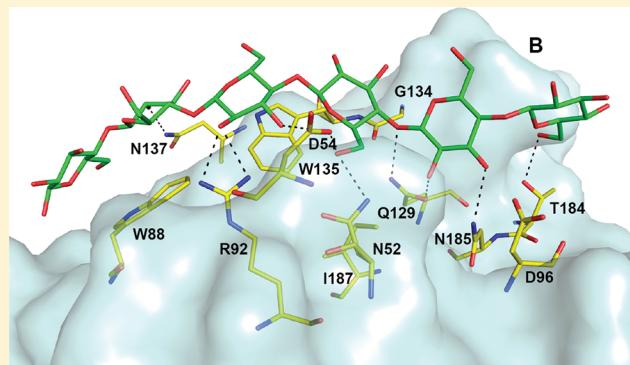


Recognition of Cello-Oligosaccharides by CBM17 from *Clostridium cellulovorans*: Molecular Dynamics Simulation

Ruyin Cao, Yongdong Jin, and Dingguo Xu*

Key Laboratory of Green Chemistry & Technology, Ministry of Education, College of Chemistry, Sichuan University, Chengdu, Sichuan 610064, P. R. China

ABSTRACT: The basic function of carbohydrate binding module (CBM) is believed to enhance local concentration of glycosidases on the carbohydrate molecule, and thus facilitates the subsequent degradation of carbohydrate. Full understanding of the recognition mechanism of carbohydrates by CBM can be helpful to enhance the enzyme activity. In this work, the detailed recognition specificity of two soluble cello-oligosaccharide substrates, cellobetaose and cellohexaose, by a family 17 CBM from *Clostridium cellulovorans* was investigated by molecular dynamics simulation. Calculated binding free energies using molecular mechanics/generalized Born and surface area (MM/GBSA) approach are in excellent agreement with experimental values. Overall, based on the decomposition of total binding free energy, nonpolar terms are shown to have favorable contributions to the binding, while polar interactions make unfavorable contributions, no matter significant hydrogen bond network is formed between substrate and protein. On the basis of computational alanine scanning and per-residue free energy decomposition, Trp88 and Trp135 are shown to be two most important residues in the cellobetaose binding mainly via hydrophobic interactions. The calculated subtotal contributions for those polar residues, D54, R92, Q129, and N185, can compare very well with experimental data.



1. INTRODUCTION

Enzymatic degradation of β -1,4-glycosidic bond is one of the most important reactions on earth. Its application has been extended to many fields, such as paper, food, pharmaceutical, and biofuel industries. Those enzymes that can catalyze the cleavage of glycosidic bond have been grouped into glycoside hydrolase (GH) superfamily, which has more than 110 subfamily members to date. Glycoside hydrolase is a kind of modular enzyme, which usually contains three modules, catalytic domain (CD), noncatalytic carbohydrate binding module (CBM) and a linker region connecting CD and CBM.

CBMs are currently classified into 64 families in the CAZy database (<http://www.cazy.org/Carbohydrate-Binding-Modules.html>)¹ on the basis of their amino acid sequences. To our best knowledge, there are over 200 PDB entries that can be located with the survey of “carbohydrate binding module” in protein data bank so far. In order to understand detailed contributions to the catalysis and corresponding binding pattern of CBMs, there are a lot of crystallographic structures complexed with various oligosaccharides have been reported, e.g., family 2,² 6,^{3–5} 9,⁶ 17,⁷ 22,⁸ 29,⁹ 32,¹⁰ 35,¹¹ 61,¹² 62,¹³ etc. These structures can provide valuable insights into understand the binding specificity of CBMs. It was suggested^{1,14} that CBM can be further divided into seven structural families (β -sandwich, β -trefoil, cystein knot, unique, OB fold, hevein fold and unique containing hevein-like fold). The recognition of carbohydrate by CBM plays the key role in

some important biological processes, e.g., cell signaling, and host-pathogen interactions. The primary function of the CBM has been postulated that it can facilitate the accessibility of glycosidic bond by enzyme active site, thus enhance the catalytic efficiency. Indeed, it has been shown that the removal of CBM from the enzyme could dramatically lead to decreasing of the enzymatic activity.¹⁵ The understanding of the CBM–polysaccharide recognition and corresponding binding process are fundamental in studying polysaccharide degradation.

In the plant cell wall, the cellulose is the most abundant component compared with mannan, chitin or xylan. The binding specificity of glycosidases for the cellulose deserves special research interests. As we know, some subfamilies of current resolved CBMs are specific to bind cellulose, e.g., family 1 from Cel7A and Cel6A, family 17 and 28 from Cel5A. In this work, we will try to understand the function of one of cellulose-specific CBMs, the family 17 CBM from *Clostridium cellulovorans*.¹⁶ Those cellulose-specific CBMs prefer binding either crystalline cellulose or noncrystalline and soluble cellulose. CBM17 from *C. cellulovorans* 5A has been shown that it binds only noncrystalline cellulose at an extended binding site. The crystallographic structure of CBM17 complexed with cellobetaose molecule was reported in 2001,⁷

