JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

Protein Allostery at the Solid−Liquid Interface: Endoglucanase Attachment to Cellulose Affects Glucan Clenching in the Binding Cleft

Yuchun Lin,[†] Jordi Silvestre-Ryan,[‡] Michael E. Himmel,[§] Michael F. Crowley,[§] Gregg T. Beckham,[∥,⊥] and Jhih-Wei Chu*[,†]

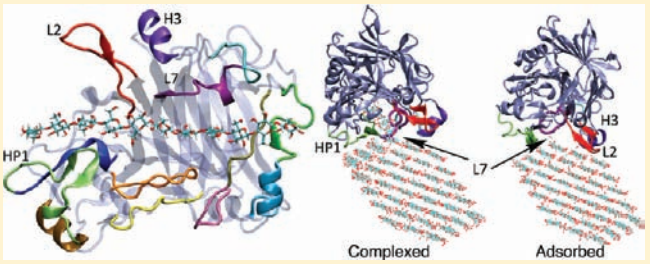
[†]Department of Chemical and Biomolecular Engineering and [‡]Department of Bioengineering, University of California, Berkeley, Berkeley, California, United States

[§]Biosciences Center and [∥]National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colorado, United States

[⊥]Department of Chemical Engineering, Colorado School of Mines, Golden, Colorado, United States

Supporting Information

ABSTRACT: At phase boundaries, physical activities of enzymes such as substrate complexation play critical roles in driving biocatalysis. A prominent example is the cellulase cocktails secreted by fungi and bacteria for deconstructing crystalline cellulose in biomass into soluble sugars. At interfaces, molecular mechanisms of the physical steps in biocatalysis remain elusive due to the difficulties of characterizing protein action with high temporal and spatial resolution. Here, we focus on endoglucanase I (Cel7B) from the fungus *Trichoderma reesei* that hydrolyzes glycosidic bonds on cellulose randomly. We employ all-atom molecular dynamics (MD) simulations to elucidate the interactions of the catalytic domain (CD) of Cel7B with a cellulose microfibril before and after complexing a glucan chain in the binding cleft. The calculated mechanical coupling networks in Cel7B−glucan and Cel7B−microfibril complexes reveal a previously unresolved allosteric coupling at the solid−liquid interface: attachment of the Cel7B CD to the cellulose surface affects glucan chain clenching in the binding cleft. Alternative loop segments of the Cel7B CD were found to affix to intact or defective surface structures on the microfibril, depending on the complexation state. From a multiple sequence alignment, residues in surface-affixing segments show strong conservation, highlighting the functional importance of the physical activities that they facilitate. Surface-affixing residues also demonstrate significant sequence correlation with active-site residues, revealing the functional connection between complexation and hydrolysis. Analysis of the Cel7B CD exemplifies that the mechanical coupling networks calculated from atomistic MD simulations can be used to capture the conservation and correlation in sequence alignment.



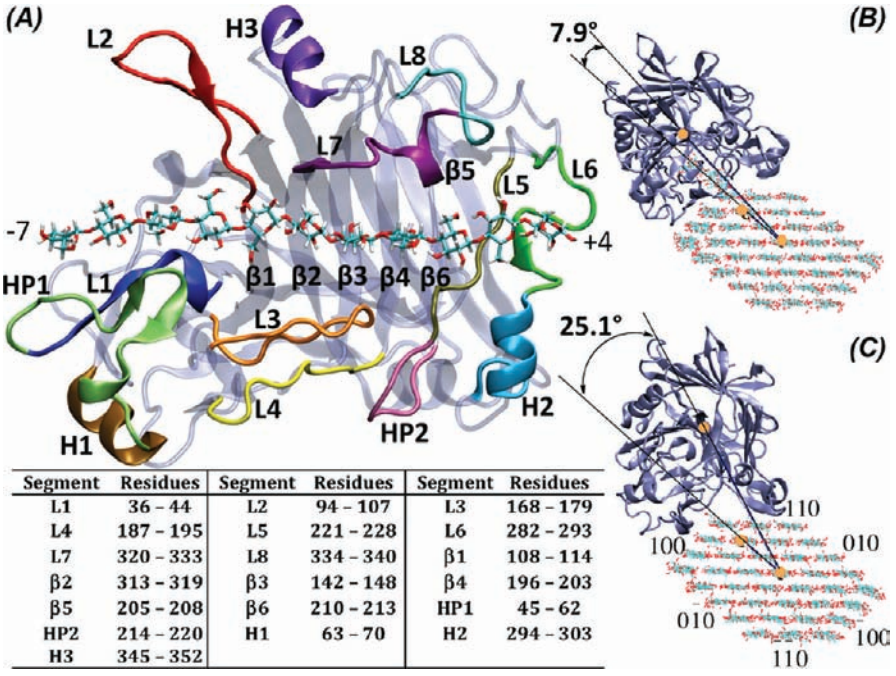
1. INTRODUCTION

Proteins tend to evolve allostery in their structures to fulfill machinelike activities, such as signal transduction and work generation.[1−7] Experimental and theoretical studies have revealed many of the underlying mechanisms.[8−16] The topology of native structures is recognized as a robust indicator of the likely shapes of conformational changes.[17−24] Sequence-specific details, such as correlation in sequence alignment, can also reveal the potential pathways of allosteric coupling in a protein structure.[25−27] In some cases, sequence correlation between residues can be captured on the basis of their coupling strengths in the protein structure network calculated from all-atom MD simulations.[28] Intrinsic structural transitions such as local unfolding could also play important roles in causing allosteric responses.[29−32]

At phase boundaries, on the other hand, protein allostery[33,34] and the physical steps in biocatalysis are poorly understood due to the difficulties of characterizing protein action with high spatial and temporal resolution. Cellulase cocktails secreted by fungi and bacteria to deconstruct crystalline cellulose in biomass into soluble sugars are a prominent example of interfacial enzymes with the physical activities playing critical roles for function.[35−40] In plant cell walls, cellulose appears as slender aggregates of glucan chains called microfibrils[41] that contain three components in the interaction network: intrachain OH···O hydrogen bonds (HBs), interchain OH··· HBs within a flat sheet, and intersheet CH···O interactions (pseudo HBs) between staggered sheets.[41−45] The recalcitrance of cellulose microfibrils presents a major difficulty in utilizing biomass as a viable feedstock for producing fuels.

