Journal of Renewable Materials
DOI:10.32604/jrm.2020.010830
Article

Tech Science Press

Docking and Molecular Dynamics Study of the Carbohydrate Binding Module from *Trichoderma reesei* Cel7A on the Surfaces of the Cellulose III_I Crystal

Toshifumi Yui[1,*] and Takuya Uto[2]

[1]Faculty of Engineering, University of Miyazaki, Miyazaki, 889-2192, Japan
[2]Organization for Promotion of Tenure Track, University of Miyazaki, Miyazaki, 889-2192, Japan
*Corresponding Author: Toshifumi Yui. Email: tyui@cc.miyazaki-u.ac.jp
Received: 30 March 2020; Accepted: 11 May 2020

Abstract: We report the systematic survey of the binding free energies at the interface between a carbohydrate binding module (CBM) of Cel7A and the cellulose III_I crystal model using grid docking searches and molecular dynamics simulations. The two hydrophobic crystal surfaces were involved in the distinct energy minima of the binding free energy. The complex models, each with the CBM at the minimum energy position, stably formed in the solution state. The binding free energies of the cellulose III_I complex models, based on both static and dynamics states, were comparable to those of the native cellulose complex models. However, the cellulose III_I crystal had a larger binding surface, which is compatible with the observed high enzymatic activity of Cel7A for the cellulose III_I substrate.

Keywords: Carbohydrate binding module; Cel7A; cellulose III_I; docking analysis

1 Introduction

Cellulose forms various crystalline allomorphs originating from the two native crystalline phases classified as either cellulose I_α or I_β [1–3]. Among the cellulose allomorphs, cellulose III_I can be obtained by treatment of the cellulose I_β form with liquid ammonia or amines, and is readily converted back to the latter form by hot-water treatment. Conventional uses of cellulosic material are based on its highly crystalline nature, which makes it difficult to degrade into its component glucose residues. The most effective method to degrade crystalline cellulose is probably enzymatic degradation using cellulases, a generic term for enzymes hydrolyzing β-1,4-glucosidic linkages. Cellobiohydrolases (CBHs) can hydrolyze crystalline cellulose [4] to produce mainly cellobioses. They consist of a catalytic domain (CD) and a small carbohydrate binding module (CBM) connected by a highly glycosylated linker peptide [5–8]. The enzymes are adsorbed on the surface of crystalline cellulose at the initial stage of the reaction through a CBM, followed by hydrolysis by a CD at the substrate surface [9–11]. *Trichoderma reesei*, an asexually reproducing filamentous fungus and a clonal line derived from the tropical ascomycetes *Hypocrea jecorina* [12], secretes two CBHs (called Cel7A and Cel6A) along with several endoglucanases [13]. Igarashi and co-workers proposed that cellulose III_I is more susceptible to degradation by cellulases than native cellulose [14], and that Cel7A exhibits both higher affinity and specific activity for cellulose III_I, although the negative cooperative effect of its adsorption is higher on cellulose III_I than on

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

cellulose I_α [15]. The same group, using a high speed atomic force microscopy observation, revealed that molecular congestion, or "traffic jams", occur on the cellulose crystal surfaces [16]. More recently, it was proposed that higher susceptibility of cellulose III_I to Cel7A than that of cellulose I_α can be caused by the surface properties of the cellulose allomorphs at high enzymatic concentration of micromolar [17]. Ammonia treatment and the related procedures to increase the cellulose III_I phase are applicable to effectively enhance the enzymatic hydrolyzability of cellulosic biomass for bioethanol production and biorefining. Atomistic understanding of interfacial interactions between the CBM and the cellulose crystal will promote further improvement of the pretreatment procedures for a cellulose crystalline substrate.

The two-fold helix symmetry of a cellulose molecular chain constitutes a flat ribbon shape, where the two groups of the non-polar protons (H1, H3, and H5, and H2 and H4) alternatively appear on both faces of the ribbon and the polar functional groups align on both edges. The cellulose chains join together at the polar sides by intermolecular hydrogen bonds to form a cellulose chain sheet with hydrophobic faces. Consequently, the crystal structures of the native cellulose crystal are characterized by layered cellulose chain sheets joined by hydrophobic interactions (see Fig. 1b). We have reported systematic computational docking studies of Cel7A and Cel6A CBMs with respect to the cellulose I_α crystal surface [18,19]. The potential energy surfaces of the binding free energy between the CBM and the crystal surface exhibited two distinct potential wells on the (110) hydrophobic crystal surface, reflecting the two-fold helix symmetry of the cellulose chain. The complex models of the CBM bound at the minimum position of the cellulose crystal retained their binding interface during subsequent molecular dynamics (MD) simulation in the solution state. The present study reports similar docking searches and MD simulations of the Cel7A CBM with the cellulose III_I crystal surfaces as a comparative study with our previous studies of the cellulose I_α complex models [18].

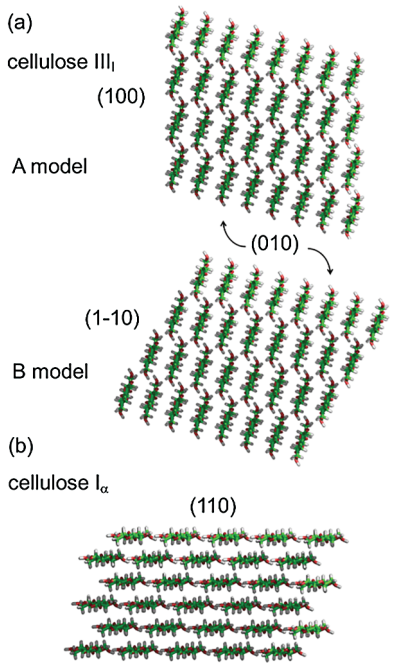


Figure 1: *ab* projections of (a) the two types of cellulose III_I and (b) the cellulose I_α crystal models. The Miller indexes of the constituent lattice planes are given

2 Computational Methods

2.1 Crystal Model Building

Fig. 1a shows the *ab* base planes of the cellulose III_I crystal models used in the present study, each of which (labeled as either the **A** or **B** model) comprises 32 cellulose chains with 20 repeating units (DP = 20). The cellulose chains were arranged to construct the crystal model based on a set of atomic coordinates and the crystal symmetry reported for the cellulose III_I crystal structure [20]. The cellulose I_α crystal model (Fig. 1b) was constructed in a similar fashion based on the crystallographic data [21]. The crystal model comprised six layers of (110) chain sheets containing five cellulose chains with DP = 20.