Received: February 1, 2012

Revised: April 11, 2012

Published: May 14, 2012

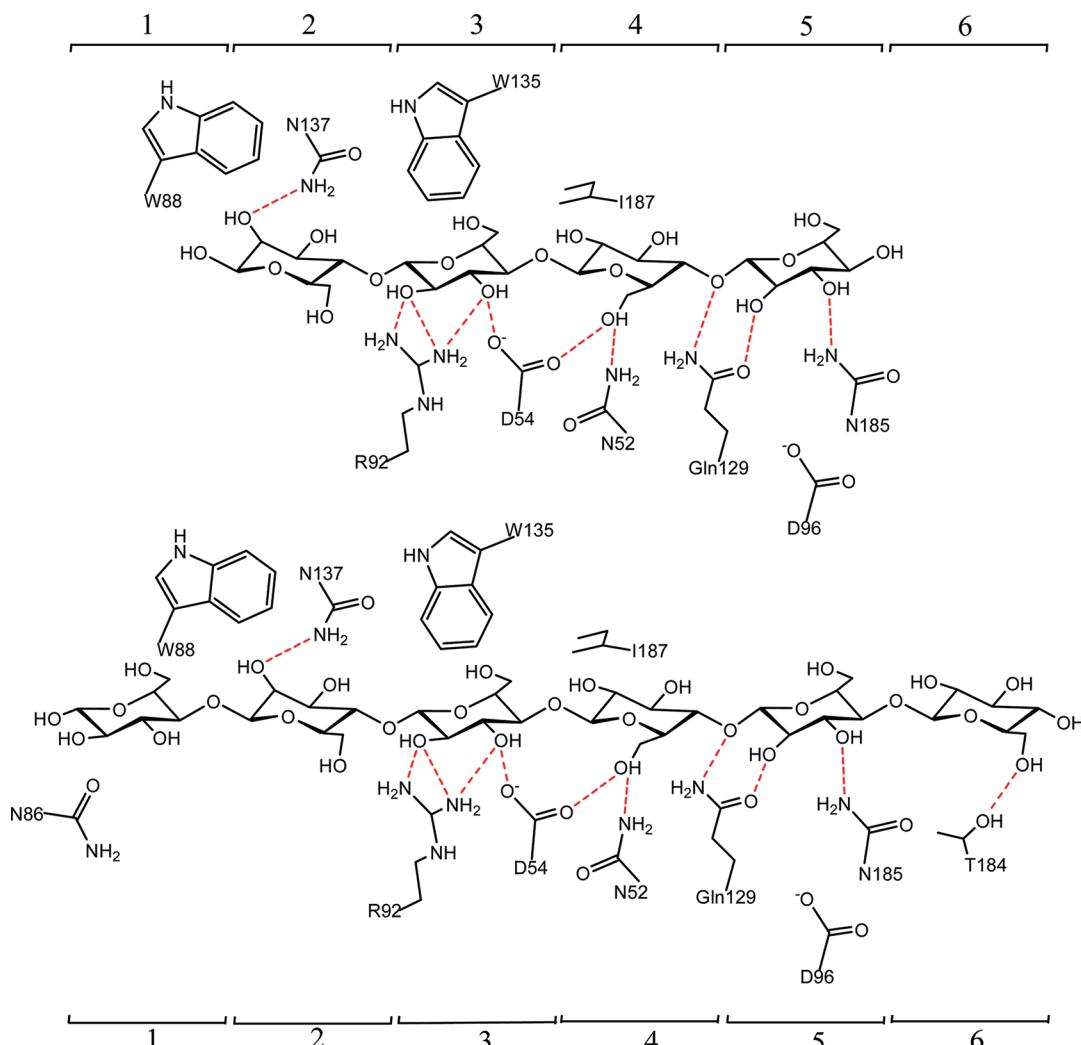


Figure 1. Interactions between CBM binding site and substrate molecules (cellotetraose and cellohexaose). Possible hydrogen bonds are labeled in red color.

which is the first structure that the bounded species is longer than disaccharide. This provides a very good starting point for further investigation of the binding specificity. On the basis of this structure, Notenboom et al⁷ further carried out isothermal titration calorimetry and alanine scanning mutagenesis studies. Some important residues around the binding cleft are identified. Hydrogen bond network between cellobeta-tetraose and enzyme residues are shown to play significant role in the substrate binding. The typical binding free energy contribution of each polar residues of N52, D54, R92, Q129, N137, and N185 is about 1–2 kcal/mol.⁷

As suggested by Boraston et al,¹⁶ the CBM17 has a binding site which spans as long as 31 Å. However, the crystallized cellobeta-tetraose molecule is not long enough to cover the whole binding cleft. Apparently, some important binding characteristics could not be directly identified. In the comprehensive review article by Boraston et al,¹ three types of binding-site “platforms” formed by the aromatic amino acid residues are suggested to play the key role in the carbohydrate recognition specificity. CBM17 belongs to the type B or “twisted” platform. Indeed, two tryptophan residues (W88 and W135) were proposed by Notenboom et al⁷ that they might play the important role in the recognition of longer polysaccharide, such as cellobeta-hexaose, because they are the only two residues which

contain aromatic groups near the binding site. It was further observed that the cellobeta-hexaose molecule could not be bound in the binding site for the mutations of W88A and W135A. Considering hydrogen bond network provided by protein, it is then interesting to perform theoretical simulations to address the detailed influences of two tryptophan residues or other factors on the oligosaccharide binding, since experiment alone could not provide sufficient information. Indeed, it was also shown by some crystallographic studies that the hydrogen bond network is not that essential in the binding for some CBMs, whereas hydrophobic stacking interactions between aromatic residues (Trp or Phe) and sugar rings are dominant.¹⁷ More importantly, detailed study of the CBM binding carbohydrate can be helpful to understand the processive mechanism of glycosidases, or to perform protein engineering study to improve the catalytic activity. Indeed, combined with mutagenesis of the catalytic domain, mutation studies on CBM for Cel5A from *Thermotoga maritima* did show a 14–18 fold increase of the catalytic efficiency.¹⁸

In this article, we built two models for the binding of cellobeta-tetraose and cellobeta-hexaose molecules based on the reported X-ray structure. Extensive molecular dynamics (MD) simulations were carried out to address the binding specificity and functions of residues around the binding site.

2. COMPUTATIONAL DETAILS

2.1. MD Simulations. The initial structure was adopted from the crystal structure of CBM17 from *C. cellulovorans*, which is a complex structure with a cellotetraose molecule (PDB code 1J84).⁷ Interestingly, the crystallized cellotetrasaccharide molecule occupies the binding site from 2 to 5 as shown in Figure 1. According to Boraston's proposal that the binding cleft of CBM17 could accommodate six glucose units, the coordinates of cellobiohexaose molecule were then obtained by simply adding two glucose units at the subsite 1 and 6 based on the X-ray structure. Possible interactions with surrounding residues and the definition of the glucose unit for cellobiohexaose and cellotetraose are then given in Figure 1. For glucose units of cellotetraose molecule, they are named as G2 to G5 according to the binding position.

The resulted systems were then solvated in a pre-equilibrated rectangular box of TIP3P¹⁹ water with sodium ion added to neutralize the system. A typical system consists of over 23000 atoms with the dimension of approximate 60 Å × 73 Å × 66 Å. The periodic boundary conditions and 10 Å cutoff for nonbond interactions were applied. The particle mesh ewald (PME) algorithm²⁰ was used to describe the long-range electrostatic interactions. The positions of water molecules were relaxed by 500 steps of steepest descent (SD) and 1000 steps of conjugate gradient (CG) minimization approach with all of protein and ligand molecules fixed at their positions. Further 10000 steps of CG full minimization were carried out for total system. The obtained systems were gradually heated to 300 K in 20 ps in the NVT ensemble, followed by 10 ns equilibration simulation under 1 atm pressure. Subsequently, further 20 ns MD simulations in the NPT ensemble were performed for data analysis. Newton's equations of atomic motion were integrated by the Verlet algorithm with 2 fs time step. SHAKE algorithm²¹ was applied to constraint bond stretching of the valence bonds involving hydrogen atoms. All calculations were performed using the AMBER9²² suite of programs together with AMBER FF03 force field.²³ GLYCAM 06 carbohydrate parameter set was employed to describe sugars.²⁴

2.2. Binding Free Energy Calculations. To quantitatively assess the binding affinity of oligosaccharides by CBM17, it is necessary to calculate the binding free energies for the bounded oligosaccharides. Several methods have been proposed for calculating absolute binding free energy, e.g., linear response approximation (LRA),²⁵ linear interaction energy (LIE),^{26–28} molecular mechanics Poisson–Boltzmann (or generalized Born) surface area (MM-PB/GBSA),^{29,30} solvated interaction energy (SIE),³¹ or free energy pathway method.³² In this work, the binding free energies are calculated using MM-GBSA²⁹ method.