The fungus *Trichoderma reesei* (*Hypocrea jecorina*) is of particular interest, as it secretes an effective enzyme cocktail at very high protein concentrations.[38−40] *T. reesei* cellulases, such as endoglucanase I (Cel7B), typically contain a catalytic domain (CD) and a carbohydrate-binding module (CBM) connected by a flexible, O-glycosylated linker.[46−49] The CBM of *T. reesei* cellulases significantly enhances the activity toward crystalline cellulose.[46−49] In addition to raising local enzyme concentration and elongating

Received: July 20, 2011
Published: August 30, 2011



Segment	Residues	Segment	Residues	Segment	Residues
L1	36 - 44	L2	94 - 107	L3	168 - 179
L4	187 - 195	L5	221 - 228	L6	282 - 293
L7	320 - 333	L8	334 - 340	β1	108 - 114
β2	313 - 319	β3	142 - 148	β4	196 - 203
β5	205 - 208	β6	210 - 213	HP1	45 - 62
HP2	214 - 220	H1	63 - 70	H2	294 - 303
H3	345 - 352				

Figure 1. Structures of the Cel7B CD. (A) Ribbon representation with structural segments around the binding cleft highlighted; their residues are listed in the table below. (B) Snapshot of the complexed-on-microfibril simulation after equilibration. An 11-residue glucan chain at a corner of the microfibril is detached from the reducing end (inward) and complexed in the binding cleft. The relative orientation of the center of mass of the Cel7B CD with respect to the microfibril is shown. (C) Snapshot of the adsorbed-on-microfibril simulation. The glucan chain of interest is intact in its crystallinic arrangement. The relative orientation of the center of mass of the Cel7B CD with respect to the microfibril is also shown. Surfaces of the microfibril are labeled according to their crystalline indices.

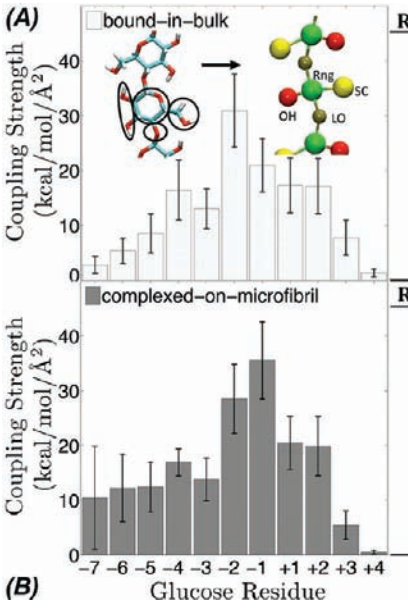
enzyme residence time on the solid substrate, whether CBM exhibits an active role in disrupting cellulose structures is still under debate.[50−52] The O-glycosylated linker might also play specific roles in Cel7B function.[53] In any case, the 371-residue Cel7B CD is responsible for hydrolytic cleavage of glycosidic bonds on cellulose microfibrils. The 50 Å long active-site cleft of Cel7B (Figure 1A) suggests that a glucan chain from the cellulose surface needs to be complexed in for hydrolysis.[35−37]

The loops extending from the binding cleft of the Cel7B CD[35] are likely responsible for complexing a targeted glucan chain to form a catalytically active complex on the cellulose microfibril. This theory is supported by the structural similarity between the Cel7B CD and the CD from the processive *T. reesei* Cel7A enzyme.[36,37] The two CDs have similar structures, except that the Cel7A CD has four additional loops with which to enclose the active site to form a tunnel instead of a cleft.[35−37] Therefore, the common structural segments in the CDs of Cel7A and Cel7B are likely sufficient to complex a glucan chain from cellulose. The extra segments in the Cel7A CD may impart processivity via increased ligand binding free energy, whereas the relatively lower ligand binding free energy in Cel7B imparts the ability to complex the cellulose surface, hydrolyze glucan chains, and then diffuse away from the surface of cellulose randomly.[40,54]

As a first step to elucidate the mechanism of complexing glucan chains on cellulose, we analyzed how attachment of the Cel7B CD to the cellulose microfibril affects glucan chain clenching in the active-site cleft using explicit-solvent, all-atom molecular dynamics (MD) simulations. On a model microfibril based on the X-ray structures and atomic force microscopy (AFM) images of cellulose I_α,[41−44] we created a free end in an edge glucan chain, as corner chains can be accessed with less free-energy cost,[55] detached 11 glucose residues from the reducing end, and complexed the chain into the binding cleft of the Cel7B CD (see Figure 1). The oligosaccharide-bound structure of the Cel7A CD[37] was used to inform the docking position of the detached glucan chain in the initial structure; other details of the Cel7B-microfibril simulation models are described in Supporting Information. We also performed separate simulations of the Cel7B CD adsorbed onto the microfibril with the glucan chain of interest retained in its crystalline configuration as well as simulations of the Cel7B CD in bulk solution with and without a bound glucan chain. A total of four sets of Cel7B CD simulation were performed: (a) complexed-on-microfibril, (b) adsorbed-on-microfibril, (c) bound-in-bulk, and (d) apo-in-bulk.