2.2 Program and Parameters for Energy Minimization and Molecular Dynamics

The minimizations and MD simulations were performed using the PMEMD module of the AMBER 11 package [22] combined with the ff99SB force field for a protein system [23,24] and the GLYCAM06 parameter set for the cellulose crystal system [25]. In the MD simulations, the 1–4 electrostatic (SCEE) and non-bonded (SCNB) scaling factors were set to unity for the cellulose crystal and the default values (SCEE = 1.2 and SCNB = 2.0) were used for the protein.

2.3 Grid Docking Search

A similar procedure to that of our previous study [18] was used for the present docking searches of Cel7A CBM to the cellulose III_I crystal surfaces. In brief, the three-dimensional (3D) data of Cel7A CBM, characterized by a wedge-like shape with three tyrosine residues (Tyr5, Tyr31, and Tyr32) at the substrate binding face, was obtained from the Protein Data Bank (1CBH) [26]. The CBM was docked to either the (100) or (010) surface of the A model, and to the (1–10) surface of the B model. The CBM was translated along both the fiber axis direction (x) and its vertical direction (y) in 0.5 Å steps over the ranges of 12 Å × 6 Å for the (010) surface and 12 Å × 9 Å for the (100) and (1–10) surfaces. At each grid point, the entire structure of the CBM–cellulose crystal complex was optimized with the implicit solvent of the generalized Born (GB) system [27,28] to evaluate the binding free energy between the CBM and the crystal surface. When a tip of the CBM is oriented towards the reducing terminal of a cellulose chain, the orientation is referred to as parallel (*P*), and when the CBM is oriented towards the non-reducing terminal it is referred to as antiparallel (*AP*).

2.4 Molecular Dynamics Simulations

The minimized complex models obtained from the grid docking search were placed in a rectangular periodic box filled with about 700–2,000 TIP3P water models [29] depending on the complex model. For the **A** crystal complex models, constant volume NVT dynamic simulations were carried out for 300 ps accompanied with a gradual increase of the temperature up to 300 K at a constant heating rate of 1 K/ps. The production NPT runs were performed for 50 ns at 1 bar and 300 K. The same heating and production MD procedures resulted in crystal deformation involving the chain sliding along the fiber axis [30,31] in some of the **B** crystal complex models. During the heating run from 200 to 300 K, a weak positional constraint with a constant force of 1 kcal/(mol $Å^2$) was imposed on all of the eight terminal residues of the four chains, each consisting of the crystal edge. The constraint force was then reduced to 0.1 kcal/(mol $Å^2$) during the following production run. The CBM–cellulose I_α complex models were constructed by placing the CBM at the minimum position of the (110) surface determined from the previous grid docking search [18]. The MD simulations of the complex models were carried out with a positional constraint of 50 kcal/(mol $Å^2$) imposed on the two inner chain layers to inhibit significant twisting of the crystal surfaces, which is an intrinsic feature of the native cellulose crystal models [32,33]. In the complex models, the NVT heating runs were carried out with a constant heating rate of 0.5 K/ps, followed by NPT production runs at 300 K and 1 bar for 50 ns.

Newton's equations of atomic motion were integrated by the Verlet algorithm [34] with a 2 fs time step combined with the SHAKE algorithm [35] to fix the bond stretching of the valence bonds involving hydrogen atoms. Non-bonding interactions were cut off at 10 Å, and the particle mesh Ewald method [36] was used to determine the long-range electrostatic interactions.

2.5 Thermodynamics Calculations

Variation of the binding free energy (ΔG_{bind}) between the CBM and the cellulose crystal surface was calculated based on the complex structures obtained from either the grid docking search or the MD trajectories by evaluating the non-covalent interaction terms of the AMBER potential energy [37] and the solvation free energy as follows:

$$\Delta G_{bind} = \Delta E_{nc,bind} + \Delta G_{solv,bind,} \quad (1)$$

where $\Delta E_{nc,bind}$ and $\Delta G_{solv,bind}$ represent the differences in non-covalent interactions and the solvation free energy on binding, respectively. The former term is also expressed in the difference in the gas phase energy, ($\Delta G_{gas,bind}$) so that

$$\Delta G_{bind} = \Delta G_{gas,bind} + \Delta G_{solv,bind.} \quad (2)$$

In the above expressions, the entropy contributions to the binding energy are ignored. For the change in the conformational entropy on binding, the side-chain orientations of Tyr31 and Tyr32 of the isolated CBM essentially exhibited a single conformation and that of Tyr5 only diverged. However, in our previous MD study of the Cel7A CBM and cellulose I_α complex models, we observed that the side chains of the three tyrosine residues were restricted to a single conformation in the complex structures and that the surface hydroxymethyl groups of the cellulose crystal models were allowed to rotate to the three representative conformations at the interfaces with the CBM [18]. Therefore, we assumed that the difference in the conformational entropy on binding may be insignificant. Changes in other entropic terms, such as translation, rotation, and vibration, were also assumed to be similar among the present complex models. The solvation free energy, which includes an entropic contribution related to water structuring, was approximated using implicit solvent models. Each of the ΔG_{solv} values was estimated throughout the MD trajectories using a combination of either the GB [27,28] or Poisson–Boltzmann (PB) [38] method with the surface area (SA) [39] option, and is subsequently referred to as either the GBSA or PBSA solvation free energy, respectively.

The ΔG_{bind} values were then decomposed into the contributions per side chain of each amino acid residue of the CBM. The 1–4 electrostatic and non-bonded interactions were included in the internal potential energy as calculating energy components.

These ΔG_{bind} calculations together with both the GBSA and PBSA methods and their decompositions were performed using the MM-PBSA.py module [40]. The default values of the option variables were used unless otherwise noted.

2.6 Programs for Visualization of Molecules

VMD 1.8 software was used for molecular visualization and animation of the trajectory data [41]. Molecular graphics were produced using PyMOL 1.5 software (Schödinger, LLC, Portland, OR, USA) .