For the calculation of binding free energy in MM-GBSA framework, it has been discussed extensively.^{33–37} Only a short description is summarized here:

$$\Delta G_{\text{binding}} = \Delta G_{\text{complex}} - \Delta G_{\text{protein}} - \Delta G_{\text{ligand}} \quad (1)$$

$$G = E_{\text{gas}} + G_{\text{sol}} - TS \quad (2)$$

$$E_{\text{gas}} = E_{\text{int}} + E_{\text{vdW}} + E_{\text{ele}} \quad (3)$$

$$G_{\text{sol}} = G_{\text{GB}} + G_{\text{np}} \quad (4)$$

$\Delta G_{\text{complex}}$, $\Delta G_{\text{protein}}$, and ΔG_{ligand} are free energies of the complex, the protein and the ligand, respectively. Each term can be obtained according to eq 2. Practically, they are

calculated as the statistical averages over frames extracted from MD trajectories. The solvation free energy (G_{sol}) can be divided into polar (G_{GB}) and nonpolar (G_{np}) contributions. The polar solvation contribution is calculated by solving the Generalized Born (GB) equation.³⁸ Dielectric constants (ϵ) for the solute and solvent were selected to be 1 and 80, respectively. The nonpolar contribution due to the cavity formation and van der Waals interactions between the solute and solvent can be estimated by the equation of $\gamma SA + b$, where $\gamma = 0.0072 \text{ kcal}/\text{\AA}^2$, $b = 0.0 \text{ kcal/mol}$. The SA is defined as the solvent accessible surface area, which was estimated using the program MSMS.³⁹ For each complex system, binding energies were averaged over 1000 frames of the 20 ns MD trajectory. Not surprisingly, the inclusion of entropic effect in the calculation of total binding free energy can dramatically decrease the gap between theoretical and experimental values.⁴⁰ Entropy contributions are from changes in the degrees of freedom including translation, rotation, and vibration. The translational, rotational, and vibrational entropy terms are functions of the mass and moments of the inertia of molecule, thus can be calculated using the standard equations of statistical mechanics.⁴¹ Specifically, contributions to the vibrational entropy were calculated using the normal-mode analysis approach.⁴² The normal-mode analysis is high computationally demanding, so that $-T\Delta S$ was averaged over only 20 snapshots of the MD trajectory.

2.3. Computational Alanine-Scanning. Computational alanine scanning approach has been widely used in structure-based drug design and protein engineering as an inexpensive way to quantitatively probe key residues in the binding site and understand the substrate binding specificity.^{43–47} The alanine mutant structures were generated by altering the coordinates of the wild-type trajectory, in which we simply replaced the target residue with an alanine residue.^{43,46} Corresponding binding free energies were averaged over 1000 frames at 20 ps time interval were selected from the last 20 ns MD trajectory. The change of binding free energy between the mutant and wild-type complexes is defined as:

$$\Delta\Delta G = \Delta G_{\text{mut}} - \Delta G_{\text{wt}} \quad (5)$$

where ΔG_{wt} , ΔG_{mut} refer to the binding free energies of wild-type and mutant complexes, respectively. Some of key residues around the binding cleft of CBM17 were chosen for the alanine scanning analysis, such as N52, D54, N86, R92, D96, Q129, N137, T184, N185, and I187. Meanwhile, we further carried out explicit MD simulations for three mutants, D54A, R92A, and Q129A, in which we used the same parameters as we did in the wild-type MD simulation.

3. RESULTS AND DISCUSSION

3.1. Cello-Oligosaccharides-CBM17 Complex Systems.

To understand the detailed dynamic effect of oligosaccharide molecules recognized by CBM17, we have in this work carried out extensive molecular dynamics simulations to investigate the recognition specificity of cello-oligosaccharides by CBM17.

First of all, it would be interesting to examine the stability of the apo CBM17, for which the initial structure is extracted from the protein data bank (PDB code 1J83). The setup protocol is essentially the same as we did for complex systems. Throughout the 20 ns MD simulation, the overall protein structure was kept very well, evidenced by the root-mean-square deviation (rmsd) of $1.18 \pm 0.13 \text{ \AA}$ for backbone atoms as shown in Figure 2. Surprisingly, along the trajectory, the indole platforms of two

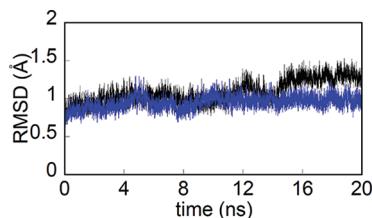


Figure 2. RMSDs for both apo CBM17 (black) and CBM17-cellotetraose complex (blue) as a function of time.

tryptophan residues (W88 and W135) fluctuate drastically as shown in Figure 3. Indeed, these two tryptophan residues have

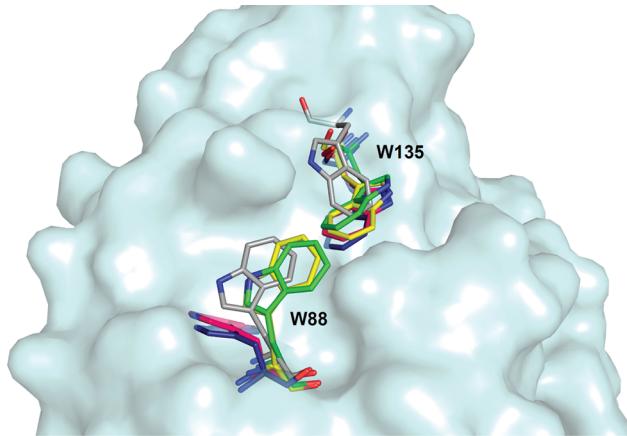


Figure 3. Conformational changes of the twisted aromatic platform in the binding cleft of apo CBM17. Protein backbone is shown in surface. Conformations of Trp88 and Trp135 at the initial time, fifth ns, 10th ns, 15th ns, and 20th ns are shown in gray, green, yellow, magenta, and blue, respectively.

been proposed to play the key role in the cellulose binding via hydrophobic interaction with sugar rings.⁷ Such considerable flexibility could indicate that their side chain groups can recognize the substrate molecule by easy adjustment of their position and orientation.

Furthermore, to validate the simulation protocol we used in this work, 20 ns MD simulation on the cellotetraose molecule recognized by CBM17 were carried out. Although there might be three binding possibilities for cellotetraose molecule, we in

this work only investigated the system based on the X-ray structure in which the cellotetraose occupies the binding sites from 2 to 5. The structure is maintained quite stably, judged by the rmsd of 0.98 ± 0.06 Å for the backbone atoms as plotted in Figure 2. One snapshot is depicted in Figure 4A which was randomly extracted from the MD trajectory to display the binding pattern. It can be seen that the cellotetraose molecule contacts protein through significant hydrogen bond network. All of possible hydrogen bonds for the binding of cellotetraose, which were measured by the hydrogen bond occupancy, are listed in Table 1. Larger occupancy means stronger hydrogen

Table 1. Hydrogen Bonds Network Analysis for Interactions between CBM17 and Cellotetraose and Cellohexaose^a

H-bond	occupancy (%)		
	cellohexaose	cellotetraose	expt ^b
Asn137 N _{δ2} ...G2 O ₂	19.6	27.1	2.57
Arg92 N _{η1} ...G3 O ₂	58.6	91.9	2.79
Arg92 N _{η2} ...G3 O ₂	42.5	52.7	3.27
Asp54 O _{δ1(δ2)} ...G3 O ₃	15.6	51.2	2.79
Asn52 N _{δ2} ...G4 O ₆	45.7	72.8	3.39
Asp54 O _{δ1(δ2)} ...G4 O ₆	36.9	94.2	2.99
Gln129 N _{ε2} ...G4 O ₄	49.2	71.1	2.96
Gln129 N _{ε1} ...G5 O ₂	69.6	66.0	2.49
Gln129 N _{ε2} ...G5 O ₂	46.5	14.8	3.62
Asn185 N _{δ2} ...G5 O ₃	61.5	64.5	2.99
Thr184O _{γ1} ...G6 O ₆	19.0	—	—
Gly134 O...Gln129N _{ε2}	98.39	99.09	2.81

^aExperimental values are extracted from the X-ray structure for comparison. Occupancy is in unit of percentage of the investigated time period (20 ns) during which specific hydrogen bonds are formed. The hydrogen bond is defined as the distance of acceptor and donor atoms shorter than 3.2 Å, and the internal angle of acceptor-H-donor larger than 120°.

bond. Clearly, our simulation could largely reproduce the experimental binding characteristics of cellotetraose by CBM17. For G2 unit, only one weak hydrogen bond is provided by protein, i.e., about 27.1% of the simulation time formed with N137. G3 unit is mainly hydrogen bonded with D54 and R92. Typically, it is about 91.9% for the hydrogen bond between O2 hydroxyl group of G3 unit and N_{η1} atom of R92, and 52.7% between O2 hydroxyl group of G3 unit and N_{η2} atom of R92.