2. COMPUTATIONAL METHODS

2.1. All-Atom MD Simulations of the Cel7B Catalytic Domain. The X-ray structure from Kleywegt and co-workers[35] (PDB ID 1EG1) was used as the initial configuration. The glucan-bound structure of Cel7A CD (PDB ID 6CEL)[37] was best fit to that of Cel7B as the initial structure of the glucan chain in the Cel7B CD. The CHARMM22 all-atom force field with the CMAP correction[56,57] and the TIP3P water model were used for MD simulations. Langevin dynamics with a damping coefficient of 0.5 ps^{-1} was used to maintain the system at the optimal temperature of Cel7B at 313 K,[35] and the Langevin piston method was used to maintain pressure at 1 atm with the same damping coefficient.[35] All-atom MD simulations were performed with the NAMD software.[58] System setup and other analyses were performed with the CHARMM software[59] and in-house codes. Figures of protein structures were prepared via VMD.[60] More details of setting up the simulation models and MD protocols can be found in Supporting Information.



Residue	Axial	Equatorial
-7	62%: W40	38%: W40, E73
-6	75%: W40	25%: W40, R39
-5	56%: Y38, R39	44%: Y38, R39
-4	62%: Y38	38%: Y38, S106
-3	66%: S106, R108	34%: R108, S106
-2	55%: W320, R108	45%: W320, S318
-1	46%: W320, S144	54%: S144, E201
+1	50%: T210, H212	50%: T210, E201
+2	61%: T210, G223	39%: T210, G223
+3	71%: G225, A222	29%: F226, C224

Residue	Axial	Equatorial
-7	65%: S48, W40	35%: S48, W40
-6	54%: W40, V51	46%: V51, W40
-5	60%: Y38, V51	40%: R39, W40
-4	68%: Y38	32%: V105, N37
-3	67%: Y38, R108	33%: Y38, S106
-2	51%: W320, R108	49%: W320, R108
-1	58%: W320	42%: S144, W320
+1	58%: T210, E196	42%: T210, E201
+2	53%: T210, S221	47%: S221, G225
+3	55%: C224, G225	45%: C224, G225

Figure 2. Strengths of mechanical coupling (in kcal/mol/Å^2) of the Cel7B CD to glucose residues in the glucan chain bound in the binding cleft. The coupling strength between a residue pair is the sum of all the force constants of the elastic bonds that connect their CG sites. The coarse graining of a glucose residue is shown in the inset. Over a 100-ns trajectory, an ENM is calculated every 4 ns for determining the averages and standard deviations of the coupling strengths between protein and glucose residues. (A) Coupling strengths in the bound-in-bulk simulation. (B) Coupling strengths in the complexed-on-microfibril simulation. The percentages of equatorial coupling to the hydroxyl group (OH) and side chain (SC) sites and axial coupling to the sugar ring (Rng) and linker oxygen (LO) sites for each glucose residue are listed, with the top two protein residues contributing to most of the coupling strength.

2.2. Coarse-Grained Elastic Network Model. Only the coarse-grained (CG) sites mapped from protein and glucose atoms are employed in the elastic network model (ENM).[61,62] The procedures of calculating bond lengths and force constants from atomic fluctuations follow those reported in a previous work.[28] In our implementation, each protein residue in the ENM has a backbone site based on the positions of C_α atoms and a side-chain site based on the centers of masses of side-chain atoms; glycine has a single site. Each glucose residue has four CG sites as shown in Figure 2 to represent sugar rings, linker oxygens, hydroxyl groups, and side chains separately. In total, the Cel7B−microfibril complex is represented by 3003 CG sites, within which 699 CG sites result from Cel7B CD. From each all-atom simulation, a distance cutoff of 7.5 Å is used to generate the pool of elastic bonds for fluctuation matching. Since the force constants are adjusted iteratively to match the statistics of intersite distances from all-atom MD, the results of fluctuation matching are not sensitive to the distance cutoff used for assigning initial connectivity.

2.3. Multiple Sequence Alignment and Statistical Coupling Analysis. Homologues of Cel7B were gathered from the NCBI's non-redundant database by use of GGSEARCH in the FASTA suite[63] as well as Pfam.[64] Sequences from the Pfam endoglucanase I family (P07981) were combined with the results from GGSEARCH with the duplicated sequences removed. Since GGSEARCH returns only globally alignable sequences, both the mature protein sequence (371 aa) and the sequence including the N-terminal signal peptide and propeptide (459 aa) were used as queries. For both the GGSEARCH and Pfam sequences, an initial alignment was constructed with MAFFT[65] and then truncated to positions in the mature *T. reesei* Cel7B structure (PDB 1EG1). Sequences were selected from these truncated alignments on the basis of number of alignable positions (no more than 185 gaps). After removal of redundant sequences (≥95% similarity) identified by BLASTClust,[66] the sequences were realigned and the resulting alignment of 340 sequences was used for calculating sequence conservation and conducting the statistical coupling analysis. The statistical coupling matrix was created as described in Halabi et al.,[27] and eigenvectors 2 and 3 of this matrix were used to assign three sectors. Our Mathematica code is based on the MATLAB code from ref 27 and is available upon request.