3 Results and Discussion

3.1 Binding Free Energy Maps

Unlike the flat (110) chain sheets of the native cellulose crystal structures, the (100) and (1–10) lattice planes of the cellulose III_I models, each consisting of arrays of pyranose rings diagonally arranged with respect to the lattice plane, partly expose the polar functional groups (see Fig. 1), introducing a

hydrophilic nature to the corresponding crystal surfaces. The (1–10) surface appears to be more hydrophobic than the (100) surface because the non-polar pyranose faces of the former are slightly more exposed. Therefore, both the (100) and (1–10) surfaces can be characterized as "semi-hydrophobic" compared with the hydrophobic (110) surface of cellulose I_α. The (010) crystal surface is the most hydrophilic. Fig. 2 shows the GBSA binding free energy (ΔG^{GBSA}_{bind}) maps of both the *P* and *AP* orientations of the CBM derived from the grid docking search. The (010) crystal surface generated relatively shallow potential wells comprising multiple local minima. The ΔG^{GBSA}_{bind} maps of the (100) and (1–10) surfaces suggest higher affinity of CBM for these surfaces than the (010) surface, as indicated by the two distinct minima on the surfaces that are related to the two-fold helix symmetry of a cellulose chain along the fiber axis. The two minima are not strictly equivalent, reflecting an alternative arrangement of the two types of non-polar faces. Similar features were also observed for the hydrophobic (110) maps of the cellulose I_α crystal complex models [18,19].

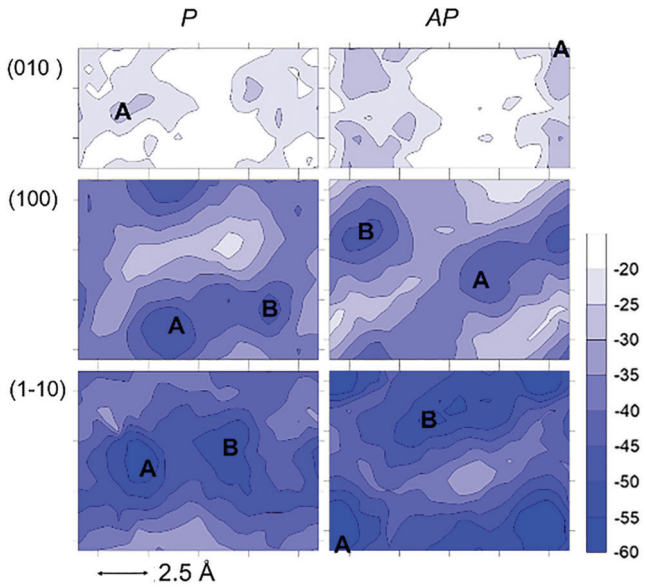


Figure 2: GBSA binding free energy maps with respect to the translational positions of the CBM in Å over the three crystal surfaces of (010), (100), and (1–10). The CBM is oriented in either a parallel (*P*) or antiparallel (*AP*) orientation with respect to the fiber axis of the cellulose chain. The potential energy surfaces are indicated by contour lines with an interval of 5 kcal/mol. The potential minima are denoted by either A or B

Tab. 1 compares the binding free energies (ΔG^{GBSA}_{bind}) at the minima and their canonical ensemble average values ($<\Delta G^{GBSA}_{bind}>$). The average values of the CBM–cellulose I_α complex models were calculated from the energy maps reported in our previous study [18]. From the $<\Delta G^{GBSA}_{bind}>$ values, among the cellulose III_I complex models, the CBM should most stably bind to the (1–10) surface followed by the (100) surface. Most importantly, the $<\Delta G^{GBSA}_{bind}>$ values of the former surface are comparable with those of the (110) surface of the cellulose I_α complex model. The results also indicate that the CBM prefers an *AP* direction of the cellulose III_I (1–10) and cellulose I_α (110) surfaces, which is consistent with the proposed processing direction of Cel7A [42].

3.2 Binding Free Energies of the MD Simulations

The cellulose III_I complex models with the CBM bound at the minimum energy position were then subjected to solvated MD simulations for 50 ns. The change of the binding free energies of the cellulose

Table 1: GBSA binding free energies of the minimum energy complex models and the ensemble average energies over the GBSA potential energy surfaces (kcal/mol)

Docking crystal surface	Orientation of CBM[a]	ΔG^{GBSA}_{bind} at the minimum energy position[b]		$<\Delta G^{GBSA}_{bind}>^{c}$
		A	B	
Cellulose III_I surfaces				
(010)	*P*	−29	—	−29
	AP	−32	—	−31
(100)	*P*	−50	−49	−49
	AP	−48	−45	−46
(1–10)	*P*	−59	−55	−58
	AP	−60	−57	−59
Cellulose I_α surface				
(110)	*P*	−56	−56	−55
	AP	−61	−57	−59

Notes: [a]*P*, parallel; *AP*, antiparallel.
[b]**A**, the lowest energy minimum; **B**, the second lowest energy minimum.
[c]Canonical ensemble average of GBSA binding free energies over all of the grid docking points calculated by

$$\langle \Delta G^{GBSA}_{bind} \rangle = \frac{\sum_i \Delta G^{GBSA}_{bind,i} e^{\Delta G^{GBSA}_{bind,i}/RT}}{\sum_i e^{\Delta G^{GBSA}_{bind,i}/RT}},$$

where *R* is the gas constant and *T* is the temperature (300 K).

III_I and I_α complex models with simulation time evaluated using the PBSA method (ΔG^{PBSA}_{bind}) are compared in Figs. 3 and 4, respectively. The changes of the two partitioned binding energies ($\Delta G_{gas,bind}$ and ΔG^{PBSA}_{bind}) along with ΔG^{PBSA}_{bind} are given in Appendix A1. The binding energies varied throughout the MD simulations as a result of structural changes caused by dynamic motion at the docking interfaces. The cellulose III_I (010) complex models exhibited positive values of ΔG^{PBSA}_{bind} throughout the MD simulations, as shown in Appendix A1, whereas the $\Delta G_{gas,bind}$ values remained negative. The CBMs were stably bound to all of the crystal surfaces, including the cellulose III_I (010) surface. The two *AP*-**A** complex models were more preferred not only because they corresponded to the global minimum models predicted by the binding free energy maps, but also because the directions of the CBM were consistent with the proposed processing direction of Cel7A. The ΔG^{PBSA}_{bind} energies fluctuated less within the values ranging from about −35 to −45 kcal/mol, and are comparable to those of the cellulose I_α *AP*-**A** complex models. In our previous study of the cellulose I_α complex models, we proposed the distinctive stability of the *AP*-**A** complex model in the 5 ns MD simulations. The extended MD simulations shown in Fig. 4 revealed crossing of the ΔG^{PBSA}_{bind} energies during about 20–35 ns. It should be noted that the $\Delta G_{gas,bind}$ value of the *AP*-**A** complex model shown in Appendix A1 was lower than that of the *P*-**A** complex model during most of the MD simulation.

