The impact of *Trichoderma reesei* Cel7A carbohydrate binding domain mutations on its binding to a cellulose surface: a molecular dynamics free energy study

Tong Li · Shihai Yan · Lishan Yao

Received: 2 March 2011 / Accepted: 27 June 2011 © Springer-Verlag 2011

Abstract A critical role of the Family 7 cellobiohydrolase (Cel7A) carbohydrate binding domain (CBD) is to bind to a cellulose surface and increase the enzyme concentration on the surface. Several residues of *Trichoderma reesei* Cel7A CBD, including Y5, N29, Y31, Y32 and Q34, contribute to cellulose binding, as revealed by early experimental studies. To investigate the interactions between these important residues and cellulose, we applied a thermodynamic integration method to calculate the cellulose–Cel7A CBD binding free energy changes caused by Y5A, N29A, Y31A, Y32A and Q34A mutations. The experimental binding trend was successfully predicted, proving the effectiveness of the complex model. For the two polar residue mutants N29A and Q34A, the changes in the electrostatics are comparable to those of van der Waals, while for three Y to A mutants, the free energy differences mainly come from van der Waals interactions. However, in both cases, the electrostatics dominates the interactions between individual residues and cellulose. The side chains of these residues are rigidified after the complex is formed. The binding free energy changes for the two mutants Y5W and Y31W were also determined, and for these the van der Waals interaction was strengthened but the electrostatics was weakened.

Keywords Cellobiohydrolase · Cellulose · Thermodynamic integration · Molecular dynamics · Free energy

Introduction

Biofuels produced from biomass are a promising and environmentally friendly alternative to fossil fuels, but the process of producing biofuels faces substantial challenges, mainly due to its high cost compared to traditional energy production. One approach is to hydrolyze cellulose—the main polymer in biomass—to sugars by cellulases, which are then fermented to ethanol. Cellulose is the most abundant bioresource on Earth and the major structural component of plant cell walls. In nature, cellulose molecular chains are linear polymers composed of β-1,4 linked D-glucose units.

The enzymes involved in the hydrolysis of cellulose consist of endoglucanases (EG), exoglucanases or cellobiohydrolases (CBH), and β-glucosidases. Among them, endoglucanases are responsible for breaking down the intramolecular β-1,4 glucosidic bonds of cellulose chains to produce new chain ends; exoglucanases hydrolyze cellulose chains from their ends and release soluble cellobiose; and β-glucosidases catalyze cellobiose to glucose [1, 2]. The enzymes secreted by the filamentous fungus *Trichoderma reesei* (*T. reesei*), which are highly effective against native and crystalline cellulose [3], are well studied. *T. reesei* produces a mixture of cellulases consisting of two CBHs (CBH I or Cel7A and CBH II or Cel6A) and several EGs [4, 5]. The enzyme Cel7A is composed of a large catalytic domain (CD) and a small cellulose-binding domain (CBD) connected via a glycosylated peptide linker [6]. The CD of Cel7A, containing a 50 Å tunnel-shaped active site, binds a cellulose chain and hydrolyzes the glycosidic bond to release cellobiose [3, 7–9].

The three-dimensional solution structure of the Cel7A CBD from *T. reesei* has been determined by nuclear magnetic resonance (NMR) spectroscopy [10], which
revealed that several aromatic residues, Y5, Y31 and Y32, form a hydrophobic surface. However, there is no experimental structure of the Cel7A–CBD–cellulose complex due to the insoluble nature of cellulose. The experimental study by Linder and coworkers [4] has shown that these three tyrosine residues are critical for binding to crystalline cellulose, while N29 and Q34 also contribute but are less important. Recently, the CBD cellulose complex has been modeled in order to investigate the binding mechanism by molecular dynamics (MD) simulations. It has been suggested that CBD may assist with the recognition of the cellulose chain end and facilitate the processivity of Cel7A [11–13]. Beckham et al. predicted that electrostatic interactions dominate the energy changes in the Cel7A CBD processivity [12]. However, it is very important to validate the CBD–cellulose complex model against experimental data.

Free energy methods provide a rigorous way to obtain important and readily measurable thermodynamic observables, such as binding free energy, from MD simulations [14, 15]. Chemical alchemy is an important and accurate class of free energy calculations. This approach involves simulating a transformation of one chemical system to a different one along a nonphysical path to compute the free energy change associated with the transformation [14]. Among the alchemical methods, thermodynamic integration (TI) [15–19] has grown in popularity over the last few years. A parameter \( \lambda \), ranging from zero to one, is introduced to define the intermediate states, with \( \lambda = 0 \) (1) corresponding to the initial (final) state. A series of simulations are performed for all the states, and then the partial derivative of the Hamiltonian over \( \lambda \) is estimated for each state and integrated over \( \lambda \) to yield the free energy change of the transformation. Although the computational demand is still large, TI has come of age with increased computer power and the wide use of parallel computers. One important application of the TI method is the study of the consequences of protein side-chain mutations, which has shown promising results [16, 20–22].

In this work, we calculated the relative binding free energy differences between the mutants (N29A, Q34A, Y5A, Y31A and Y32A) and wild-type CBD that bind to cellulose. Qualitative agreement between the experimental and computed free energy differences was observed, validating the effectiveness of the CBD cellulose-binding model. The absolute binding free energy calculation of CBD involves the annihilation of CBD on the cellulose surface and in water, which is a formidable task, so it was not pursued in this work. On the other hand, the absolute binding free energy was also difficult to determine (if not impossible) experimentally because of the insolubility of cellulose.