3. RESULTS AND DISCUSSION

3.1. Conformational Change in the Cel7B Catalytic Domain Induced by Glucan Chain Binding. Structural segments of the Cel7B CD are labeled in Figure 1A. Glucose residues in the bound glucan chain are numbered according to the conventional scheme.[35] The bound-in-bulk and apo-in-bulk simulations are both performed for 110 ns, with the last 100 ns used for analysis. In both simulations, the C_α RMSDs (root-mean-square-differences) of protein configurations from the initial X-ray structure are small (∼2 Å; Figure S1, Supporting Information). In the bound-in-bulk simulation, L7 (Trp320 to Gly333, Figure 1) clenches the glucan chain in the binding cleft. In the apo-in-bulk simulation, L7 moved away from L8 toward L3 to partially close the binding cleft, indicating that the loop dynamics in the Cel7B CD depends on glucan chain binding. Temporal evolution of the $C_\alpha-C_\alpha$ distances between Gln325 in L7 and Gln174 in L3 and between Gln325 and Ser340 in L8 is shown in Figure S2 (Supporting Information). Crossing in their values reveals the conformational change in L7. In the bound-in-bulk and the two interfacial simulations, these conformational changes did not occur, denoting that the cleft remains open when the ligand is present (Figure S2, Supporting Information). The conformational change of L7 indicates that interactions with the glucan chain in the binding cleft would affect the couplings of nearby protein segments and vice versa.

3.2. Glucan Chain Clenching in the Binding Cleft of the Cel7B Catalytic Domain. By coarse graining an all-atom MD trajectory into a sequential set of ENMs, the coupling strengths between protein and glucose residues are calculated by summing

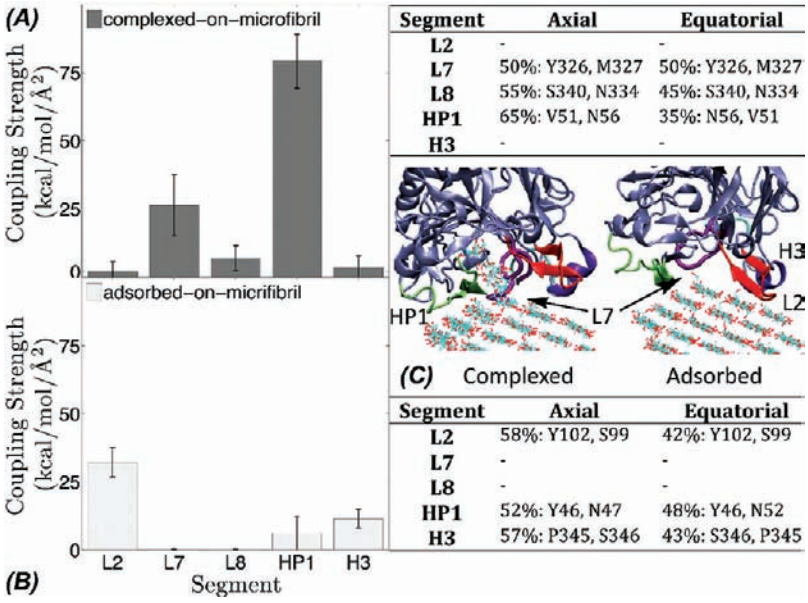


Figure 3. Strengths of mechanical coupling (in kcal/mol/Å^2) of structural segments of the Cel7B CD to glucose residues on the microfibril. Glucose residues in the bound glucan chain were excluded in the calculations. Coupling strengths were calculated from (A) complexed-on-microfibril and (B) adsorbed-on-microfibril simulations. The percentages of equatorial and axial coupling are listed, along with the top two protein residues contributing most of the coupling strength. (C) Representative snapshots of Cel7B−microfibril interactions observed in the complexed and adsorbed simulations.

the force constants of the elastic bonds that connect them. In clenching the glucan chain bound in the active-site cleft of the Cel7B CD, the elastic bonds to sugar rings or linker oxygens are considered axial coupling, and those to hydroxyl groups and side chains are equatorial coupling. From an all-atom MD trajectory, averages of site−site distances are calculated as bond lengths in the ENM, and variances of intersite distances are employed to determine the force constants of elastic bonds via a fluctuation-matching method.[24,28] Each bond in the ENM has an independent set of parameters to represent the sequence and structural specificity in mechanical coupling.

Over a long MD trajectory, we computed a separate ENM for every 4-ns segment to effectively capture the anharmonicity of protein structure fluctuations, and the temporal evolution of the ENM is termed the "fluctuogram" of protein dynamics.[28] Sequential ENMs are also used to compute the means and variances of coupling strengths between protein and glucose residues. The results shown in Figures 2 and 3 indicate that axial coupling of the Cel7B CD to the bound glucan chain tends to be stronger than equatorial coupling, consistent with the earlier observation that intersheet interactions in cellulose, with sugar rings as donors and oxygen-containing groups such as linker oxygens as acceptors, are more robust than interchain interactions between hydroxyl groups and side chains of glucose residues.[44]

In the bound-in-bulk simulation, clenching the glucan chain by the Cel7B CD is represented by the coupling strengths between protein and glucose residues. In Figure 2A, the coupling strength of each glucose residue in the bound chain is shown with the percentages coming from axial and equatorial couplings, as well as the top two Cel7B residues that contribute most of the coupling strength. To glucose residues −1 and −2 around the catalytic site, the highly conserved Trp320 dominates the coupling strength (see Figure 2A). Glucan chain clenching gradually becomes weaker toward the nonreducing end of increasingly negative residue numbers, and the main contributing protein residues are the highly conserved Tyr38 and Trp40. To fill in the gap between Tyr38 and Trp320, polar residues in nearby segments, such as Ser106 in L2 and Arg108 in β1, participate in clenching the bound glucan chain (see Figures 2A and 1A). To the reducing end of increasingly positive residue numbers, polar residues in β4, β6, and L5 of the Cel7B CD contribute significant coupling strengths.