From the energies in Appendix A1, exothermic $\Delta G_{gas,bind}$ and endothermic $\Delta G_{solv,bind}$ contribute to ΔG_{bind}. The $|\Delta G_{gas,bind}|$ and $|\Delta G_{solv,bind}|$ values ranged from about 50 to 100 kcal/mol and from 100 to 150 kcal/mol, respectively, which resulted in a typical exothermic ΔG_{bind} value of about −50 kcal/mol. In contrast to $\Delta G_{gas,bind}$ where $\Delta E_{nc,bind}$ is derived from the well-established force field functions and parameters, estimation of $\Delta G_{solv,bind}$ requires a certain approximation. The GB and PB solvents used in the present study are implicit solvent models to estimate the solvent-mediated electrostatic and the non-electrostatic interactions, consisting of energetically favorable van der Waals interactions and the

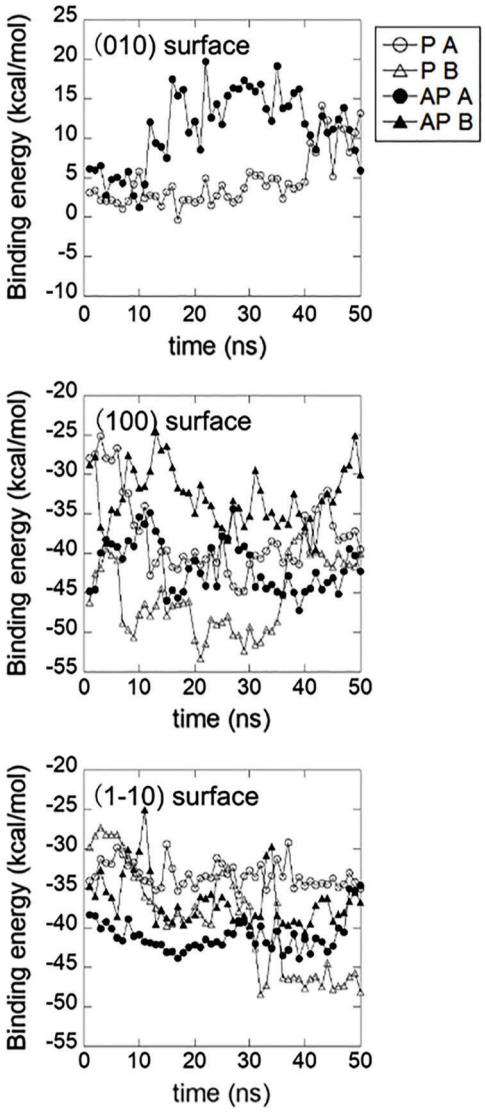


Figure 3: Trajectories of the PBSA binding free energy (ΔG_{bind}) calculated for the complex models with the CBM at the minimum position on the (010), (100), and (1–10) surfaces of the cellulose III_I crystal models

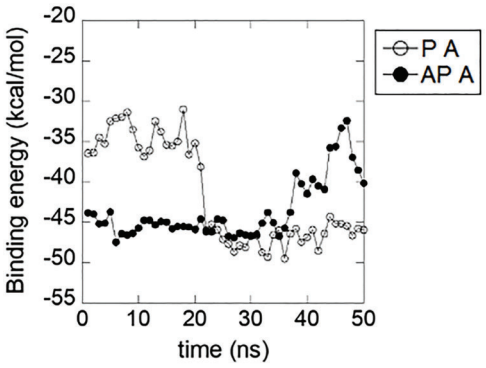


Figure 4: PBSA binding free energy calculated for the complex models with the CBM at the minimum energy position on the (110) surface of the cellulose I_α crystal model plotted against simulation time

energetically unfavorable cost of deforming the water structuring, and are modeled with a term proportional to the solvent accessible area. From the results of MD simulations, Brady and co-workers suggested that hydrophobic hydration of such extended non-polar surfaces significantly contributes to the binding between the non-polar faces of a D-glucose residue and the planer aromatic groups [43–45]. This may limit the validity and accuracy of the solvation free energy predicted by the implicit solvent model methods, and the methods tended to underestimate the endothermic solvation effect, resulting in a large exothermic value of ΔG_{bind}. In fact, the ΔG_{bind} values predicted for the interactions between D-glucose and the planer aromatic groups were about −2.5 to −10 kcal/mol and −0.5 kcal/mol from MD simulations with an umbrella sampling [43] and estimation of the equilibrium constant of binding [44], respectively. Our previous MD simulations of the Cel7A CBM–cellulose I_α complex models predicted much smaller ΔG_{bind} values of about −20 kcal/mol using a 3D reference interaction site model theory [46–48] to estimate the solvation free energy than those predicted using the GBSA method [18].

Appendix A2 shows the CBM structures obtained at either the initial (0 ns) or final (50 ns) frame of the MD simulations of the cellulose III_I complex models superimposed on the cellulose surfaces. Despite the large amount of variations in the $\Delta G^{\text{PBSA}}{}_{\text{bind}}$ energies in Fig. 3, movement of the CBM appears to be slight on each of the crystal surfaces. Most of the CBMs rotated with respect to the normal axis to the crystal surface with a negligible amount of translation, except for the (1−10) *AP*-**B** complex model that moved towards the top right in Appendix A2. Tab. 2 lists the root-mean square deviation (RMSD) values of the CBM backbone atoms between the initial and final MD structures. Fitting and non-fitting values of the RMSD indicate some backbone deformation and that combined with relative movement of the CBM, respectively. Larger fitting values were observed for the *P* and *AP* CBMs at the **B** minima on the cellulose III_I (100) surface than at the **A** minima. The CBM of the cellulose I_α *AP*-**A** complex model deformed the least in terms of both the fitting and non-fitting RMSD values. The fitting value of the cellulose I_α *P*-**A** complex model was even smaller than those of most of the cellulose III_I complex

Table 2: RMSD values of the CBM backbone structures between the initial (0 ns) and final (50 ns) frames of the MD simulations (Å)

Docking crystal surface	Orientation of CBM[a]	Minimum energy position[b]	RMSD (Å)	
			Fitting[c]	Non-fitting[d]
Cellulose III_I complex models				
(100)	*P*	**A**	0.9	2.8
		B	1.9	2.7
	AP	**A**	0.7	2.9
		B	1.5	4.6
(1–10)	*P*	**A**	1.0	2.6
		B	1.1	2.7
	AP	**A**	1.0	2.0
		B	0.7	4.1
Cellulose I_α complex models				
(110)	*P*	**A**	0.8	3.8
	AP	**A**	0.7	1.4

Notes: [a]See the footnote *a* of Tab. 1 for the definition of orientation of CBM.
[b]See the footnote *b* of Tab. 1 for the minimum position.
[c]Amount of deformation of the CBM backbone structure.
[d]Amount of the backbone deformation, rotation, and translation of the CBM from the initial position.

models. Compared with the hydrophobic flat surfaces of the native cellulose crystal, the cellulose III_I crystal has corrugated surface topologies consisting of arrays of polar functional groups and hydrophobic pyranose rings of the CBM. Such backbone deformation and positional deviation of the CBM can also be interpreted as a sort of induced fit of a non-inherent substrate of Cel7A to the cellulose III_I crystal surface.