**Methods**

**Preparation of the system**

The starting coordinates for the CBD of Cel7A were obtained from the NMR structure [10] (PDB code 1CBH). The cellulose slab was modeled based on the cellulose I\( \beta \) crystal structure presented by Nishiyama et al. [23]. As shown in Fig. 1, the cellulose is three layers deep with five chains in each layer, and each chain includes 12 glucose units. Consistent with early studies [12, 13], the CBD is aligned on a cellulose surface with the aromatic residues Y5, Y31 and Y32 within 3 Å of the cellulose surface simultaneously, and the Y5, Y31 and Y32 rings are stacked over the glucose rings of cellulose. The complex model is stable during the MD equilibration (see the section “Results and discussion”).

**Molecular dynamics simulations**

The simulations were performed using the SANDER module in the AMBER 10 program [24], with the AMBER99sb [25] force field used for the protein and the GLYCAM_06 force field [26] [27] for the cellulose. The complex was soaked in a rectangular box with 12 Å TIP3P water molecules on each side using the LEAP module in AMBER. Na\(^+\) ions were added to neutralize the system. These yielded about 35,000 atoms for the simulation system. The 1–4 electrostatic (SCEE) and nonbonded (SCNB) scaling factors were set to their default values, with SCEE=1.2 and SCNB=2.0. It was realized later that the SCEE and SCNB scaling factors of cellulose should be set to one to yield the correct rotamer population, which is not done properly in AMBER 10. However, in our work, cellulose is restrained by using a harmonic potential with a force constant of 2.0 kcal mol\(^{-1}\) Å\(^{-2}\) on the C1 and C4 atoms, essentially removing the dihedral fluctuations along the C1–O4 and C4–O4 bonds.

Energy minimization was performed to remove bad contacts, and then the system was heated from 0 K to

**Fig. 1** Structural model of the CBD and cellulose complex. The side chains of residues Y5, N29, Y31, Y32 and Q34 are shown in stick representation, with the carbon atoms colored in cyan and the oxygen atoms in red. The cellulose was modeled by three layers of cellulose chains.
300 K in a 50 ps constant volume MD simulation, after which a 1 ns NPT simulation was run to equilibrate the water so that the water density reached 1.0 kg/L. The particle mesh Ewald (PME) method [28] was applied to calculate long-range electrostatics interactions. The SHAKE method [29] was applied to constrain all the covalent bonds involving hydrogen atoms. Periodic boundary conditions were applied. A nonbonded pair list cutoff of 12.0 Å was used. The time step used to integrate the equations of motion was 0.002 ps. Coordinates were saved every 0.4 ps along the 20 ns production run, and the trajectories were analyzed with the PTRAJ module of AMBER.

Free energy calculation by thermodynamics integration

According to the thermodynamic cycle, four states should be considered during the calculation of the relative binding free energy. The relative binding free energy difference \( \Delta \Delta G_{\text{bind}} \) between the wild-type CBD cellulose complex (WT-CBD-Cel) and a mutant CBD cellulose complex (MU-CBD-Cel) could be obtained from \( \Delta G_B - \Delta G_A \) (Fig. 2). However, as discussed above, directly calculating the binding free energies \( \Delta G_A \) and \( \Delta G_B \) is a formidable task, which can be avoided by designing a thermodynamic cycle as shown in Fig. 2. We can get \( \Delta \Delta G_{\text{bind}} \) by calculating the difference in the free energy changes for the free (\( \Delta G_1 \)) and bound (\( \Delta G_2 \)) states.

\[
\Delta \Delta G_{\text{bind}} = \Delta G_B - \Delta G_A = \Delta G_2 - \Delta G_1
\]

(1)

\( \Delta G_1 \) (\( \Delta G_2 \)) corresponds to the free energy change in the alchemical process of converting the wild type to the mutant residue in the absence (presence) of the cellulose. It can be computed using the following equation:

\[
\Delta G_{W \rightarrow M} = \int_0^1 \langle \delta V/\partial \lambda \rangle_\lambda d\lambda,
\]

(2)

where the subscripts W and M denote the wild type and mutant state, \( V \) is the potential energy, and \( \lambda \) is a parameter that ranges from 0 (wild type) to 1 (mutant). The binding affinities of the five mutants, including Y5A, N29A, Y31A, Y32A and Q34A, to cellulose were determined experimentally. The free energy calculations for the mutation of each of these five residues to an alanine were performed by dividing the process into three steps, as commonly employed in TI simulations. The first step is to remove the partial charges of the side chain atoms of the residue subject to mutation, except for the \( C_\beta \) atom and the two \( H_\beta \) hydrogens that exist in the alanine residue. In step two, the van der Waals (VDW) interactions with the atoms whose charges are removed in step one are eliminated (this is described in more detail below). In step three, the third \( H_\beta \) hydrogen atom attached to the \( C_\beta \) of the alanine residue gets the partial charge back. During the VDW transformation in step two, a linear mixing soft core potential is applied in which nonbonded VDW interactions are represented by a \( \lambda \)-dependent modified LJ equation [14],

\[
V_{\lambda, \text{disappearing}} = 4 \varepsilon (1 - \lambda) \left[ \left( \frac{1}{\alpha \lambda + \left( \frac{r_{ij}}{\sigma} \right)^6 \lambda} \right) - \left( \frac{1}{\alpha \lambda + \left( \frac{r_{ij}}{\sigma} \right)^6} \right) \right],
\]

(3)

where \( \varepsilon \) is the well depth and \( \sigma \) is the collision diameter, \( \alpha \) is an adjustable constant which is set to 0.5 in this study, and \( r_{ij} \) is the distance between atoms \( i \) and \( j \). Soft core scaling provides smooth \( \lambda \) curves that are best suited for commonly used numerical integration schemes [14]. Equation 3 is the potential function used for atoms that disappear during a VDW transformation. When \( \lambda = 0 \), it has the same form as the normal LJ equation; for \( \lambda \) values close to 1, it provides a smooth interaction function that allows other atoms to be situated on top of a vanishing atom, with a finite energy penalty that goes to zero at \( \lambda = 1 \) [14].