3.3. Enhanced Glucan Chain Clenching in the Binding Cleft Due to Attachment of the Cel7B Catalytic Domain to the Cellulose Microfibril. In the complexed-on-microfibril simulation, a corner chain on the hydrophobic (110) surface of the microfibril is detached and complexed into the Cel7B CD. The orientation of the Cel7B CD with respect to the microfibril equilibrated after ∼35 ns is shown (Figure 1B). Similar Cel7B−microfibril orientation and surface coupling were also observed in three independent MD simulations starting with different initial velocities. One of the simulations was extended to 140 ns, and the 41−140 ns portion was used for analysis. The calculated coupling strengths of the attachment of the Cel7B CD to cellulose surface are plotted in Figure 3A. For each Cel7B segment listed in Figure 3, the top two protein residues contributing most of the coupling strength in surface attachment and the percentages of axial versus equatorial coupling are also specified.

Figure 3A,C illustrate that, among the structural segments of the Cel7B CD (see Figure 1A), HP1 (Asn45−Asp62) and L7 (Trp320−Gly333) strongly affix to cellulose surface in the complexed-on-microfibril simulation; other segments have only very weak coupling in surface attachment. HP1 situates at the junction where the detached chain is linked to the microfibril. Val51 in HP1 affixes to the detached chain and glucose residues on the (110) surface, while other HP1 residues attach to the (100) surface.

Detaching the corner chain leaves an extended step defect on the cellulose surface and exposes the glucose residues in the next layer. This topology allows residues in L7 (Tyr326 and Met327) to affix to the surface defect for attachment. With Tyr326 and

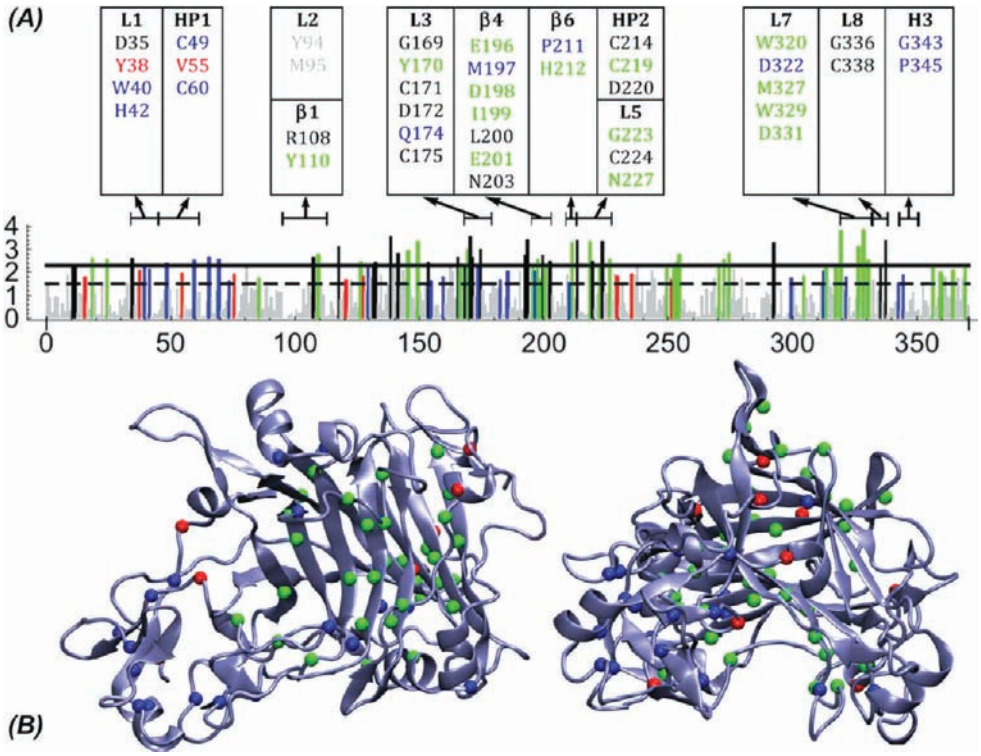


Figure 4. Conservation and correlation in the multiple sequence alignment, with the Cel7B CD as the query sequence. (A) Conservation (relative entropy to a reference probability distribution) of each amino acid. The solid line (2.3) is the cutoff for residues with strong conservation but low correlation with other residues. The dashed line (1.5) is the conservation cutoff for residues demonstrating significant sequence correlation with other residues. The three sectors in sequence correlation are clustered on the basis of the contributing values of each residue in the statistically significant eigenvectors of the sequence correlation matrix.[27] Residues with significant sequence correlation and a conservation higher than 1.5 are colored according to their sectors; residues with conservation higher than 2.3 but low sequence correlation are colored in black. Other residues are colored in gray. The structural segments that are responsible for glucan chain clenching in the binding cleft and attaching to the microfibril are labeled with the residues that have strong conservation or significant correlation. (B) Sector residues listed in panel A are shown as spheres in the X-ray structure of the Cel7B CD.

Met327 in L7 strongly affixing to the surface defect, glucan chain clenching in the binding cleft by Trp320, which is also in L7, may be affected. Similarly, with the attachment of HP1 to surface at the junction where the detached glucan is linked to the microfibril, the nearby glucan chain clenching in the binding cleft may become stronger as well. Indeed, as shown in Figure 2, coupling strengths of the Cel7B CD to the bound glucan chain are stronger in the complexed-on-microfibril simulation than those in the bound-in-bulk simulation, except at the reducing end. All-atom MD simulations thus establish that the L7 segment (Trp320−Gly333) plays at least two roles in complexation on cellulose: clenching the targeted glucan chain at Trp320 and affixing to the surface defect with residues around Tyr326 and Met327. HP1 also demonstrates dual physical activities of glucan chain clenching and surface attaching in the complexed simulation.