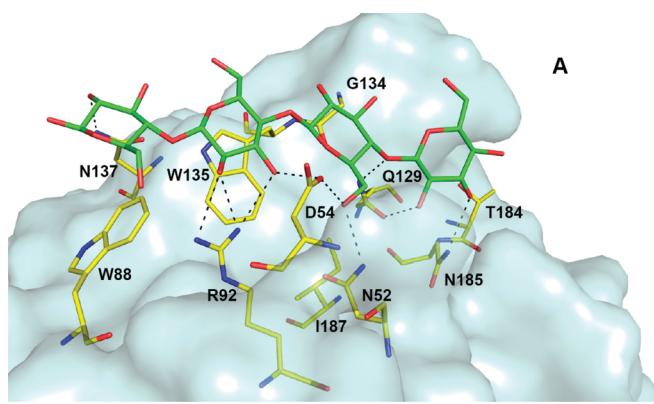


Figure 4. Snapshots of CBM17 bound with cellotetraose (panel A) and cellohexaose (panel B). For better view, hydrogen atoms are not plotted, and hydrogen bonds are represented in black dash line between heavy atoms. The carbon atoms are in green color for sugar, while yellow for protein. Oxygen atoms are in red, and nitrogen atoms are shown in blue.

Table 2. Binding Free Energy for CBM17-Cellobetetaose Complex and Decomposition to Electrostatic Interaction, van der Waals Interaction, Solvation Free Energies, and Entropy^a

	complex	receptor	ligand	delta
E_{ele}	-4163.40(54.89)	-4398.39(53.69)	302.78(6.88)	-67.80(12.62)
E_{vdW}	-755.72(19.98)	-724.98(19.50)	3.06(2.51)	-33.81(5.10)
E_{int}	2716.76(38.37)	2627.61(38.52)	89.15(6.56)	-0.00(0.00)
E_{gas}	-2202.37(62.34)	-2495.76(62.44)	395.00(8.75)	-101.61(10.93)
$G_{sol,np}$	59.66(0.63)	58.44(0.59)	6.50(0.06)	-5.28(0.36)
$G_{sol,GB}$	-1890.13(46.95)	-1884.63(47.08)	-81.84(4.04)	76.33(7.89)
$E_{gas} + G_{sol}$	-4032.84(39.01)	-4321.96(38.60)	319.66(6.80)	-30.55(4.36)
$-TS_{total}$	-1984.81(3.92)	-1922.06(4.22)	-88.70(0.19)	25.95(3.79)
ΔG_{bind}^{cal}				-4.60
ΔG_{bind}^{exp}				-4.19

^aEnergies are in kcal/mol. Numbers in parentheses are standard deviations. Experimental binding free energy is obtained from ref 7.

Table 3. Binding Free Energy for CBM17-Cellobhexaose Complex and Decomposition to Electrostatic Interaction, van der Waals Interaction, Solvation Free Energies and Entropy Terms^a

energy	complex	receptor	ligand	delta
E_{ele}	-4052.81(45.25)	-4431.50(43.64)	441.14(9.43)	-62.44(12.12)
E_{vdW}	-782.34(21.09)	-735.93(19.73)	4.58(3.28)	-50.99(6.28)
E_{int}	2761.99(40.13)	2627.93(38.63)	134.06(8.07)	-0.00(0.00)
E_{gas}	-2073.16(56.54)	-2539.50(53.44)	579.77(11.17)	-113.43(13.58)
$G_{sol,np}$	58.89(0.71)	57.24(0.53)	9.16(0.07)	-7.52(0.55)
$G_{sol,GB}$	-1873.25(38.88)	-1845.89(37.15)	-110.50(5.42)	83.15(10.06)
$E_{gas} + G_{sol}$	-3887.52(40.58)	-4328.15(38.70)	478.43(8.54)	-37.80(5.08)
$-TS_{total}$	-2013.40(3.81)	-1919.65(3.38)	-125.52(0.28)	31.77(3.62)
ΔG_{bind}^{cal}				-6.03
ΔG_{bind}^{exp}				-6.55

^aEnergies are in kcal/mol. Numbers in parentheses are standard deviations. Experimental binding free energy is obtained from ref 7

Such obvious occupancy difference indicates different strength of hydrogen bonds, which also agrees well with the experimental distance difference between O2 atom and these two nitrogen atoms (2.79 vs 3.29 Å). O6 hydroxyl group of G4 unit is hydrogen bonded with N52 (72.8% of the simulation time) and D54 (94.2% of the simulation time). At the same time, O4 hydroxyl group of G4 unit is also stabilized by Q129 with relatively strong hydrogen bond of 71.1% occupancy. Over 60% of MD simulation time, the G5 unit is stabilized by hydrogen bonds with Q129 and N185. It can be seen that the hydrogen bond network should be one of the key issues in the cellobetetaose binding. From Figure 4A, the hydrophobic interactions between substrate and protein might be also important, since the G3 sugar ring is almost in parallel position to the indole group of W135. This can be judged by the dihedral angle between the indole group of W135 and G3 sugar ring. In this work, we first selected C₁, C₃, and C₅ atoms from the glucose unit, and C_{δ2}, C_{β2}, and C_{ζ3} atoms from the indole group of the tryptophan residue, which can form two planes. If not otherwise stated, the dihedral angle between the glucose unit and tryptophan will be defined as the angle between these two planes. At the same time, the distance between the glucose unit and tryptophan residue is then defined using the distance between centers of mass of the glucose ring and the side chain of tryptophan. For the binding of cellobetetaose, this dihedral angle is calculated about 12.2 ± 6.5° between the G3 unit and indole group of W135, with the distance of 4.11 ± 0.20 Å between two rings. The presence of the hydrophobic interaction between G3 and W135 could provide additional contribution to the substrate binding.