The last configuration of the CBD-cellulose complex from the 20 ns simulation was selected as the starting structure for the alchemical simulations to calculate \( \Delta G_2 \). A CBD system without cellulose was prepared with a size of ~11000 atoms for the \( \Delta G_1 \) calculations. In general, for each \( \lambda \) in the three-step transformation, we carried out a 100 ps constant-volume equilibration run followed by a 400 ps production run where the energy derivative \( \partial V/\partial \lambda \) for each \( \lambda \) was collected. \( \lambda \) values were set to 0.025n (\( n = 1, 2, \ldots, 39 \)) in each of the three steps; therefore, there were a total of 117 windows for each mutation. The free energy change associated with each of the three steps was calculated by numerically integrating the derivative data. The integration was done using an in-house perl script. The errors were estimated by splitting the trajectories into four equal blocks and calculating the corresponding free energy differences.

Besides the five single mutations (Y5A, N29A, Y31A, Y32A and Q34A), we also performed two more alchemy simulations (Y5W and Y31W) to determine the binding affinity change after tryptophan substitution. Presumably,
the tryptophan ring prefers to be parallel to the cellulose surface to maximize the VDW interaction, similar to the tyrosine phenol ring conformations on a cellulose surface. From this reasoning, we designed two conformations with \((\chi_1, \chi_2)\) dihedrals equal to \((-36^\circ, 113^\circ)\) and \((-36^\circ, -66^\circ)\) for W5, and \((-69^\circ, 83^\circ)\) and \((-69^\circ, -96^\circ)\) for W31, respectively (see Fig. 3), as the starting structures, but they were not restrained during the TI simulations. In total, there were nine TI simulations with a total simulation time of \(\sim 1 \mu s\).

Side-chain conformational entropy

The side-chain conformational entropy, \(S_{\text{conf}}\), was calculated using T-Analyst \([30]\). \(S_{\text{conf}}\) is defined as

\[
S_{\text{conf}} = -R \sum p_i \ln(p_i),
\]

where \(R\) is the gas constant, \(p_i\) is the probability distribution of angle \(i\), which refers to the side chain \((\chi_1, \chi_2)\) of the mutated residues in this work. The distributions were built from 20 ns molecular dynamics trajectories of free and cellulose-bound CBD. For each residue mutation, the entropy change was estimated using the T-Analyst program, and the entropic contribution to the binding free energy was calculated with the temperature set to 300 K. The bin size for each \(p_i\) was \(5^\circ\) for side-chain dihedrals \([30]\).

Results and discussion

20 ns MD

A putative binding position for CBD was proposed where the aromatic rings of the three tyrosine residues are nearly collinearly aligned along the cellulose fiber axis \([31]\). The close contact of the tyrosine and sugar rings is in concert with the fact that three tyrosine residues are critical in the CBD cellulose binding based on the experimental study. The Cel7A processivity action against insoluble highly crystalline cellulose is unambiguously toward glucose nonreducing ends from reducing ends \([9]\), which effectively restrains the alignment of the CBD relative to cellulose. Figure 1 shows the modeled initial structure of the CBD–cellulose complex for the 20 ns simulation. During the simulation, the RMSD values of all atoms compared to the starting structure reach \(\sim 0.8\) Å after several ps and are maintained at around 1 Å without large deviations (Fig. 4a).

In addition, the distance between the center of mass (COM) of the CBD and the COM of cellulose remains stable (Fig. 4b). Figure 4c shows the projection of the CBD COM on the cellulose surface during the 20 ns MD simulation, which is narrowly distributed. The results indicate that the system is equilibrated after the 20 ns simulation and can serve as the input geometry for the TI simulations.

Free energy calculations

The mutations of functionally important amino acids such as Y5, N29, Q34, Y31 and Y32 in the CBD domain to alanine have been shown experimentally to reduce the binding affinity and activity of Cel7A to crystalline cellulose \([4, 32, 33]\). To elucidate the fundamental principles governing the interaction between the CBD and cellulose, a three-step TI method was used to calculate the binding free energy differences between the WT–CBD-Cel and MU–CBD-Cel (see “Methods”). The calculations from the first and third steps are related to electrostatic effects, and those from the second step are the VDW interactions. In the second step, soft core Lennard–Jones (LJ) potentials were used to provide smooth \(\lambda\) curves that are best suited to commonly used numerical integration schemes \([14]\).