3.4. Complexation-Dependent Surface Attachment of the Cel7B Catalytic Domain to the Cellulose Microfibril. Before complexing with a glucan chain, the Cel7B CD adsorbs onto the targeted microfibril. To shed light on Cel7B−microfibril interactions in adsorption, we performed the adsorbed-on-microfibril simulation with the same initial structure as that in the complexed-on-microfibril simulation but with the detached glucan chain in its crystalline configuration. Without the step defect on the cellulose surface, as in the complexed simulation, the orientation of the Cel7B CD with respect to the microfibril equilibrated after ∼35 ns at a different value in the adsorbed-on-microfibril simulation (Figure 1C). Similar orientation and surface coupling were reproduced in three independent simulations with different initial conditions. One of the simulations was extended to 140 ns, and the 41−140 ns portion was used for analysis.

In the adsorbed-on-microfibril simulation, the calculated coupling strengths of the Cel7B CD to glucose residues on the microfibril are plotted in Figure 3B to contrast with those calculated from the complexed-on-microfibril simulation shown in Figure 3A. It is clear that the strong surface attachment of HP1 and L7 in the complexed simulation become very weak in the adsorbed-on-microfibril simulation. Instead, L2 and H3 strongly attach to the intact surface, but their coupling strengths to the defect surface in the complexed simulation are very weak. All-atom MD simulations at the solid−liquid interface of the cellulose microfibril thus illustrate that alternative loop segments of the Cel7B CD can affix to intact or defect surface structures depending on the complexation state.

3.5. Sequence Conservation, Coevolution, and the Mechanical Coupling Network in the Cel7B Catalytic Domain. Figure 4 shows the profile of sequence conservation among the 340 homologues of the Cel7B CD collected from the NCBI database. We also characterized the correlation in sequence alignment via a statistical coupling analysis (SCA) and clustered three sectors based on the statistically significant eigenvectors of the sequence correlation matrix.[27] In Figure 4, the 65 residues

having high correlation in sequence alignment with other residues (cutoff of 0.09 in the magnitude of eigenvector components) as well as high conservation (cutoff of 1.5 in relative entropy) are colored according to their sectors (blue, red, and green). The top 15% of the residues that are most conserved (cutoff 2.3 in relative entropy) but do not show significant sequence correlation with other residues are colored black. Together, a total of 90 residues in the Cel7B CD with high conservation and/or significant correlation in sequence alignment are highlighted in Figure 4 with explicit labeling of the glucan-clenching and surface-affixing segments discussed earlier.

Sequence conservation provides an indication of the importance of a protein residue in contributing to the required properties for function. In addition to hydrolyzing glycosidic bonds, complexing a glucan chain from cellulose is an important physical step for the Cel7B CD to perform endoglucanase activity. This theory is supported by the results of a multiple sequence alignment that, in addition to the active site and the binding cleft, the segments strongly affixing to the intact or defect surface structures on cellulose in MD simulations, such as L1, HP1, L7, and H3 (see Figure 4), also contain highly conserved residues.

Furthermore, correlation between residues in sequence alignment provides an indicator of their coupling in facilitating protein function.[25−27] Via division of residues into sectors according to the statistically significant eigenvector components of the sequence correlation matrix (Figure S3, Supporting Information), the random matrix theory provides a way to dissect the functional connection between protein residues.[25−27] The largest sector identified in the sequence correlation of the Cel7B CD is colored green in Figure 4. The green sector contains residues at the active site in β4 as well as in L7. Trp320, which strongly clenches the bound glucan chain at the active-site cleft, and Met327, which strongly affixes to the surface defect in the complexed simulation, are both in L7 and in the green sector. In combination with the results of all-atom MD simulations, the significant sequence correlation between residues in the active site and in L7 signals the functional connection between complexation and hydrolysis. The second largest sector in the Cel7B CD is colored blue in Figure 4, and many blue-sector residues are in L1 and HP1 that couple to the bound glucan chain at the junction where the chain is linked to the microfibril, including Trp40, His42, and the Cys49−Cys70 and Cys60−Cys66 disulfide bonds. Blue-sector residues are also found in β4 (Met197) at the active site and in L7 (Asp322) (Figure 4). This result provides further evolutionary signals in the functional couplings between complexation and hydrolysis. The smallest sector in the sequence correlation of the Cel7B CD is colored red in Figure 4, and red-sector residues scatter on the surfaces of the Cel7B CD, including Tyr38 in L1 and Val55 in HP1. Tyr38 in the binding cleft strongly couples to the bound glucan chain, whereas residues near Val55 strongly interact with the glucan chain at the junction where it links to the microfibril (Figures 2 and 3).

To bridge all-atom MD simulations with the conservation and correlation observed in sequence alignment, mechanical coupling strengths between CG sites provide a physics-based metric.[28] The force constants calculated by matching the intersite fluctuations sampled in all-atom MD simulations are used to determine the effective coupling strengths between protein and glucose residues. The net coupling strength associated with each amino acid in the Cel7B CD is then computed by summing the force constants of the associated elastic bonds. On the basis of the calculated coupling strengths, we devise criteria to screen for protein residues and examine if coupling strengths could be used as predicative metrics for identifying the highly conserved and significantly correlated residues shown in Figure 4.