3.3 Decomposition of the Binding Free Energies on a Per Side Chain Basis

Decomposing ΔG^{PBSA}_{bind} in terms of the contributions from structural subunits provides the source of the binding interactions on a residue basis [49]. Fig. 5 shows the decomposition of the ΔG^{PBSA}_{bind} values on a per side chain basis for the amino acid residues with $|\Delta G^{PBSA}_{bind}| \geq 1$ kcal/mol for all of the cellulose III_I

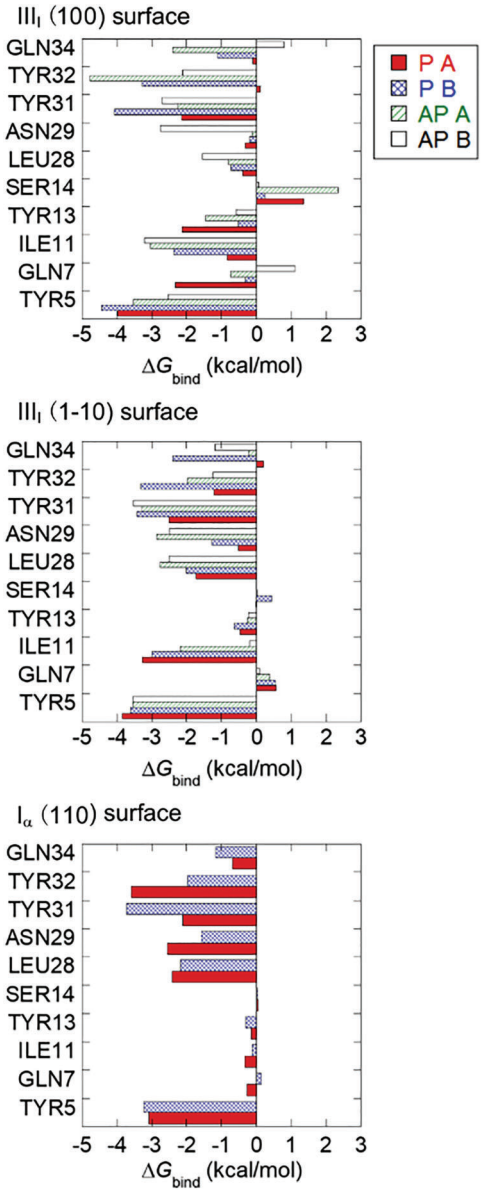


Figure 5: Decomposition of the PBSA binding free energy on a per residue basis for residues with $|\Delta G_{bind}^{PBSA}| \geq 3$ kcal/mol for the complex models with the CBM at the minimum energy position on the (100) and (1–10) surfaces of the cellulose III_I crystal models, and the (110) surface of the cellulose I_α crystal model

(100) and (1–10), and cellulose I_α complex models. The side chains of the three linearly aligned tyrosine residues on the docking face (Tyr5, Tyr31, and Tyr32) were the three largest exothermic contributions to the total ΔG^{PBSA}_{bind} in all of the complex models. The additional exothermic contributions with $|\Delta G^{PBSA}_{bind}| > 1$ kcal/mol were found at the side chains of Leu28 for the CBMs of all of the cellulose I_α and cellulose III_I (1−10) complex models and the cellulose III_I (100) *AP*-**B** complex model. Similarly, the side chains of Asn29 contributed significant exothermic interactions to the CBMs of the cellulose I_α and some of the cellulose III_I complex models. The side chains of Ile11 did not significantly interact with the cellulose I_α crystal surface, but the binding energies of Ile11 were similar to those of the three tyrosine side chains for most of the cellulose III_I (100) and (1–10) complex models. Although more diverse binding was found in the cellulose III_I complex models, the three tyrosine side chains contributed the most to ΔG^{PBSA}_{bind} through hydrophobic stacking between the aromatic side chains and pyranose rings irrespective of the cellulose crystal surface.

4 Conclusions

The main objective of the present study was to compare the substrate binding specificities of the CBM of Cel7A to cellulose III_I with that to the native cellulose crystal. The docking search and MD simulation results suggested that the (100) and (1–10) crystal surfaces of the cellulose III_I crystal contained distinct binding free energy minima as well as affinities comparable to the hydrophobic surface of native cellulose. This enables the CBM to effectively bind to the crystal substrate upon enzymatic degradation [14]. The cellulose III_I crystal, which has two binding surfaces for the CBM, can more effectively form a complex with the CBM than the native cellulose crystal, which has only one binding surface. Igarashi et al. [16] suggested that the cellulose III_I fibril is involved in a larger number of processive lanes than cellulose I_α based on high-speed atomic force microscopy observations of the enzymatic degradation of cellulose I_α and III_I samples by Cel7A. The additional feature suggested by the docking search was a slight preference for the *AP* direction of the CBMs on the cellulose III_I (1–10) surfaces, as was observed in our docking study of the cellulose I_α complex models [18]. However, in the following MD simulations, dynamic motion and solvation by explicit water molecules caused significant variation in the ΔG_{bind} energies, making the directional preference of the CBM less distinct. Some of the cellulose III_I complex models showed slight but significant movement of the CBM accompanied with its backbone deformation, probably as a result of induced fit. Decomposition analyses of ΔG^{PBSA}_{bind} suggested that the three linearly aligned tyrosine side chains play a significant role in the binding of the CBM to the cellulose III_I (100) and (1–10) crystal surfaces, as well as the cellulose I_α crystal surface. The results are consistent with an adsorption study of the fusion protein of Cel7A CBM with a red-fluorescent protein [15]. The present decomposition analysis also showed more diverse interactions involving additional amino acid residues at the binding interfaces of cellulose III_I.