Plots of \(\partial V/\partial \lambda\) versus \(\lambda\) for steps one and two are depicted in Fig. 5. Switching on the charge of the third hydrogen attached to \(C_\beta\), as done in step three, provides a significantly smaller contribution to the free energy changes for all five mutations, so the corresponding \(\partial V/\partial \lambda\) plot is not shown here. The \(\partial V/\partial \lambda\) values versus time for step two at \(\lambda = 0.125\) are shown in Fig. S1 as a typical energy derivative plot (see the “Electronic supplementary material”), and this indicates that the systems are equilibrated during the production period. It is worth noting that in the VDW soft core alchemical process, a Y to A mutation generally displays larger \(\partial V/\partial \lambda\) fluctuations than a Q or N to A mutation does, especially for the CBD–cellulose complex (Fig. 5). To further investigate the cause of the energy derivative fluctuations, we selected two neighboring windows with the largest \(\partial V/\partial \lambda\) difference (13.8 kcal
mol$^{-1}$), corresponding to $\lambda = 0.875$ and 0.900 of the CBD–cellulose complex Y32A VDW alchemy step. It can be seen (Fig. S2) that the side-chain $\chi_2$ distribution is quite broad (ranging from $-180^\circ$ to $180^\circ$) for both windows, suggesting that tyrosine ring rotation occurs in the simulations, which can cause a dramatic fluctuation in the VDW interaction between the tyrosine and a glucose ring of the cellulose chain. Furthermore, a shift of the $\chi_1$ distribution was observed for the $\lambda = 0.875$ window compared to that for $\lambda = 0.900$ (Fig. S2). Thus, from a structural point of view, the bulky size of tyrosine and a slightly different local environment around the mutated site are the main cause of the large $\partial V/\partial \lambda$ fluctuation in the Y to AVDW alchemical process.

Total free energies were obtained by summing the contributions of all three steps. Table 1 shows the calculated and experimental relative free energy differences between the WT-CBD-Cel and MU-CBD-Cel. According to the TI simulations, the calculated $\Delta \Delta G_{\text{cal}}$ values for Q34A, N29A, Y5A, Y31A and Y32A are 3.42, 3.68, 5.40, 6.35 and 10.76 kcal mol$^{-1}$, respectively, suggesting that all five of these residues play roles in the binding to the cellulose surface. The corresponding standard deviations were 0.57, 0.74, 0.82, 0.59 and 0.87 kcal mol$^{-1}$, respectively.

The three aromatic residues Y5, Y31 and Y32 show larger binding free energy contributions than the residues N29 and Q34 do, which is in good agreement with the experimental results [4]. Among the five mutations, Y32A...
they should be higher than 1.74 kcal mol\(^{-1}\), the calculated glucose units. We repeated the TI calculations for Q34A removing the restraints, except for those on the terminal motions. The restraints on the C1 and C4 atoms also limit the conformational sampling of glucose units in contact with the cellulose, which is modeled here. However, as it is stated by Lehtio that “it cannot be ruled out that the initial, perhaps lower affinity, binding occurs in the (100) and (010) faces followed by tighter binding at the (110) plane as the hydrolysis proceeds,” the binding on other faces is most likely less sensitive to the mutations. (2) The cellulose model built here has an ideal flat surface composed of polyglucan chains; this is restrained during the TI simulations to simplify the system and yield better convergence; however, in reality, the surface may twist due to molecular motions. The restraints on the C1 and C4 atoms also limit the conformational sampling of glucose units in contact with the CBD side chains. This effect was studied by removing the restraints, except for those on the terminal glucose units. We repeated the TI calculations for Q34A and Y5A. The calculated \(\Delta G\) values decrease slightly, by 0.48 kcal mol\(^{-1}\) for the former and 1.08 kcal mol\(^{-1}\) for the latter—comparable to the uncertainties of the calculations (Table 1). This effect alone is unlikely to alter the predicted binding trend or account for the differences between experimental and computational \(\Delta G\) values. (3) The calculated binding free energies may also depend on the cellulose binding surface size. Presumably, the smaller surface should yield looser binding. This effect is not considered in our work. (4) The orientation of CBD relative to cellulose sampled in the MD simulations is limited. The TI calculations adopted here are very time-consuming and costly, impeding us from exploring this factor. (5) Force field errors for the cellulose may also contribute to the deviations. The GLYCAM06 force field has not been widely tested for binding free energy calculations of carbohydrates binding to protein receptors. Further investigation is needed to clarify these factors and yield quantitative agreement with the experimental data.

Although free energy decomposition is only an approximate process [35–37], it offers useful information about the molecular mechanism of the CBD–cellulose binding. The three-step TI simulation conveniently separates the electrostatic and VDW contributions to the free energy. We can see that \(\Delta G_{\text{vdw}}\) for the three aromatic residue mutations Y5A, Y31A and Y32A are 4.13, 4.39 and 8.97 kcal mol\(^{-1}\), respectively, corresponding to 76%, 69% and 83% of the total free energy changes, which indicates that the VDW interactions of the three residues play a more important role in the tight binding to cellulose than the electrostatic interactions. For the N29A and Q34A mutations, \(\Delta G_{\text{vdw}}\) accounts for 37% and 48% of the corresponding total free energy differences. Considering that tyrosine residues contribute more to the overall binding, we can conclude that VDW interactions are more important than electrostatics in the binding of CBD to cellulose.