For the binding of CBM17 for cellobhexaose, during the simulation, the cellobhexaose molecule was found to stay in the protein binding surface very tightly. One of snapshots is given in Figure 4B. Every glucose recognition site has its own special binding pattern since the cellobhexaose can fully cover all six binding sites. Similar to the binding of cellobetetaose, the cellobhexaose molecule contacts CBM17 via both hydrogen bond network and hydrophobic interactions. All of possible hydrogen bonds measured by the hydrogen bond occupancy were also summarized in Table 1. For G1 and G2 units, major interactions with the protein are provided from the hydrophobic interactions with W88. The calculated dihedral angle between G1 and the side chain of W88 is about 157.7° ± 10.4°, with the distance of 4.91 ± 0.52 Å between two centers of mass of two rings. Meanwhile, the dihedral angle between W88 and G2 unit is about 16.3° ± 9.3°, and the corresponding center of mass distance is 4.78 ± 0.40 Å. It can be seen that the indole group of Trp88 is in near parallel position to sugar rings of G1 and G2 units. Strong hydrophobic stacking interactions can thus be established. Similar situation can be found for G3 unit with W135, for which the corresponding dihedral angle is 15.6° ± 6.6°. The corresponding center of mass distance between G3 and W135 is 4.28 ± 0.28 Å. No direct hydrogen bond formed between N86 with the substrate can be found during the simulation. This agrees with experimental observation that N86A has no effect on the affinity of CBM17 for the cellobhexaose.⁷ In addition, for the stabilization of the G2 and G3 units, besides hydrophobic interactions provided by W88 and W135, some hydrogen bonds can be also observed for these two subsites. For example, G2 unit is hydrogen bonded with N137, although relatively low hydrogen occupancy, 19.6%,

Table 4. Computational Alanine Scanning of CBM17 Complexed with Cellohexaose Using the Single Trajectory Method^a

	N52A	D54A	N86A	W88A	R92A	D96A
$\langle \Delta\Delta E_{ele} \rangle$	0.56	18.79	0.34	2.45	13.00	1.26
$\langle \Delta\Delta E_{vdW} \rangle$	0.68	-1.01	0.37	8.30	1.61	0.37
$\langle \Delta\Delta E_{gas} \rangle$	1.24	17.78	0.71	10.74	14.62	1.62
$\langle \Delta\Delta G_{sol,np} \rangle$	-0.01	0.10	0.08	0.07	0.04	0.08
$\langle \Delta\Delta G_{sol,GB} \rangle$	-0.62	-12.34	-0.55	-2.78	-10.27	-1.58
$\langle \Delta\Delta G_{bind,cal} \rangle$	0.61	5.54(1.44)	0.24	8.02	4.39(2.09)	0.12
$\Delta\Delta G_{bind,exp}$	1.55	1.34	-	-	1.89	0.79
	Q129A	W135A	N137A	T184A	N185A	I187A
$\langle \Delta\Delta E_{ele} \rangle$	5.47	2.50	3.32	2.35	3.30	2.35
$\langle \Delta\Delta E_{vdW} \rangle$	2.15	9.28	1.28	0.79	0.81	0.79
$\langle \Delta\Delta E_{gas} \rangle$	7.63	11.78	4.60	3.14	4.10	3.14
$\langle \Delta\Delta G_{sol,np} \rangle$	-0.04	0.19	0.15	0.01	-0.01	0.01
$\langle \Delta\Delta G_{sol,GB} \rangle$	-3.39	-2.94	-3.30	-1.60	-2.11	-1.60
$\langle \Delta\Delta G_{bind,cal} \rangle$	4.20(2.62)	9.03	1.45	1.55	1.98	1.55
$\Delta\Delta G_{bind,exp}$	1.77	-	1.58	-	1.91	-

^aEnergies are in kcal/mol. Experimental data is obtained from ref 7. Numbers in the parentheses were obtained by separate MD simulations.

can be found from Table 1. Similar with the cellobetaose, hydrogen bonds are also formed between the O2 hydroxyl group of G3 unit and the guanidinium group of R92, but lower frequency of 58.6% and 42.5% can be found. D54 is shown to mainly recognize the G3 and G4 units. Additionally, other hydrophilic residues such N52, Q129 and N185 also make strong hydrogen bonds with G3 to G5 units, respectively. This is in good agreement with that Q129A and N185A mutants could cause significant loss of the binding affinity.⁷ Interestingly, only T184 was found to be close to G6 unit, but with relatively low hydrogen bond occupancy (about 19%). Overall, the frequency of the occurrence of hydrogen bonds between the cellohexaose and protein seems to be lower than that for the binding of cellobetaose molecule. This suggests that for the binding of cellohexaose, the hydrogen bond net work might not be as critical as that in the binding of cellobetaose, and major contributions to the binding affinity for the cellohexaose should come from two tryptophan residues.

Similar binding pattern could be found in the family 15 CBM from *Pseudomonas cellulosa* xylanase Xyn10C, for which the X-ray structure has been crystallized with xylopentaose molecule.¹⁷ Interestingly, direct hydrogen bond interactions between protein and xylopentaose are found to be quite few. In contrast, two tryptophan residues, W176 and W181, are found to form strong hydrophobic stacking interactions with two xylopranose units at *n* and *n* + 2.

3.2. Binding Free Energy Calculations. To obtain insights into different energy contributions to the binding affinity of cello-oligosaccharides to CBM17, binding free energies were computed for CBM17-cellobetaose, and CBM17-cellohexaose complexes. As shown in Table 2 and 3, the calculated total binding free energies (-4.69 kcal/mol for the cellobetaose, and -6.03 kcal/mol for the cellohexaose) are in fairly good agreement with experimental values (-4.19 kcal/mol for the cellobetaose, and -6.55 kcal/mol for the cellohexaose). Such small differences should further demonstrate the significance of our results. In addition, Tables 2 and 3 also list different components for the calculation of binding free energy including van der Waals (E_{vdW}) and electrostatic interactions (E_{ele}), and polar ($G_{sol,GB}$) and nonpolar ($G_{sol,np}$) contributions to the solvation energy.

Usually, there are two major factors that could affect the substrate binding, i.e., the electrostatic ($\Delta E_{ele} + \Delta G_{sol,GB}$) and

nonpolar ($\Delta E_{vdW} + \Delta G_{sol,np}$) terms. In current systems we studied, for the cellobetaose, from statistical analysis of the data, we can see that $\Delta E_{ele} + \Delta G_{sol,GB} = 8.53$ kcal/mol, and 20.71 kcal/mol in the binding of cellohexaose. The electrostatic interactions are canceled by polar solvation contributions, such that the total electrostatic terms are shown to disfavor the recognition of cello-oligosaccharides by CBM17. In other words, it is more likely that the polar interactions (like hydrogen bonds) could only provide directional constraints for the substrate binding. In contrast, for both substrates, van der Waals interactions have dominantly favorable contributions to the binding affinity, as does the nonpolar part of solvation. However, it should be noted that this is just the total energetic effect. To obtain more detailed information, it might need some more cautious analysis for contributions by individual residues to reach final conclusion as we discussed in below.