Chen et al. [50] reported a systematic study of the effects of *O*-mannosylation on Cel7A CBM function. The 20 CBMs synthesized with mono-, di-, or trisaccharide at each of the three glycosylation sites (Thr1, Ser3, and Ser14) and their binding affinity, proteolytic stability, and thermostability were compared. It was suggested that *O*-linked mannoses at the three sites synergistically enhanced the binding affinity of the CBM to crystalline cellulose substrate. A theoretical study reported by the same group proposed the mannosylation effect to the CBM by systematic thermodynamic integration combined with long MD simulations of the various CBMs [51]. It is likely that a mannosylation also affects the binding affinity of the CBM to the cellulose III_I crystal surfaces.

Acknowledgement: The calculations were partly performed using Research Center for Computational Science, Okazaki Research Facilities, National Institutes of Natural Sciences (NINS), Japan.

Funding Statement: This work was supported by JSPS KAKENHI Grant Number 17K00409.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

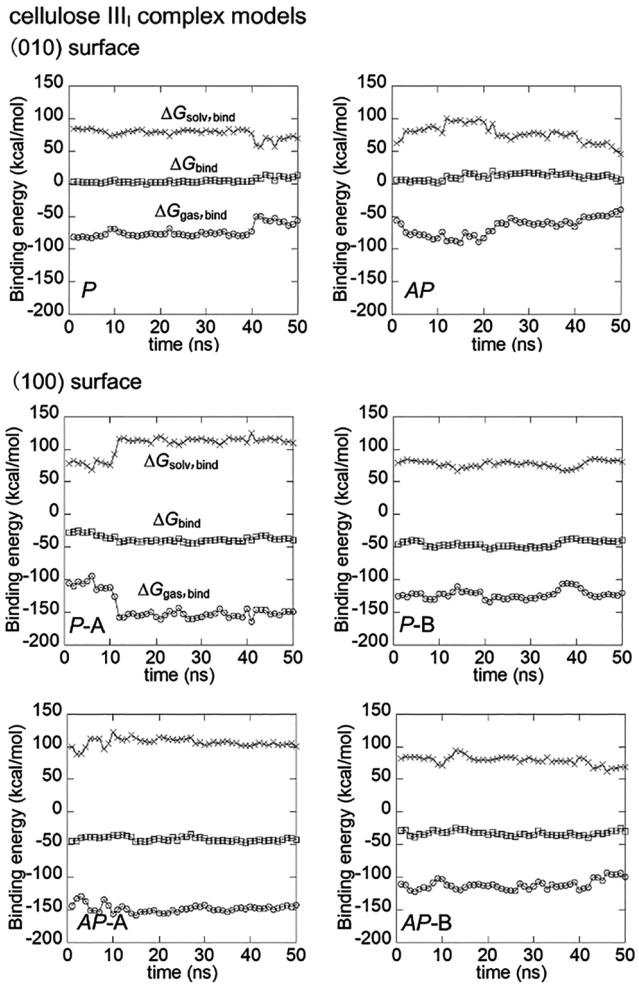
References

1. Atalla, R. H., Vanderhart, D. L. (1984). Native cellulose: a composite of two distinct crystalline forms. *Science, 223(4633),* 283–285. DOI 10.1126/science.223.4633.283.
2. VanderHart, D. L., Atalla, R. H. (1984). Studies of microstructure in native celluloses using solid-state carbon-13 NMR. *Macromolecules, 17(8),* 1465–1472. DOI 10.1021/ma00138a009.
3. Sugiyama, J., Vuong, R., Chanzy, H. (1991). Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. *Macromolecules, 24(14),* 4168–4175. DOI 10.1021/ma00014a033.
4. Teeri, T. (1997). Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends in Biotechnology, 15(5),* 160–167. DOI 10.1016/S0167-7799(97)01032-9.
5. Abuja, P. M., Schmuck, M., Pilz, I., Tomme, P., Claeyssens, M. et al. (1988). Structural and functional domains of cellobiohydrolase I from *Trichoderma reesei*. *European Biophysics Journal, 15(6),* 339–342. DOI 10.1007/BF00254721.
6. Johansson, G., Ståhlberg, J., Lindeberg, G., Engström, Å., Pettersson, G. (1989). Isolated fungal cellulose terminal domains and a synthetic minimum analogue bind to cellulose. *FEBS Letters, 243(2),* 389–393. DOI 10.1016/0014-5793(89)80168-1.
7. Shoemaker, S., Schweickart, V., Ladner, M., Gelfand, D., Kwok, S. et al. (1983). Molecular cloning of exo-cellobiohydrolase I derived from *Trichoderma reesei* strain-L27. *Nature Biotechnology, 1(8),* 691–696. DOI 10.1038/nbt1083-691.
8. Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J. et al. (1988). Studies of the cellulolytic system of *Trichoderma reesei* QM, 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. *European Journal of Biochemistry, 170(3),* 575–581. DOI 10.1111/j.1432-1033.1988.tb13736.x.
9. Lee, Y., Fan, L. T. (1982). Kinetic studies of enzymatic hydrolysis of insoluble cellulose: analysis of the initial rates. *Biotechnology and Bioengineering, 24(11),* 2383–2406. DOI 10.1002/bit.260241107.
10. Lee, Y. H., Fan, L. T. (1983). Kinetic studies of enzymatic hydrolysis of insoluble cellulose: (II). Analysis of extended hydrolysis times. *Biotechnology and Bioengineering, 25(4),* 939–966. DOI 10.1002/bit.260250406.
11. Ståhlberg, J., Johansson, G., Pettersson, G. (1991). A new model for enzymatic hydrolysis of cellulose based on the two-domain structure of cellobiohydrolase I. *Nature Biotechnology, 9(3),* 286–290. DOI 10.1038/nbt0391-286.
12. Kuhls, K., Lieckfeldt, E., Samuels, G. J., Kovacs, W., Meyer, W. et al. (1996). Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina*. *Proceedings of the National Academy of Sciences of the United States of America, 93(15),* 7755–7760. DOI 10.1073/pnas.93.15.7755.
13. Wood, T. M. (1992). Fungal cellulases. *Biochemical Society Transactions, 20(1),* 46–53. DOI 10.1042/bst0200046.
14. Igarashi, K., Wada, M., Samejima, M. (2007). Activation of crystalline cellulose to cellulose III_I results in efficient hydrolysis by cellobiohydrolase. *FEBS Journal, 274(7),* 1785–1792. DOI 10.1111/j.1742-4658.2007.05727.x.
15. Sugimoto, N., Igarashi, K., Wada, M., Samejima, M. (2012). Adsorption characteristics of fungal family 1 cellulose-binding domain from *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose: negative cooperative adsorption via a steric exclusion effect. *Langmuir, 28(40),* 14323–14329. DOI 10.1021/la302352k.
16. Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S. et al. (2011). Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science, 333(6047),* 1279–1282. DOI 10.1126/science.1208386.
17. Shibafuji, Y., Nakamura, A., Uchihashi, T., Sugimoto, N., Fukuda, S. et al. (2014). Single-molecule imaging analysis of elementary reaction steps of *Trichoderma reesei* cellobiohydrolase I (Cel7A) hydrolyzing crystalline cellulose Iα and III_I. *Journal of Biological Chemistry, 289(20),* 14056–14065. DOI 10.1074/jbc.M113.546085.

