent HS-AFM experiments (3).

Overall, our simulation results suggest that the initial binding (up to position −5) of the cellulose is largely thermodynamically downhill, and in *exo*-mode initiation, this may aid in the cellulose ligand binding at the tunnel entrance. Most CBHs exhibit directional specificity, which can arise not only from structural arrangement of the enzyme active site that only allows for catalytic reaction to occur, or from a preferential binding of CBM on substrate surfaces, but also from an initial recognition of a specific cellulose chain end. Although not absolutely required for Cel7A to achieve its directional processive activity, the initial recognition of a specific cellodextrin chain end may lead to fewer nonproductive binding events, thus enhancing the overall hydrolytic efficiency of the enzyme.

SUPPORTING MATERIAL

Seven supporting figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(13)00070-2.

REFERENCES

1. Vasella, A., G. J. Davies, and M. Böhm. 2002. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* 6:619–629.

2. Bayer, E. A., H. Chanzy, …, Y. Shoham. 1998. Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* 8:548–557.

3. Igarashi, K., A. Koivula, …, M. Samejima. 2009. High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose. *J. Biol. Chem.* 284:36186–36190.

4. Igarashi, K., T. Uchihashi, …, M. Samejima. 2011. Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science.* 333:1279–1282.

5. Jalak, J., and P. Väljamäe. 2010. Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis. *Biotechnol. Bioeng.* 106:871–883.

6. Kurašin, M., and P. Väljamäe. 2011. Processivity of cellobiohydrolases is limited by the substrate. *J. Biol. Chem.* 286:169–177.

7. Jalak, J., M. Kurašin, …, P. Väljamäe. 2012. Endo-exo synergism in cellulose hydrolysis revisited. *J. Biol. Chem.* 287:28802–28815.

8. Stahlberg, J., G. Johansson, and G. Pettersson. 1991. A new model for enzymatic-hydrolysis of cellulose based on the 2-domain structure of cellobiohydrolase-I. *Nat. Biotechnol.* 9:286–290.

9. Chanzy, H., B. Henrissat, …, M. Schulein. 1983. The action of 1,4-beta-D-glucan cellobiohydrolase on valonia cellulose micro-crystals - an electron-microscopic study. *FEBS Lett.* 153:113–118.

10. Barr, B. K., Y.-L. Hsieh, …, D. B. Wilson. 1996. Identification of two functionally different classes of exocellulases. *Biochemistry.* 35:586–592.

11. Rye, C. S., and S. G. Withers. 2000. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* 4:573–580.

12. Divne, C., J. Ståhlberg, …, T. A. Jones. 1998. High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel

of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* 275: 309–325.

13. Ubhayasekera, W., I. G. Muñoz, …, S. L. Mowbray. 2005. Structures of *Phanerochaete chrysosporium* Cel7D in complex with product and inhibitors. *FEBS J.* 272:1952–1964.

14. Ståhlberg, J., C. Divne, …, T. A. Jones. 1996. Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* 264:337–349.

15. Rouvinen, J., T. Bergfors, …, T. A. Jones. 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science.* 249:380–386.

16. Divne, C., J. Ståhlberg, …, T. A. Jones. 1994. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science.* 265:524–528.

17. Guimarães, B. G., H. Souchon, …, P. M. Alzari. 2002. The crystal structure and catalytic mechanism of cellobiohydrolase CelS, the major enzymatic component of the *Clostridium thermocellum* Cellulosome. *J. Mol. Biol.* 320:587–596.

18. Koivula, A., T. Kinnari, …, T. T. Teeri. 1998. Tryptophan 272: an essential determinant of crystalline cellulose degradation by *Trichoderma reesei* cellobiohydrolase Cel6A. *FEBS Lett.* 429:341–346.

19. von Ossowski, I., J. Ståhlberg, …, T. T. Teeri. 2003. Engineering the exo-loop of *Trichoderma reesei* cellobiohydrolase, Cel7A. A comparison with Phanerochaete chrysosporium Cel7D. *J. Mol. Biol.* 333: 817–829.

20. Fox, J. M., S. E. Levine, …, H. W. Blanch. 2012. Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes. *Biochemistry.* 51:442–452.

21. Martinez, D., R. M. Berka, …, T. S. Brettin. 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat. Biotechnol.* 26(5):553–560.

22. Kraulis, J., G. M. Clore, …, A. M. Gronenborn. 1989. Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry.* 28:7241–7257.

23. Beckham, G. T., Y. J. Bomble, …, M. F. Crowley. 2010. The O-glycosylated linker from the *Trichoderma reesei* Family 7 cellulase is a flexible, disordered protein. *Biophys. J.* 99:3773–3781.

24. Harrison, M. J., A. S. Nouwens, …, N. H. Packer. 1998. Modified glycosylation of cellobiohydrolase I from a high cellulase-producing mutant strain of *Trichoderma reesei*. *Eur. J. Biochem.* 256:119–127.

25. Henriksson, H., J. Stahlberg, …, R. Isaksson. 1997. The catalytic amino-acid residues in the active site of cellobiohydrolase 1 are involved in chiral recognition. *J. Biotechnol.* 57:115–125.

26. Kleywegt, G. J., J. Y. Zou, …, T. A. Jones. 1997. The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 A resolution, and a comparison with related enzymes. *J. Mol. Biol.* 272:383–397.

27. Koivula, A., T. Reinikainen, …, T. T. Teeri. 1996. The active site of *Trichoderma reesei* cellobiohydrolase II: the role of tyrosine 169. *Protein Eng.* 9:691–699.