In order to quickly and quantitatively identify some important residues in the substrate binding, the alanine scanning analysis has been applied to various CBM subfamilies. Experimentally, mutations to the alanine of those direct hydrogen bonding polar residues were found to have a significant decreasing in the overall binding affinity, e.g., CBM4-1 from *Cellulomonas fimi*,⁴⁸ CBM6 from *Clostridium thermocellum* Xylanase11A,⁴⁹ CBM15 from *Cellvibrio japonicus* Xyn10C⁵⁰ and CBM22-2 from *Clostridium thermocellum* Xyn10B.⁵¹ For the CBM17 complexed with cellohexaose, a variety of mutagenesis studies have also been reported, especially for those polar residues which might play some key roles in the substrate binding via hydrogen bond interaction.⁷ In this work, the computational alanine scanning was performed for the CBM17-cellohexaose complex to investigate residues around the binding groove using a single trajectory method.⁴³

All of mutagenesis simulations were calculated using MM/GBSA alanine scanning approach, in which only specific residue was manually mutated using the same trajectory we obtained in the simulation for wild type protein. Calculated relative binding free energies are listed in Table 4. The calculated relative changes of binding free energy ($\Delta\Delta G$) agree with experimental data very well. Typical energy differences are less than 1.0 kcal/mol, except three polar/charged residues, such as D54, R92, and Q129. Interestingly, for these three residues, changes in the electrostatic interaction are three largest compared to other

residues. As pointed by Hou et al.,⁴⁷ it is quite challenging to get a reasonable evaluation of electrostatic interaction and polar contribution to the solvation free energy upon mutating from polar/charged to a neutral residue using the single trajectory method, since the mutation could significantly change the local electrostatic interactions with substrate. It would be highly desirable to conduct explicit MD simulations for such kind of residues to get more reasonable binding free energies. In this work, three separate 20 ns MD simulations were performed for D54A, R92A, and Q129A, respectively. Their binding free energies are then calculated using the standard MM/GBSA procedure as we did for wild type protein. Calculated binding free energies are also listed in the parentheses of Table 4 for comparison, in which much better agreement with experimental data can be observed. For example, the relative free energy change for D54A becomes 1.44 kcal/mol from 5.54 kcal/mol, which can be comparable with experimental result of 1.34 kcal/mol. Although no direct experimental data for the mutation of W88A and W135A are reported, we do know the mutations could result in total loss of binding affinity. Thus, it would be interesting to theoretically assess their individual contributions to the binding. It can be seen from the Table 4, both residues clearly provide dominant contributions to the binding free energy versus other residues around the binding surface. This is also consistent with the suggestion that van der Waals interactions are the key for the cellulose binding as shown by our above MD simulations, since the van der Waals interactions contribute dominantly for the interactions between tryptophan residues and protein. Finally, one of important assumptions for computational alanine scanning analysis is that the introduction of the local changes to the residue should not affect the overall binding modes for the substrate to protein. To confirm this, the overlap representations of wild type and selected snapshots of D54A, R92A, and Q129A are depicted in Figure 5. As we can see, the total binding modes do not have significant changes upon mutations. This can give us confidences in the above analysis.

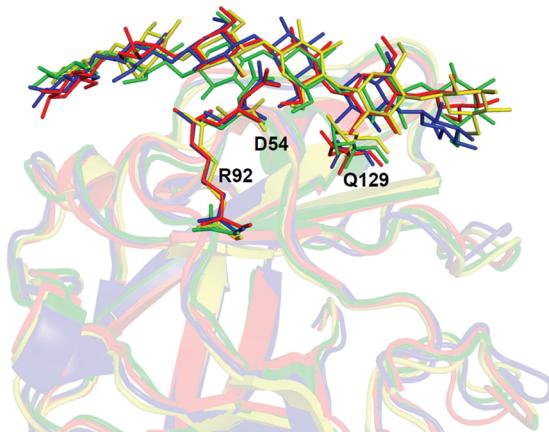


Figure 5. Snapshots overlap between wild type (red), D54A (yellow), R92A (green), and Q129A (blue).

3.3. Free Energy Decomposition per Residues. Compared to virtual alanine scanning, another inexpensive technique to evaluate the influence of each residue to the binding is the free energy decomposition analysis per residues. Basic idea is to sum per-atom contribution over the residue atoms, so that the contribution from individual residues can be

estimated at the atomistic level in a “non-mutable” way. The interaction energy between each residue and substrate was computed using the MM/GBSA decomposition protocol implemented in the *mm_pbsa* module in the AMBER9 software package. It should be noted that no entropy contribution is included in this approach. The free energy decomposition strategy at the per-residue basis allows us to analyze in depth the interaction profile between substrate and protein, and therefore determine the most important residues and driving force to the binding.

Table 5 and 6 report the side chain (ΔG) and backbone ($B\Delta G$) contributions as well as contributions of the nonpolar terms (van der Waals interactions and nonpolar solvation free energies) and polar terms (electrostatic interactions and polar solvation free energies) to the recognition of cellobetaose and cellobetaose by CBM17, respectively. Only residues with the $|\Delta G_{\text{subtotal}}| \geq 0.5$ kcal/mol are listed. Notably, those residues with significant contributions to the overall binding are mostly situated at the substrate binding groove. More interestingly, they are all shown to have favorable contributions to the substrate binding, which suggests that the recognition ability of CBM17 is quite efficient and optimal. Of course, contributions from other residues, which have no direct contacts with substrate molecule, cannot be ignored either. For example, the G134 provides about -1.1 kcal/mol stabilization energy to both substrates. The function of this residue is found to orientate the Q129 via hydrogen bond. As the Table 1 shows, the hydrogen bonds formed between the G134 and Q129 are well maintained throughout the simulation for both substrates.

In general, theoretical results compare quite well with experimental alanine scanning of a few residues for the binding of cellobetaose. In particular, as Table 6 shows, $\Delta G_{\text{subtotal}} = -1.70$ kcal/mol for D54, which is in fairly good agreement with experimental observation that D54A causes 1.34 kcal/mol loss of the binding affinity.⁷ Furthermore, similar results can be found for other residues around the binding groove. For example, $\Delta G_{\text{subtotal}}(\text{R92}) = -1.37$ kcal/mol can compare with very well with its contribution of 1.92 kcal/mol to the total binding free energy obtained by mutation to alanine. On the other hand, the subtotal contributions to the binding provided by Q129, N137, and N185 are -1.74, -1.14, and -1.96 kcal/mol, respectively. Remarkably, corresponding drops of binding free energies for these three residues caused by mutagenesis are shown to be 1.77, 1.58, and 1.91 kcal/mol, respectively. Excellent agreement can be seen here. Our computation can support that interactions provided by these hydrophilic residues have important influences on the substrate binding.

As we have discussed in above section that the nonpolar contacts play the mainly favorable role in the calculation of binding free energy. Same observations can be obtained by the free energy decomposition strategy in a per-residue way. Particularly, some hydrophobic residues are shown to have significant subtotal binding free energies. For the binding of cellobetaose, W135 makes the most favorable contribution. Similarly, W88 and W135 are shown to be the most favorable residues for the binding of cellobetaose. It can indicate that these two tryptophan residues are two key residues for the binding of cellulose with no doubt. This is also consistent with results obtained by experimental studies and theoretical studies as we mentioned above. In addition, as we can see from the Table 5 and 6, the nonpolar terms, especially van der Waals interactions, clearly are the major components in the calculation of subtotal binding free energies for most of