However, if we integrate \(\partial V/\partial \alpha\) for the electrostatics and VDW separately (Fig. 5), it becomes apparent that the electrostatic change is much larger than the VDW change in the mutation to alanine for all five residues (Fig. 5) of the CBD–cellulose complex. This result underlines the importance of electrostatics in the CBD–cellulose binding, in concert with an early theoretical study of Cel7A CBD processivity [12] as well as a more recent quantum mechanics (QM) study of a tyrosine methyl \(\beta\)-glucopyranoside (OMG) complex, which suggests that dispersive interactions are less important than the hydrogen bond between the tyrosine OH and sugar O4 [38]. This observation seems to suggest that electrostatics is more important than VDW in the binding. However, according to the thermodynamic cycle (Fig. 2), the reference state (CBD alone in solution) must also be considered. In fact, the electrostatics dominates the overall free energy change (\(\Delta G_2\)) for the free CBD forms (Fig. 5). This electrostatics presumably arises from the change in solvation energy when mutating a polar residue (e.g., N or Y) to a nonpolar alanine. The comparable electrostatic change for a mutation in CBD and the CBD–cellulose complex largely cancels out. In another words, the solvation energy loss due to the binding of CBD to cellulose is compensated by the electrostatic interactions of CBD and cellulose. As a consequence, for \(\Delta G_2 - \Delta G_1\) (the binding free energy

### Table 1 Comparison of the predicted and experimental CBD–cellulose binding free energy differences of the single mutations. All values are in kcal mol\(^{-1}\)

| Mutation | \(\Delta G_{\text{vdw}}\) | \(\Delta G_{\text{ele}}\) | \(\Delta G_{\text{cal}}\) | \(\Delta G_{\text{exp}}\) |
|----------|-----------------|-----------------|-----------------|-----------------|
| Q34A     | 1.65±0.43       | 1.77±0.58       | 3.42±0.57       | 1.17            |
| N29A     | 1.35±0.41       | 2.32±0.67       | 3.68±0.74       | 0.57            |
| Y5A      | 4.13±0.73       | 1.27±0.65       | 5.40±0.82       | >1.74           |
| Y31A     | 4.39±0.64       | 1.96±0.75       | 6.35±0.59       | 1.74            |
| Y32A     | 8.97±0.24       | 1.78±0.77       | 10.76±0.87      | >1.74           |

The free energy of binding was deduced from the equation:
\[
\Delta G = -RT \ln\left(k_{\text{mut}}/k_{\text{wt}}\right)
\]
where \(k_{\text{mut}}\) and \(k_{\text{wt}}\) are the partition coefficients of the mutant and wild type. The values cannot be extrapolated from the adsorption isotherms, but they should be higher than 1.74 kcal mol\(^{-1}\), according to the adsorption isotherms [4].

shows the largest binding free energy loss. The NMR structure of the CBD suggests that the substitution at position 32 could cause significant disruption of the flat CBD cellulose contact face due to the close proximity of the neighboring side chains, which might account for the binding importance of this residue. In general, the absolute values of the free energy changes calculated from TI simulations are higher than the experimental values extrapolated from the adsorption isotherms. This may be due to a number of factors. (1) The binding of CBD to cellulose is very complex. Lehtio and coworkers [34] showed that the primary binding site is the (110) face of cellulose, which is modeled here. However, as it is stated by Lehtio that “it cannot be ruled out that the initial, perhaps lower affinity, binding occurs in the (100) and (010) faces followed by tighter binding at the (110) plane as the hydrolysis proceeds,” the binding on other faces is most likely less sensitive to the mutations. (2) The cellulose model built here has an ideal flat surface composed of polyglucan chains; this is restrained during the TI simulations to simplify the system and yield better convergence; however, in reality, the surface may twist due to molecular motions. The restraints on the C1 and C4 atoms also limit the conformational sampling of glucose units in contact with the CBD side chains. This effect was studied by removing the restraints, except for those on the terminal glucose units. We repeated the TI calculations for Q34A and Y5A. The calculated \(\Delta G\) values decrease slightly, by 0.48 kcal mol\(^{-1}\) for the former and 1.08 kcal mol\(^{-1}\) for the latter—comparable to the uncertainties of the calculations (Table 1). This effect alone is unlikely to alter the predicted binding trend or account for the differences between experimental and computational \(\Delta G\) values. (3) The calculated binding free energies may also depend on the cellulose binding surface size. Presumably, the smaller surface should yield looser binding. This effect is not considered in our work. (4) The orientation of CBD relative to cellulose sampled in the MD simulations is limited. The TI calculations adopted here are very time-consuming and costly, impeding us from exploring this factor. (5) Force field errors for the cellulose may also contribute to the deviations. The GLYCAM06 force field has not been widely tested for binding free energy calculations of carbohydrates binding to protein receptors. Further investigation is needed to clarify these factors and yield quantitative agreement with the experimental data.

Although free energy decomposition is only an approximate process [35–37], it offers useful information about the molecular mechanism of the CBD–cellulose binding. The three-step TI simulation conveniently separates the electrostatic and VDW contributions to the free energy. We can see that \(\Delta G_{\text{vdw}}\) for the three aromatic residue mutations Y5A, Y31A and Y32A are 4.13, 4.39 and 8.97 kcal mol\(^{-1}\), respectively, corresponding to 76%, 69% and 83% of the total free energy changes, which indicates that the VDW interactions of the three residues play a more important role in the tight binding to cellulose than the electrostatic interactions. For the N29A and Q34A mutations, \(\Delta G_{\text{vdw}}\) accounts for 37% and 48% of the corresponding total free energy differences. Considering that tyrosine residues contribute more to the overall binding, we can conclude that VDW interactions are more important than electrostatics in the binding of CBD to cellulose.