In criterion A, we select protein residues on the basis of the number of strongly coupled neighbors, including glucose residues.[28] Criterion A thus uses a neighbor-number cutoff to screen protein residues for their numbers of neighbors to which the maximum coupling strength exceeds a strength cutoff. In criterion B, the Cel7B CD residues are screened for their averaged coupling strengths in a fluctuogram of protein dynamics.[28] In applying both criteria, the hit rates of the screened protein residues in capturing the highly conserved and significantly correlated residues shown in Figure 4 to within $\pm$1 in residue number increase with the strength cutoffs (see Figure S4 in Supporting Information). When criteria A and B are combined, all four fluctuograms calculated in this work give hit rates above 80%, significantly higher than those obtained by random selection. These results illustrate the functional relevance of the mechanical coupling network in protein structure. Since glucan chain clenching and microfibril attaching have distinct behaviors in different MD simulations, residues screened from the corresponding fluctuograms also vary. For example, Met327 is both highly conserved and significantly correlated as shown in Figure 4 and is only picked up by the fluctuogram of the complexed-on-microfibril simulation, in which Met327 strongly affixes to the surface defect.

4. CONCLUSIONS

At phase boundaries, physical activities such as substrate complexation play critical roles in driving biocatalysis. In deconstructing crystalline cellulose by cellulases secreted by bacteria and fungi, the physical activity of complexation is likely a limiting factor of the apparent rates of producing soluble sugars.[55] Here, we elucidated and compared the interactions of the Cel7B CD with a microfibril substrate as a function of the complexation state by all-atom MD simulations. The distinct mechanical coupling networks calculated from the simulations of Cel7B−glucan and Cel7B−microfibril complexes reveal a previously unresolved allosteric response at the solid−liquid interface: attachment of the Cel7B CD to the surface of cellulose microfibril affects glucan chain clenching in the binding cleft. Alternative segments of the Cel7B CD were found to affix to intact or defect surface structures on cellulose depending on the complexation state. Versatility in coupling to different surface structures on the solid substrate is thus likely a critical attribute for cellulases to deconstruct cellulose effectively.

The observed interfacial allostery of the Cel7B CD revealed that, in complexing a glucan chain from the microfibril, clenching the glucan chain in the binding cleft and affixing to the cellulose surface are interconnected. From a multiple sequence alignment, we also found that residues in the surface-affixing loops of the Cel7B CD not only have strong conservation but also have significant correlation with active-site residues. Such sequence-specific characters highlight the functional importance of the physical activity of attaching to surface structures and the functional connection between complexation and hydrolysis. This work further exemplifies that calculating mechanical coupling networks in protein structure and analyzing the fluctuograms of protein dynamics[28] provide a systematic framework for analyzing protein allostery at interfaces and protein−surface coupling and, thus, potentially improving the performance of industrial enzymes.

■ ASSOCIATED CONTENT

Supporting Information. Simulation details, four figures as described in the text, and complete refs 47 and 59. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail jwchu@berkeley.edu.

■ ACKNOWLEDGMENT

We acknowledge financial support from the DOE Office of the Biomass Program; subcontract ZGB-0-40593-01 from the National Renewable Energy Laboratory, and the College of Chemistry, University of California, Berkeley. We also acknowledge the computational resources provided by NERSC, which is supported by the Office of Science of the U.S. Department of Energy under Contract DE-AC02-05CH11231, the Texas Advanced Computing Center Ranger cluster under the National Science Foundation Teragrid Grant TG-MCB090159, and the Environmental Molecular Sciences Laboratory under Proposal 25651.