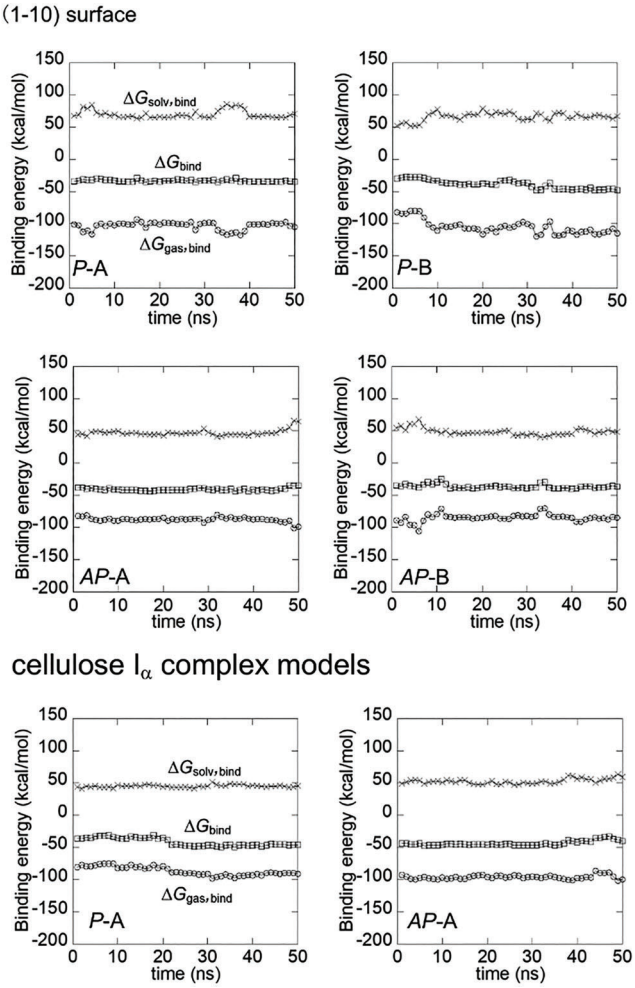
18. Yui, T., Shiiba, H., Tsutsumi, Y., Hayashi, S., Miyata, T. et al. (2010). Systematic docking study of the carbohydrate binding module protein of Cel7A with the cellulose Iα crystal model. *Journal of Physical Chemistry B, 114(1),* 49–58. DOI 10.1021/jp908249r.
19. Shiiba, H., Hayashi, S., Yui, T. (2012). Molecular simulation study with complex models of the carbohydrate binding module of Cel6A and the cellulose Iα crystal. *Cellulose, 19(3),* 635–645. DOI 10.1007/s10570-012-9671-y.
20. Wada, M., Chanzy, H., Nishiyama, Y., Langan, P. (2004). Cellulose III_I crystal structure and hydrogen bonding by synchrotron X-ray and neutron fiber diffraction. *Macromolecules, 37(23),* 8548–8555. DOI 10.1021/ma0485585.
21. Nishiyama, Y., Sugiyama, J., Chanzy, H., Langan, P. (2003). Crystal structure and hydrogen bonding system in cellulose Iα from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society, 125(47),* 14300–14306. DOI 10.1021/ja037055w.
22. Case, D. A., Darden, T. A., Cheatham, T. E. III., Simmerling, C. L., Wang, J. et al. (2010). *Amber 11*. San Francisco: University of California.
23. Wang, J., Cieplak, P., Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry, 21,* 1049–1074. DOI 10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F.
24. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. et al. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics, 65(3),* 712–725. DOI 10.1002/prot.21123.
25. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R. et al. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *Journal of Computational Chemistry, 29 (4),* 622–655. DOI 10.1002/jcc.20820.
26. Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G. et al. (2002). 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(18),* 7241–7257. DOI 10.1021/bi00444a016.
27. Hawkins, G. D., Cramer, C. J., Truhlar, D. G. (1995). Pairwise solute descreening of solute charges from a dielectric medium. *Chemical Physics Letters, 246(1–2),* 122–129. DOI 10.1016/0009-2614(95)01082-K.
28. Hawkins, G. D., Cramer, C. J., Truhlar, D. G. (1996). Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *Journal of Physical Chemistry, 100(51),* 19824–19839. DOI 10.1021/jp961710n.
29. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics, 79(2),* 926–935. DOI 10.1063/1.445869.
30. Yui, T., Hayashi, S. (2008). Structural stability of the solvated cellulose III_I crystal models: a molecular dynamics study. *Cellulose, 16(2),* 151–165. DOI 10.1007/s10570-008-9265-x.
31. Yui, T., Okayama, N., Hayashi, S. (2010). Structure conversions of cellulose III_I crystal models in solution state: a molecular dynamics study. *Cellulose, 17(4),* 679–691. DOI 10.1007/s10570-010-9422-x.
32. Matthews, J. F., Skopec, C. E., Mason, P. E., Zuccato, P., Torget, R. W. et al. (2006). Computer simulation studies of microcrystalline cellulose Iβ. *Carbohydrate Research, 341(1),* 138–152. DOI 10.1016/j.carres.2005.09.028.
33. Yui, T., Nishimura, S., Akiba, S., Hayashi, S. (2006). Swelling behavior of the cellulose Iβ crystal models by molecular dynamics. *Carbohydrate Research, 341(15),* 2521–2530. DOI 10.1016/j.carres.2006.04.051.
34. Verlet, L. (1967). Computer "Experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review, 159(1),* 98–103. DOI 10.1103/PhysRev.159.98.
35. Ryckaert, J. P., Ciccotti, G., Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics, 23(3),* 327–341. DOI 10.1016/0021-9991(77)90098-5.
36. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. et al. (1995). A smooth particle mesh Ewald method. *Journal of Chemical Physics, 103(19),* 8577–8593. DOI 10.1063/1.470117.