28. Chundawat, S. P., G. T. Beckham, …, B. E. Dale. 2011. Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* 2:121–145.

29. Beckham, G. T., Y. J. Bomble, …, M. F. Crowley. 2011. Applications of computational science for understanding enzymatic deconstruction of cellulose. *Curr. Opin. Biotechnol.* 22:231–238.

30. Beckham, G. T., J. F. Matthews, …, M. F. Crowley. 2010. Identification of amino acids responsible for processivity in a Family 1 carbohydrate-binding module from a fungal cellulase. *J. Phys. Chem. B.* 114:1447–1453.

31. Bu, L., G. T. Beckham, …, M. R. Nimlos. 2009. The energy landscape for the interaction of the family 1 carbohydrate-binding module and the cellulose surface is altered by hydrolyzed glycosidic bonds. *J. Phys. Chem. B.* 113:10994–11002.

32. Taylor, C. B., M. F. Talib, …, G. T. Beckham. 2012. Computational investigation of glycosylation effects on a family 1 carbohydrate-binding module. *J. Biol. Chem.* 287:3147–3155.

33. Nimlos, M. R., G. T. Beckham, …, M. F. Crowley. 2012. Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. *J. Biol. Chem.* 287: 20603–20612.

34. Barnett, C. B., K. A. Wilkinson, and K. J. Naidoo. 2011. Molecular details from computational reaction dynamics for the cellobiohydrolase I glycosylation reaction. *J. Am. Chem. Soc.* 133:19474–19482.

35. Barnett, C. B., K. A. Wilkinson, and K. J. Naidoo. 2010. Pyranose ring transition state is derived from cellobiohydrolase I induced conformational stability and glycosidic bond polarization. *J. Am. Chem. Soc.* 132:12800–12803.

36. Ting, C. L., D. E. Makarov, and Z.-G. Wang. 2009. A kinetic model for the enzymatic action of cellulase. *J. Phys. Chem. B.* 113:4970–4977.

37. Lin, Y., J. Silvestre-Ryan, …, J. W. Chu. 2011. Protein allostery at the solid-liquid interface: endoglucanase attachment to cellulose affects glucan clenching in the binding cleft. *J. Am. Chem. Soc.* 133:16617–16624.

38. Bu, L., G. T. Beckham, …, M. F. Crowley. 2011. Probing carbohydrate product expulsion from a processive cellulase with multiple absolute binding free energy methods. *J. Biol. Chem.* 286:18161–18169.

39. Bu, L. T., M. R. Nimlos, …, G. T. Beckham. 2012. Product binding varies dramatically between processive and nonprocessive cellulase enzymes. *J. Biol. Chem.* 287:24807–24813.

40. Phillips, J. C., R. Braun, …, K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.

41. MacKerell, A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

42. Mackerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.

43. Guvench, O., E. R. Hatcher, …, A. D. Mackerell. 2009. CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Chem. Theory Comput.* 5:2353–2370.

44. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald - an N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

45. Tuckerman, M., B. Berne, and G. Martyna. 1992. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* 97:1990–2001.

46. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* 23: 327–341.

47. Kumar, S., D. Bouzida, …, J. M. Rosenberg. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* 13:1011–1021.

48. Ashbaugh, H. S., L. R. Pratt, …, T. L. Beck. 2005. Deblurred observation of the molecular structure of an oil-water interface. *J. Am. Chem. Soc.* 127:2808–2809.

49. Lin, S. T., M. Blanco, and W. A. Goddard. 2003. The two-phase model for calculating thermodynamic properties of liquids from molecular dynamics: validation for the phase diagram of Lennard-Jones fluids. *J. Chem. Phys.* 119:11792–11805.

50. Lin, S. T., P. K. Maiti, and W. A. Goddard, 3rd. 2010. Two-phase thermodynamic model for efficient and accurate absolute entropy of water from molecular dynamics simulations. *J. Phys. Chem. B.* 114:8191–8198.

51. Petridis, L., R. Schulz, and J. C. Smith. 2011. Simulation analysis of the temperature dependence of lignin structure and dynamics. *J. Am. Chem. Soc.* 133:20277–20287.

52. Torrie, G. M., and J. P. Valleau. 1977. Non-physical sampling distributions in Monte-Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* 23:187–199.

53. Beckham, G. T., J. F. Matthews, …, M. F. Crowley. 2011. Molecular-level origins of biomass recalcitrance: decrystallization free energies for four common cellulose polymorphs. *J. Phys. Chem. B.* 115:4118–4127.

54. Gauto, D. F., S. Di Lella, …, M. A. Martí. 2009. Carbohydrate-binding proteins: dissecting ligand structures through solvent environment occupancy. *J. Phys. Chem. B.* 113:8717–8724.

55. Saraboji, K., M. Håkansson, …, D. T. Logan. 2012. The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugarlike framework of oxygens: ultra-high-resolution structures and water dynamics. *Biochemistry.* 51:296–306.

56. Kadirvelraj, R., B. L. Foley, …, R. J. Woods. 2008. Involvement of water in carbohydrate-protein binding: concanavalin A revisited. *J. Am. Chem. Soc.* 130:16933–16942.

57. Payne, C. M., Y. J. Bomble, …, G. T. Beckham. 2011. Multiple functions of aromatic-carbohydrate interactions in a processive cellulase examined with molecular simulation. *J. Biol. Chem.* 286:41028–41035.