Table 5. Free Energy Decomposition for CBM17–Cellotetraose Complex on the Per-Residue Basis

residues	ΔE_{vdW}	ΔE_{ele}	$\Delta G_{sol,GB}$	$\Delta G_{sol,np}$	$S\Delta G_{subtotal}$	$B\Delta G_{subtotal}$	$\Delta G_{subtotal}$
N52	-0.44	-0.99	0.74	-0.07	-0.78	0.01	-0.76
D54	-0.97	-8.70	9.27	-0.29	-0.42	-0.26	-0.69
W88	-1.55	-0.19	0.31	-0.19	-1.61	-0.01	-1.62
R92	-0.42	-10.07	8.20	-0.18	-2.47	0.01	-2.46
Q129	-0.95	-3.21	2.65	-0.21	-1.72	0.00	-1.72
G134	-1.19	-0.43	0.59	-0.16	-0.05	-1.13	-1.18
W135	-5.49	-1.35	1.20	-0.51	-5.72	-0.43	-6.15
N137	-0.91	-1.42	1.58	-0.22	-0.88	-0.09	-0.97
N185	-0.64	-1.38	0.94	-0.17	-1.21	-0.03	-1.24
I187	-0.90	-0.15	0.17	-0.05	-0.86	-0.06	-0.93

Table 6. Free Energy Decomposition for CBM17–Cellohexaose Complex on the Per-Residue Basis

residues	ΔE_{vdW}	ΔE_{ele}	$\Delta G_{sol,GB}$	$\Delta G_{sol,np}$	$S\Delta G_{subtotal}$	$B\Delta G_{subtotal}$	$\Delta G_{subtotal}$
D54	-0.30	-9.24	8.12	-0.27	-1.50	-0.20	-1.70
W88	-4.97	-1.14	1.14	-0.52	-5.32	-0.18	-5.50
R92	-0.86	-6.59	6.27	-0.20	-1.38	0.01	-1.37
Q129	-1.54	-2.70	2.73	-0.23	-1.63	-0.11	-1.74
G134	-1.11	-0.37	0.52	-0.15	-0.11	-1.01	-1.12
W135	-5.70	-1.19	1.26	-0.56	-5.85	-0.35	-6.20
N137	-1.04	-1.83	2.00	-0.27	-1.07	-0.08	-1.14
T184	-1.23	-1.51	1.66	-0.25	-1.02	-0.30	-1.33
N185	-1.27	-1.87	1.38	-0.21	-1.75	-0.21	-1.96
I187	-0.85	-0.16	0.18	-0.05	-0.80	-0.08	-0.88

residues. However, some interesting trends can be also found for the binding of cellotetraose and cellohexaose. For example, in the binding of the cellotetraose, the electrostatic term of R92 is about -1.8 kcal/mol ($\Delta E_{ele} + \Delta G_{sol,GB}$), which is stronger than nonpolar contribution of -0.6 kcal/mol ($\Delta E_{vdW} + \Delta G_{sol,np}$). This can suggest that the polar interactions still have some significant contribution to the binding for some specific residues, although overall electrostatic interactions are shown to be unfavorable. On the other hand, the polar contribution to the subtotal binding free energy in the case of cellohexaose is decreased to -0.32 kcal/mol, which is not as strong as -1.06 kcal/mol for the nonpolar interactions. Similar situations can be found for some other hydrophilic residues such as Q129 and N185. This can be attributed to weaker hydrogen bond network formed between the cellohexaose and protein as we have pointed in above section. Nevertheless, stronger binding affinity for cellohexaose than cellotetraose was determined by experimental study. Clearly, additional strong hydrophobic stacking interaction formed between W88 and cellohexaose should provide additional favorable contribution to the binding. Indeed, relatively large subtotal binding free energy of -5.5 kcal/mol is assigned to W88, which can compensate the loss of binding free energy caused by lowered favorable polar contributions. Interestingly, it is not unique that the tryptophan residues of the family 17 CBM that have such important contribution to the substrate recognition. In fact, for those type B CBMs which have a β -jelly roll topological binding groove, aromatic residues are the common determined characters in the binding of carbohydrates. For example, W78, W129 and F128 can be found for CBM28 from *Clostridium josui*.⁵² In the binding of cellohexaose and manohexaose by CBM29–2 from *Piromyces equi*,⁵³ the only interaction with sugar unit at the binding site 6 is the stacking interaction with W24. On the other hand, more important and prevalent CBMs (especially type A) only recognize crystalline

cellulose. Recently, Yui et al.⁵⁴ constructed a docking model of CBM from Cel7A with cellulose I α crystal surfaces. Interestingly, it was also suggested by these authors that hydrogen bonding scheme is not essential for substrate specificity. Apparently, the influence of van der Waals interactions on the binding of glycosidases for the polysaccharides deserves special attention.

4. CONCLUSION

We have, in this work, investigated the recognition specificity of cellotetraose and cellohexaose to the family 17 CBM from *C. cellulovorans*. Calculated total binding free energies are in excellent agreement with experimental results. Further analyses of components of binding free energies for both substrates indicate that van der Waals interaction and nonpolar solvation free energy make favorable contributions to the binding, while overall polar interactions disfavor the binding. Computational alanine scanning and per-residue free energy decomposition analysis suggest that only those residues located at the binding groove can have significant stabilization effect to the substrate binding. Two tryptophan residues, W88 and W135, are shown to be in near parallel position to pyranose rings, and provide the largest contributions to total binding affinity. On the other hand, to fully understand the hydrolysis mechanism of glycosidases, such as cellulases, it is not just enough to study the function of CBM. A complete model including all three modules, CD, CBM, and linker region, has to be constructed in the simulation. It is our hope that the current study could stimulate further theoretical and experimental exploration to understand the catalytic mechanism of cellulases, and thus finally improve the catalytic activity.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dgxu@scu.edu.cn. Telephone: 86-28-85412290. Fax: 86-28-85412906.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by Natural Science Foundation of China (No. 20803048 and 21073125, 31170675), and by the Program for New Century Excellent Talents in University (no. NCET-10-0606). Parts of the results described in this paper are obtained on the Deepcomp7000 of Supercomputing Center, Computer Network Information Center of Chinese Academy of Sciences.