37. Wang, J., Morin, P., Wang, W., Kollman, P. A. (2001). Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *Journal of the American Chemical Society, 123(22),* 5221–5230. DOI 10.1021/ja003834q.
38. Luo, R., David, L., Gilson, M. K. (2002). Accelerated poisson-boltzmann calculations for static and dynamic systems. *Journal of Computational Chemistry, 23(13),* 1244–1253. DOI 10.1002/jcc.10120.
39. Weiser, J., Shenkin, P. S., Still, W. C. (1999). Approximate solvent-accessible surface areas from tetrahedrally directed neighbor densities. *Biopolymers, 50,* 373–380. DOI 10.1002/(SICI)1097-0282(19991005)50:4<373::AID-BIP3>3.0.CO;2-U.
40. Miller, B. R., McGee, T. D. Jr., Swails, J. M., Homeyer, N., Gohlke, H. et al. (2012). MMpbsa.py: an efficient program for end-state free energy calculations. *Journal of Chemical Theory and Computation, 8(9),* 3314–3321. DOI 10.1021/ct300418h.
41. Humphrey, W., Dalke, A., Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics, 14(1),* 33–38. DOI 10.1016/0263-7855(96)00018-5.
42. Chanzy, H., Henrissat, B. (1985). Undirectional degradation of *Valonia* cellulose microcrystals subjected to cellulase action. *FEBS Letters, 184(2),* 285–288. DOI 10.1016/0014-5793(85)80623-2.
43. Wohlert, J., Schnupf, U., Brady, J. W. (2010). Free energy surfaces for the interaction of D-glucose with planar aromatic groups in aqueous solution. *Journal of Chemical Physics, 133(15),* 155103. DOI 10.1063/1.3496997.
44. Mason, P. E., Lerbret, A., Saboungi, M. L., Neilson, G. W., Dempsey, C. E. et al. (2011). Glucose interactions with a model peptide. *Proteins: Structure, Function, and Bioinformatics, 79(7),* 2224–2232. DOI 10.1002/prot.23047.
45. Brady, J. W., Tavagnacco, L., Ehrlich, L., Chen, M., Schnupf, U. et al. (2012). Weakly hydrated surfaces and the binding interactions of small biological solutes. *European Biophysics Journal, 41(4),* 369–377. DOI 10.1007/s00249-011-0776-2.
46. Kovalenko, A., Hirata, F. (1998). Three-dimensional density profiles of water in contact with a solute of arbitrary shape: a RISM approach. *Chemical Physics Letters, 290(1–3),* 237–244. DOI 10.1016/S0009-2614(98)00471-0.
47. Kovalenko, A., Hirata, F. (2000). Potentials of mean force of simple ions in ambient aqueous solution. I. Three-dimensional reference interaction site model approach. *Journal of Chemical Physics, 112(23),* 10391–10402. DOI 10.1063/1.481676.
48. Kovalenko, A., Hirata, F. (2000). Potentials of mean force of simple ions in ambient aqueous solution. II. Solvation structure from the three-dimensional reference interaction site model approach, and comparison with simulations. *Journal of Chemical Physics, 1121(23),* 10403–10417. DOI 10.1063/1.481677.
49. Gohlke, H., Kiel, C., Case, D. A. (2003). Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes. *Journal of Molecular Biology, 330 (4),* 891–913. DOI 10.1016/S0022-2836(03)00610-7.
50. Chen, L., Drake, M. R., Resch, M. G., Greene, E. R., Himmel, M. E. et al. (2014). Specificity of O-glycosylation in enhancing the stability and cellulose binding affinity of Family 1 carbohydrate-binding modules. *Proceedings of the National Academy of Sciences, 111(21),* 7612–7617. DOI 10.1073/pnas.1402518111.
51. Taylor, C. B., Talib, M. F., McCabe, C., Bu, L., Adney, W. S. et al. (2012). Computational investigation of glycosylation effects on a Family 1 carbohydrate-binding module. *Journal of Biological Chemistry, 287(5),* 3147–3155. DOI 10.1074/jbc.M111.270389.

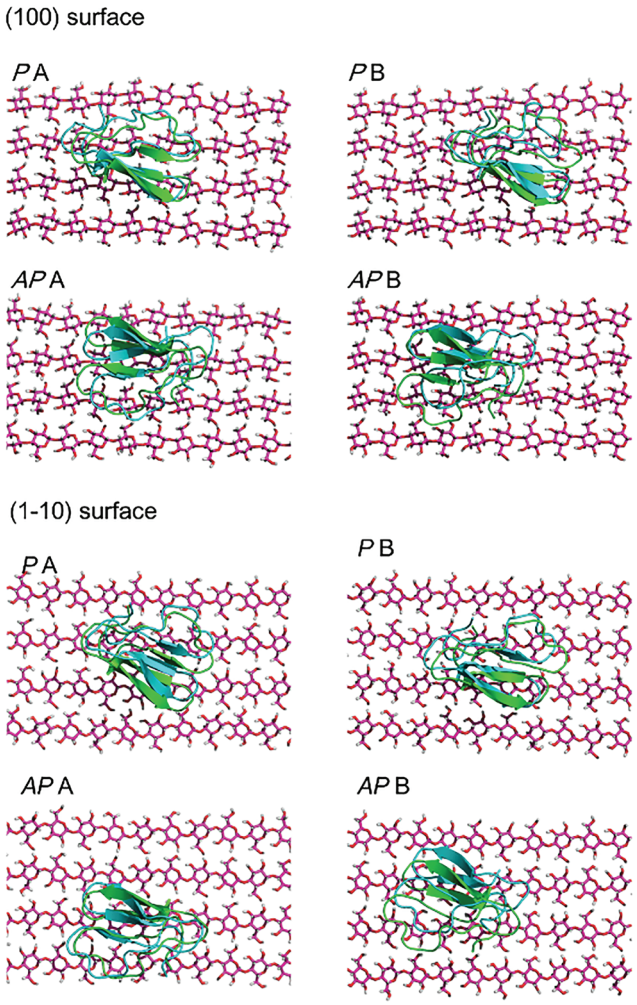
Appendix A.



Appendix A1: (Continued)



Appendix A1: Change of the PBSA binding free energy (ΔG_{bind}) and its solvation ($\Delta G_{solv,bind}$) and gas phase ($\Delta G_{gas,bind}$) components with simulation time for the complex models with the CBM at the minimum position on the (100) and (1–10) surfaces of the cellulose III_I crystal models



Appendix A2: CBM structures at 0 (light green) and 40 ns (cyan) superimposed on the surface layer of cellulose chains. *P*: parallel and *AP*: antiparallel orientations.