However, if we integrate \(\partial V/\partial \alpha\) for the electrostatics and VDW separately (Fig. 5), it becomes apparent that the electrostatic change is much larger than the VDW change in the mutation to alanine for all five residues (Fig. 5) of the CBD–cellulose complex. This result underlines the importance of electrostatics in the CBD–cellulose binding, in concert with an early theoretical study of Cel7A CBD processivity [12] as well as a more recent quantum mechanics (QM) study of a tyrosine methyl \(\beta\)-glucopyranoside (OMG) complex, which suggests that dispersive interactions are less important than the hydrogen bond between the tyrosine OH and sugar O4 [38]. This observation seems to suggest that electrostatics is more important than VDW in the binding. However, according to the thermodynamic cycle (Fig. 2), the reference state (CBD alone in solution) must also be considered. In fact, the electrostatics dominates the overall free energy change (\(\Delta G_2\)) for the free CBD forms (Fig. 5). This electrostatics presumably arises from the change in solvation energy when mutating a polar residue (e.g., N or Y) to a nonpolar alanine. The comparable electrostatic change for a mutation in CBD and the CBD–cellulose complex largely cancels out. In another words, the solvation energy loss due to the binding of CBD to cellulose is compensated by the electrostatic interactions of CBD and cellulose. As a consequence, for \(\Delta G_2 - \Delta G_1\) (the binding free energy
difference), the VDW contribution is larger and thus more important for binding. An early experimental isothermal titration microcalorimetry study of the binding of CBD from the β-1,4 exoglucanase Cex of *Cellulomonas fimi* to bacterial microcrystalline cellulose (BMCC) suggests that electrostatics does not drive the adsorption of CBDcex onto cellulose. CBDcex is larger than CBD Cel7A, but both have a flat surface composed of three aromatic residues (tyrosine or tryptophan). Our result is consistent with this experimental finding.

Our work confirms that the hydrophobic effects of the three tyrosine residues Y5, Y31 and Y32 that form the flat surface of the CBD are required for tight binding to cellulose. Experimental data show that the CBD binding free energy difference $\Delta G_{W5} - \Delta G_{Y5}$ is $-0.26$ kcal mol$^{-1}$, suggesting that the Y5W mutant has a higher binding affinity to cellulose [32]. Our computational values are $-0.24 \pm 0.55$ kcal mol$^{-1}$ for side-chain configuration 1 and $0.88 \pm 0.52$ kcal mol$^{-1}$ for configuration 2. While both configurations form a flat surface on contact with cellulose, configuration 1 is favored since it is more stable (Table 2). The free energy decomposition suggests that while the VDW interaction becomes stronger for the Y5W mutant, presumably due to its larger hydrophobic surface, the electrostatic interaction is weaker, probably because of the loss of the polar tyrosine hydroxyl group. Compared to the wild type, the Y31W mutant exhibits decreases in the binding free energy by $0.34 \pm 0.35$ and $0.62 \pm 0.40$ kcal mol$^{-1}$ for the two configurations. Similar to Y5W, the electrostatic and VDW changes offset each other. The Y32W mutation was not investigated, because substituting the Y32 phenol side chain by the larger indole group of tryptophan in the confined space could disturb the structure and make the surface less flat [39].

The experimental study of the impact of *Humicola grisea* CBHI CBD cellulose binding affinity on CBHI hydrolysis activity against cellulose has shown that the

### Table 2 Free energy differences of the single mutations Y5W or Y31W. All values are in kcal mol$^{-1}$. Two configurations are considered for each tryptophan side chain (Fig. 3)

| Mutation | $\Delta\Delta G_{\text{vdw}}$ | $\Delta\Delta G_{\text{ele}}$ | $\Delta\Delta G_{\text{cal}}$ |
|----------|-----------------|-----------------|-----------------|
| Y5W1     | $-1.83 \pm 0.41$ | $1.59 \pm 0.34$ | $-0.24 \pm 0.55$ |
| Y5W2     | $-0.75 \pm 0.54$ | $1.63 \pm 0.33$ | $0.88 \pm 0.52$  |
| Y31W1    | $-0.59 \pm 0.30$ | $0.93 \pm 0.10$ | $0.34 \pm 0.35$  |
| Y31W2    | $-0.79 \pm 0.51$ | $1.42 \pm 0.17$ | $0.62 \pm 0.40$  |

Fig. 6 Scatter plots of $\chi_1$ versus $\chi_2$ for the side chains of the five residues: a N29, b Q34, c Y5, d Y31, e Y32. The labels (1) and (2) stand for free and cellulose-bound CBD states, respectively.
tighter the binding, the higher the catalytic activity [39]. Our work suggests that one way to increase the binding affinity is to strengthen either the VDW or the electrostatic interactions, but not to sacrifice the other.

Changes in the dynamics of side chains due to binding

The protein side-chain dynamics represents a major component of protein conformational entropy [40]. Loss of side-chain conformational entropy could be an important influence that reduces a substrate’s binding affinity to an enzyme. The torsional entropy of side chains is thought to provide a good approximation for the side-chain entropy. To investigate the side-chain dynamics, we monitored the changes in side-chain dihedral angles and entropy of the five CBD residues before and after binding to cellulose. Figure 6 shows scatter plots of Chi1 (\(\chi_1\)) and Chi2 (\(\chi_2\)) for the five residues Y5, N29, Y31, Y32 and Q34 from the 20 ns free CBD and cellulose-bound CBD simulations. Table 3 lists the averages of and fluctuations in the \(\chi_1\) and \(\chi_2\) angles of the five residues in free and bound states. In general, as expected, the \(\chi_1\) and \(\chi_2\) values of the side-chain residues in the free CBD spread over a larger area than those of the cellulose-bound CBD, which indicates that the conformations of the side chains are restrained due to the binding to cellulose.