REFERENCES

- (1) Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *Biochem. J.* **2004**, *382*, 769–781.
- (2) van Bueren, A. L.; Ghinet, M. G.; Gregg, K.; Fleury, A.; Brzezinski, R.; Boraston, A. B. *J. Mol. Biol.* **2009**, *385*, 131–139.
- (3) van Bueren, A. L.; Morland, C.; Gilbert, H. J.; Boraston, A. B. *J. Biol. Chem.* **2005**, *280*, 530–537.
- (4) Vandermarliere, E.; Bourgois, T. M.; Winn, M. D.; van Campenhout, S.; Volckaert, G.; A., D. J.; Strelkov, S. V.; Courtin, C. M. *Biochem. J.* **2009**, *418*, 39–47.
- (5) Henshaw, J.; Horne-Bitschy, A.; van Bueren, A. L.; Money, V. A.; Bolam, D. N.; Czjzek, M.; Ekborg, N. A.; Weiner, R. M.; Hutcheson, S. W.; Davies, G. J.; Boraston, A. B.; Gilbert, H. J. *J. Biol. Chem.* **2006**, *281*, 17099–17107.
- (6) Notenboom, V.; Boraston, A. B.; Kilburn, D. G.; Rose, D. R. *Biochemistry* **2001**, *40*, 6248–6256.
- (7) Notenboom, V.; Boraston, A. B.; Chiu, P.; Freeland, A. C. J.; Kilburn, D. G.; Rose, D. R. *J. Mol. Biol.* **2001**, *314*, 797–806.
- (8) Najmudin, S.; Pinheiro, B. A.; Prates, J. A. M.; Gilbert, H. J.; Romao, M. J.; Fontes, C. M. G. A. *J. Struct. Biol.* **2010**, *172*, 353–362.
- (9) Charnock, S. J.; Bolam, D. N.; Nurizzo, D.; Szabo, I.; McKie, V. A.; Gilbert, H. J.; Davies, G. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14077–14082.
- (10) Abbott, D. W.; Hrynuik, S.; Boraston, A. B. *J. Mol. Biol.* **2007**, *367*, 1023–1033.
- (11) Montanier, C.; van Bueren, A. L.; Dumon, C.; Flint, J. E.; Correia, M. A.; Prates, J. A.; Firbank, S. J.; Lewis, R. J.; Grondin, G. G.; Ghinet, M. G.; Gloster, T. M.; Herve, C.; Knox, J. P.; Talbot, B. G.; Turkenburg, J. P.; Kerouvo, J.; Brzezinski, R.; Fontes, C. M. G. A.; Davies, G. J.; Boraston, A. B.; Gilbert, H. J. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3065–3070.
- (12) Cid, M.; Pedersen, H. L.; Kaneko, S.; Coutinho, P. M.; Henrissat, B.; Willats, W. G.; Boraston, A. B. *J. Biol. Chem.* **2010**, *285*, 35999–36009.
- (13) Montanier, C. Y.; Correia, M. A.; Flint, J. E.; Zhu, Y.; Basle, A.; McKee, L. S.; Prates, J. A. M.; Polizzi, S. J.; Coutinho, P. M.; Lewis, R. J.; Henrissat, B.; Fontes, C. M. G. A.; Gilbert, H. J. *J. Biol. Chem.* **2011**, *286*, 22499–22509.
- (14) Hashimoto, H. *Cell. Mol. Life Sci.* **2006**, *63*, 2954–2967.
- (15) Coutinho, J. B.; Gilkes, N. R.; Kilburn, D. G.; Warren, R. A.; Miller, R. C., Jr. *FEMS Microbiol. Lett.* **1993**, *113*, 211–217.
- (16) Boraston, A. B.; Chiu, P.; Warren, R. A. J.; Kilburn, D. G. *Biochem.* **2000**, *39*, 11129–11136.
- (17) Szabo, L.; Jamal, S.; Xie, H. F.; Charnock, S. J.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *J. Biol. Chem.* **2001**, *276*, 49061–49065.
- (18) Mahadevan, S. A.; Wi, S. G.; Lee, D.-S.; Bae, H.-J. *FEMS Microbiol. Lett.* **2008**, *287*, 205–211.
- (19) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (20) Darden, T. A.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (21) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (22) Case, D. A.; Darden, T. A.; T. E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Hayik, B. W. S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER9*; University of California: San Francisco, CA, 2006.
- (23) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (24) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-outeiriño, J.; Daniels, C. R.; Foley, B. L. W.; R., J. *J. Comput. Chem.* **2008**, *29*, 622–655.
- (25) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. *Protein Eng.* **1992**, *5*, 215–228.
- (26) Åqvist, J.; Medina, C.; Samuelsson, J. E. *Protein Eng.* **1994**, *7*, 385–391.
- (27) Bren, U.; Martínek, V.; Florián, J. *J. Phys. Chem. B* **2006**, *110*, 10557–10566.
- (28) Perdih, A.; Bren, U.; Solmajer, T. *J. Mol. Model.* **2009**, *15*, 983–996.
- (29) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.
- (30) Lee, M. S.; Salsbury, F. R.; Olson, M. A. *J. Comput. Chem.* **2004**, *25*, 1967–1978.
- (31) Naím, M.; Bhat, S.; Rankin, K. N.; Dennis, S.; Chowdhury, S. F.; Siddiqi, I.; Drabik, P.; Sulea, T.; Bayly, C. I.; Jakalian, A.; Purisima, E. O. *J. Chem. Inf. Model.* **2007**, *47*, 122–133.
- (32) Gilson, M. K.; Zhou, H. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
- (33) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, I.; Lee, M.; Lee, T.; Duan, Y.; Wang, W. D., O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III *Acc. Chem. Res.* **2000**, *33*, 889–897.
- (34) Hou, T.; Wang, J.; Li, Y.; Wang, W. *J. Comput. Chem.* **2011**, *32*, 866–877.
- (35) Rizzo, R. C.; Toba, S.; Kuntz, I. D. *J. Med. Chem.* **2004**, *47*, 3065–3074.
- (36) Strockbine, B.; Rizzo, R. C. *Proteins: Struct. Funct. Bioinform.* **2007**, *67*, 630–642.
- (37) Yang, C.; Sun, H.; Chen, J.; Coleska, Z. N.; Wang, S. *J. Am. Chem. Soc.* **2009**, *131*, 13709–13721.
- (38) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
- (39) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305–320.
- (40) Zoete, V.; Meuwly, M.; Karplus, M. *Proteins* **2005**, *61*, 79–93.
- (41) McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, CA, 2000.
- (42) Case, D. A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285–290.
- (43) Massova, I. K., P. A. *J. Am. Chem. Soc.* **1999**, *121*, 8133–8143.
- (44) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. B* **2006**, *110*, 10962–10969.
- (45) Damborsky, J.; Brezovsky, J. *Curr. Opin. Chem. Biol.* **2009**, *13*, 26–34.
- (46) Li, T.; Froeyen, M.; Herdewijn, P. *J. Mol. Graph. Model.* **2008**, *26*, 813–823.
- (47) Hou, S.; Massova, I.; Kollman, P. A. *J. Comput. Chem.* **2002**, *23*, 15–27.
- (48) Kormos, J. J., P. E.; Brun, E.; Tomme, P.; McIntosh, L. P.; Haynes, C. A.; Kilburn, D. G. *Biochemistry* **2000**, *39*, 8844–8852.
- (49) Czjzek, M.; Bolam, D. N.; Mosbah, A.; Allouch, J.; Fontes, C. M.; Ferreira, L. M.; Bornet, O.; Zamboni, V.; Darbon, H.; Smith, N. L.; Black, G. W.; Henrissat, B.; Gilbert, H. J. *J. Biol. Chem.* **2001**, *276*, 48580–48587.
- (50) Pell, G.; Williamson, M. P.; Walters, C.; Du, H.; Gilbert, H. J.; Bolam, D. N. *Biochemistry* **2003**, *42*, 9316–9323.

- (51) Charnock, S. J.; Bolam, D. N.; Turkenburg, J. P.; Gilbert, H. J.; Ferreira, L. M. A.; Davies, G. J.; Fontes, C. M. *Biochemistry* **2000**, *39*, 5013–5021.
- (52) Tsukimoto, K.; Takada, R.; Araki, Y.; Suzuki, K.; Karita, S.; Wakagi, T.; Shoun, H.; Watanabe, T.; Fushinobu, S. *FEBS Lett.* **2010**, *584*, 1205–1211.
- (53) Charnock, S. J.; Bolam, D. N.; Nurizzo, D.; Szabo, L.; McKie, V. A.; Gilbert, H. J.; Davies, G. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14077–14082.
- (54) Yui, T.; Shiiba, H.; Tsutsumi, Y.; Hayashi, S.; Miyata, T.; Hirata, F. *J. Phys. Chem. B* **2010**, *114*, 49–58.