■ REFERENCES

(1) Grant, B. J.; Gorfe, A. A.; McCammon, J. A. *Curr. Opin. Struct. Biol.* **2010**, *20*, 142.
(2) Howard, J. *Mechanics of Motor Proteins and the Cytoskeleton*; Sinauer Associates, Inc.: Sunderland, MA, 2001.
(3) Deweese, J. E.; Osheroff, M. A.; Osheroff, N. *Biochem. Mol. Biol. Educ.* **2009**, *37*, 2.
(4) Yu, E. W.; Koshland, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9517.
(5) Cecchini, M.; Houdusse, A.; Karplus, M. *PLoS Comput. Biol.* **2008**, *4*, No. e1000129.
(6) Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. *Annu. Rev. Biophys.* **2010**, *39*, 23.
(7) Whitley, M. J.; Lee, A. L. *Curr. Protein Pept. Sci.* **2009**, *10*, 116.
(8) Chu, J.-W.; Voth, G. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13111.
(9) Mittermaier, A.; Kay, L. *Science* **2006**, *312*, 224.
(10) Zheng, W. J.; Brooks, B. R.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7664.
(11) Hanson, J. A.; Duclerstacit, K.; Watkins, L. P.; Bhattacharyya, S.; Brokaw, J.; Chu, J.-W.; Yang, H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18055.
(12) Koshland, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98.
(13) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88.
(14) Cui, Q.; Karplus, M. *Protein Sci.* **2008**, *17*, 1295.
(15) Csermely, P.; Palotai, R.; Nussinov, R. *Trends Biochem. Sci.* **2010**, *35*, 539.
(16) Liu, J.; Nussinov, R. *J. Mol. Biol.* **2010**, *396*, 1508.
(17) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9319.
(18) Zheng, W. J.; Brooks, B. R. *Biophys. J.* **2005**, *89*, 167.
(19) Maragakis, P.; Karplus, M. *J. Mol. Biol.* **2005**, *352*, 807.
(20) Chu, J.-W.; Voth, G. A. *Biophys. J.* **2007**, *93*, 3860.
(21) Bahar, I.; Chennubhotla, C.; Tobi, D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 633.
(22) Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. *Chem. Rev.* **2010**, *110*, 1463.
(23) Li, W.; Wolynes, P. G.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3504.
(24) Chu, J.-W.; Voth, G. A. *Biophys. J.* **2006**, *90*, 1572.
(25) Suel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. *Nat. Struct. Biol.* **2003**, *10*, 59.
(26) Shulman, A. I.; Larson, C.; Mangelsdorf, D. J.; Ranganathan, R. *Cell* **2004**, *116*, 417.
(27) Halabi, N.; Rivoire, O.; Leibler, S.; Ranganathan, R. *Cell* **2009**, *138*, 774.
(28) Silvestre-Ryan, J.; Lin, Y.; Chu, J.-W. *PLoS Comput. Biol.* **2011**, *7*, No. e1002023.
(29) Miyashita, O.; Onuchic, J.; Wolynes, P. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12570.
(30) Whitten, S.; Garcia-Moreno, B.; Hilser, V. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4282.
(31) Brokaw, J. B.; Chu, J.-W. *Biophys. J.* **2010**, *99*, 3420.
(32) Flynn, E. M.; Hanson, J. A.; Alber, T.; Yang, H. *J. Am. Chem. Soc.* **2010**, *132*, 4772.
(33) Park, J. H.; Scheerer, P.; Hofmann, K. P.; Choe, H.-W.; Ernst, O. P. *Nature* **2008**, *454*, 183.
(34) Vantilbeurgh, H.; Egloff, M.; Martinez, C.; Rugani, N.; Verger, R.; Cambillau, C. *Nature* **1993**, *362*, 814.
(35) Kleywegt, G.; Zou, J.; Divne, C.; Davies, G.; Sinning, I.; Ståhlberg, J.; Reinikainen, T.; Srisodsuk, M.; Teeri, T.; Jones, T. *J. Mol. Biol.* **1997**, *272*, 383.
(36) Divne, C.; Stahlberg, J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J.; Teeri, T.; Jones, T. *Science* **1994**, *265*, 524.
(37) Divne, C.; Ståhlberg, J.; Teeri, T.; Jones, T. *J. Mol. Biol.* **1998**, *275*, 309.
(38) Bansal, P.; Hall, M.; Realff, M. J.; Lee, J. H.; Bommarius, A. S. *Biotechnol. Adv.* **2009**, *27*, 833.
(39) Levine, S. E.; Fox, J. M.; Blanch, H. W.; Clark, D. S. *Biotechnol. Bioeng.* **2010**, *107*, 37.
(40) Chundawat, S. P. S.; Beckham, G. T.; Himmel, M. E.; Dale, B. E. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2*, 121.
(41) Ding, S.; Himmel, M. *J. Agric. Food. Chem.* **2006**, *54*, 597.
(42) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074.
(43) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300.
(44) Gross, A. S.; Chu, J.-W. *J. Phys. Chem. B* **2010**, *114*, 13333.
(45) Matthews, J. F.; Bergenstråhle, M.; Beckham, G. T.; Himmel, M. E.; Nimlos, M. R.; Brady, J. W.; Crowley, M. F. *J. Phys. Chem. B* **2011**, *115*, 2155.
(46) Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *Biochem. J.* **2004**, *382*, 769.
(47) Beckham, G. T.; et al. *Biophys. J.* **2010**, *99*, 3773.
(48) Linder, M.; Teeri, T. *J. Biotechnol.* **1997**, *57*, 15.
(49) Suurnäkki, A.; Tenkanen, M.; Siika-Aho, M.; Niku-Paavola, M.-L.; Viikari, L.; Buchert, J. *Cellulose* **2000**, *7*, 189.
(50) Esteghlalian, A.; Srivastava, V.; Gilkes, N.; Kilburn, D.; Warren, R.; Saddler, J. *Appl. Biochem. Biotechnol.* **2001**, *91−3*, 575.
(51) Herve, C.; Rogowski, A.; Blake, A. W.; Marcus, S. E.; Gilbert, H. J.; Knox, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15293.
(52) Beckham, G. T.; Matthews, J. F.; Bomble, Y. J.; Bu, L.; Adney, W. S.; Himmel, M. E.; Nimlos, M. R.; Crowley, M. F. *J. Phys. Chem. B* **2010**, *114*, 1447.
(53) Ma, B.; Tsai, C.-J.; Haliloglu, T.; Nussinov, R. *Structure* **2011**, *19*, 907.
(54) Beckham, G. T.; Bomble, Y. J.; Bayer, E. A.; Himmel, M. E.; Crowley, M. F. *Curr. Opin. Biotechnol.* **2011**, *22*, 231.
(55) Beckham, G. T.; Matthews, J. F.; Peters, B.; Bomble, Y. J.; Himmel, M. E.; Crowley, M. F. *J. Phys. Chem. B* **2011**, *115*, 4118.
(56) Mackerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584.
(57) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400.
(58) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.
(59) Brooks, B. R.; et al. *J. Comput. Chem.* **2009**, *30*, 1545.

(60) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.
(61) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905.
(62) Bahar, I.; Atilgan, A. R.; Jernigan, R. L.; Erman, B. *Proteins: Struct., Funct., Bioinf.* **1997**, *29*, 172.
(63) Pearson, W. R.; Lipman, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2444.
(64) Finn, R. D.; Mistry, J.; Tate, J.; Coggill, P.; Heger, A.; Pollington, J. E.; Gavin, O. L.; Gunasekaran, P.; Ceric, G.; Forslund, K.; Holm, L.; Sonnhammer, E. L. L.; Eddy, S. R.; Bateman, A. *Nucleic Acids Res.* **2010**, *38*, D211.
(65) Katoh, K.; Kuma, K.; Toh, H.; Miyata, T. *Nucleic Acids Res.* **2005**, *33*, 511.
(66) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J. H.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 3389.

■ NOTE ADDED AFTER ASAP PUBLICATION

This article was published ASAP on September 15, 2011. The last sentence of the second paragraph has been corrected. The corrected version was posted on September 26, 2011.