As shown in Fig. 6, after binding to cellulose, the number of the major side-chain states reduces from 5 and 3 to 1 for Tyr5 and Tyr31, respectively. The fluctuations in \(\chi_1\) and \(\chi_2\) for both Y5 and Y31 decease to \(\leq\)10° (Table 3). The considerably restricted side-chain rotations due to binding are further reflected in the notable drop in the entropic contribution of the side chain to the free energy (\(T\Delta S\), with differences between the ligand-free and bound states for Y5 and Y31 of −1.71 and −1.02 kcal mol\(^{-1}\), respectively (Fig. 7). Unlike Y5 and Y31, the distributions of \(\chi_1\) and \(\chi_2\) values for Y32 are similar in the free and bound states, with only a small fluctuation of −10° (Table 3), suggesting that this side chain is confined in both states. The \(T\Delta S\) value between the free and bound states is only −0.11 kcal mol\(^{-1}\) for this residue. In comparison, the side-chain torsional entropy losses for N29 and Q34 are slightly larger: −0.75 and −0.30 kcal mol\(^{-1}\), respectively (Fig. 7). It is worth noting that the entropy change considered here is very rough, and the sampling of the dihedral space with a 20 ns simulation is rather limited. However, qualitatively, it is clear that for some residues the entropy changes are relatively large and partially offset the gain in interaction energy after CBD binding.

Conclusions

In this paper, we studied the effect of a single mutation of Cel7A CBD on binding to the cellulose surface using a TI simulation method. Analyses of the computed free energy differences provided useful information for understanding the binding mechanism between CBD and cellulose. Our results suggest that residues Y5, Y31 and Y32 are more important than N29 and Q34 for the binding of CBD to cellulose. The VDW interactions between the amino acid Y5, Y31 and Y32 and the cellulose surface account for a large proportion of the binding affinity. Our calculations yielded relative binding affinities with a similar trend, but overestimated the experimental \(\Delta\Delta G\) data. The current model yields only qualitative agreement with experiments. A more realistic cellulose model with properly weighted CBD binding modes will likely yield a more quantitative agreement, which will be explored in a further study. The changes in the binding affinities of the mutants Y5W and Y31W to cellulose were calculated. The mutant Y5W shows slightly improved affinity, which is in agreement with the experimental results. Although the VDW interaction between CBD and cellulose is enhanced for the Y to W mutations, weakening of the electrostatics largely offsets this effect. The side-chain torsional entropy decreases for Y5, N29, Y31, Y32 and Q34 due to rotamer rigidification after binding to cellulose.

Table 3 Averages of and fluctuations in the angles \(\chi_1\) and \(\chi_2\) for the five residues in the free and bound states

| Residue | Free state | Bound state |
|---------|------------|-------------|
|         | \(\chi_1(\degree)\) | \(\chi_2(\degree)\) | \(\chi_1(\degree)\) | \(\chi_2(\degree)\) |
| N29     | 86±40      | 25±51       | −167±31          | −89±16          |
| Q34     | 93±56      | 178±11      | 76±41            | 177±11          |
| Y5      | −124±77    | −1±91       | −45±10           | 107±10          |
| Y31     | −45±26     | −22±83      | −68±9            | 80±8            |
| Y32     | −169±9     | −90±12      | −173±10          | −98±10          |

Fig. 7 Side-chain dihedral entropic contributions to the free energies TS (T=300 K) of the five residues in the free and cellulose-bound states. The difference between the two states for each residue is shown as a blue bar.
References

1. Henriissat B (1994) Cellulases and their interaction with cellulose. Cellulose 1:169–196
2. Teeri TT (1997) Crystalline cellulose degradation: new insight into the function of celllobiohydrolases. Trends Biotechnol 15:160–167
3. Divne C, Stahlberg J, Teeri TT, Jones TA (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 A long tunnel of celllobiohydrolase I from *Trichoderma reesei*. J Mol Biol 275:309–325
4. Linder M, Mattinen ML, Kontteli M, Lindeberg G, Stahlberg J, Drakenberg T, Reikainen T, Pettersson G, Annila A (1995) Identification of functionally important amino acids in the cellulose-binding domain of *Trichoderma reesei* celllobiohydrolase I. Protein Sci 4:1056–1064
5. Teeri TT, Koivula A, Linder M, Wohlfahrt G, Divne C, Jones TA (1998) *Trichoderma reesei* celllobiohydrolases: why so efficient on crystalline cellulose? Biochem Soc Trans 26:173–178
6. Tomme P, Van Tilbeurgh H, Pettersson G, van Damme J, Vandenkerckhove J, Knowles J, Teeri T, Claeyssens M (1998) Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two celllobiohydrolases by limited proteolysis. Eur J Biochem 170:575–581
7. Divne C, Stahlberg J, Reikainen T, Ruohonen L, Pettersson G, Knowles JK, Teeri TT, Jones TA (1994) The three-dimensional crystal structure of the catalytic core of celllobiohydrolase I from *Trichoderma reesei*. Science 265:524–528
8. Barr BK, Hsieh YL, Ganem B, Wilson DB (1996) Identification of two functionally different classes of exocellulases. Biochemistry 28:7241–7247
9. Imai T, Boiset C, Samejima M, Igarashi K, Sugiyama J (1998) Unidirectional processive action of celllobiohydrolase Cel7A on *Tolotia* cellulose microcrystals. FEBS Lett 423:113–116
10. Kraulis J, Clore GM, Nilges M, Jones TA, Pettersson G, Knowles J, Gronenborn AM (1989) Determination of the three-dimensional solution structure of the C-terminal domain of celllobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. Biochemistry 28:7241–7257
11. Bu L, Beckham GT, Crowley MF, Chang CH, Matthews JF, Bombre YJ, Adney WS, Himmel ME, Nimlos MR (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
12. Beckham GT, Matthews JF, Bombre YJ, Bu L, Adney WS, Himmel ME, Nimlos MR, Crowley MF (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
13. Nimlos MR, Matthews JF, Crowley ME, Walker RC, Chukkapalli G, Brady JW, Adney WS, Cleary JM, Zhong L, Himmel ME (2007) Molecular modeling suggests induced fit of Family I carbohydrate-binding modules with a break-chain cellulose surface. Protein Eng Des Sel 20:179–187
14. Steinbrecher T, Moley DL, Case DA (2007) Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. J Chem Phys 127:214108
15. Kirkwood JG (1935) Statistical mechanics of fluid mixtures. J Chem Phys 3:300–313
16. Sattelle BM, Sutcliffe MJ (2008) Calculating chemically accurate redox potentials for engineered flavoproteins from classical molecular dynamics free energy simulations. J Phys Chem A 112:13053–13057
17. Perez MA, Fernandes PA, Ramos MJ (2010) Substrate recognition in HIV-1 protease: a computational study. J Phys Chem B 114:2525–2532
18. Shirts MR, Pitera JW, Swope WC, Pande VS (2003) Extremely precise free energy calculations of amino acid side chain analogs: comparison of common molecular mechanics force fields for proteins. J Chem Phys 119:5740–5761
19. Straatsma TP, Mccammon JA (1992) Computational alchemy. Annu Rev Phys Chem 43:407–435
20. Cheng Y, Cheng X, Radic Z, Mccammon JA (2007) Acetylcholinesterase: mechanisms of covalent inhibition of wild-type and H447I mutant determined by computational analyses. J Am Chem Soc 129:6562–6570
21. Nam K, Verdine GL, Karplus M (2009) Analysis of an anomalous mutant of MutM DNA glycosylase leads to new insights into the catalytic mechanism. J Am Chem Soc 131:18208–18209
22. Seelig D, de Grooth BL (2010) Protein thermostability calculations using alchemical free energy simulations. Biophys J 98:2309–2316
23. Nishiyama Y, Langan P, Chanzy H (2002) Crystal structure and hydrogen-bonding system in cellulose 1 beta from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc 124:9074–9082
24. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Crowley MF, Walker RC, Zhang W, Merz KM, Wang BC, Hayik S, Roitberg A, Seabra G, Kolossvary I, Wong KE, Paesani F, Vainicke J, Wu X, Brozell SR, Steinbrecher T, Golikhe H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA (2008) AMBER 10. University of California, San Francisco
25. Hornak V, Abel R, Okar A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712–725
26. Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeiro J, Daniels CR, Foley BL, Woods RJ (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem 29:622–655
27. Kirschner KN, Woods RJ (2001) Solvent interactions determine carbohydrate conformation. Proc Natl Acad Sci USA 98:10541–10545
28. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593
29. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341
30. Ai R, Qaiser Fatmi M, Chang CE (2010) T-Analyst: a program for efficient analysis of protein conformational changes by torsion angles. J Comput Aided Mol Des 24:819–827
31. Yui T, Shibai H, Totsuzumi Y, Hayashi S, Miyata T, Hirata F (2010) Systematic docking study of the carbohydrate binding module of protein of Cel7A with the cellulose Ialpha crystal model. J Phys Chem B 114:49–58
32. Linder M, Lindeberg G, Reikainen T, Teeri TT, Pettersson G (1995) The difference in affinity between two fungal cellulose-
binding domains is dominated by a single amino acid substitution. FEBS Lett 372:96–98
33. Srisodsuk M, Lehtio J, Linder M, Margolles-Clark E, Reinikainen T, Teeri TT (1997) *Trichoderma reesei* cellobiohydrolase I with an endoglucanase cellulose-binding domain: action on bacterial microcrystalline cellulose. J Biotechnol 57:49–57
34. Lehtio J, Sugiyama J, Gustavsson M, Fransson L, Linder M, Teeri TT (2003) The binding specificity and affinity determinants of Family 1 and Family 3 cellulose binding modules. Proc Natl Acad Sci USA 100:484–489
35. Smith PE, van Gunsteren WF (1994) When are free energy components meaningful? J Phys Chem 98:13735–13740
36. Mark AE, van Gunsteren WF (1994) Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. J Mol Biol 240:167–176
37. Bresch S, Karplus M (1995) The meaning of component analysis: decomposition of the free energy in terms of specific interactions. J Mol Biol 254:801–807
38. Mohamed MN, Watts HD, Guo J, Catchmark JM, Kubicki JD (2010) MP2, density functional theory, and molecular mechanical calculations of C-H…π and hydrogen bond interactions in a cellulose-binding module-cellulose model system. Carbohydr Res 345:1741–1751
39. Takashima S, Ohno M, Hidaka M, Nakamura A, Masaki H (2007) Correlation between cellulose binding and activity of cellulose-binding domain mutants of *Humicola grisea* cellobiohydrolase I. FEBS Lett 581:5891–5896
40. Pickett SD, Sternberg MJ (1993) Empirical scale of side-chain conformational entropy in protein folding. J Mol Biol 231:825–839