Supporting Information

Protein Allostery at the Solid-Liquid Interface: Endoglucanase Attachment to Crystalline Cellulose Affects Glucan Chain 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, University of California, Berkeley, Berkeley, California, United States*

[‡]*Department of Bioengineering, University of California, Berkeley,*

Berkeley, California, United States

[§] *Biosciences Center, National Renewable Energy Laboratory, Golden, CO, United States*

[¶]*National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO, United States*

[‖]*Department of Chemical Engineering, Colorado School of Mines, Golden, CO, United States*

[*](Email: jwchu@berkeley.edu)

Details of all-atom MD simulations. The protonation states of titratable residues are adjusted based on a solution with pH 4 and pKa calculations using a Generalized Born implicit solvent model and the isoelectric point and active site chemistry.[1]

The CHARMM22 all-atom force field [2] and the TIP3P water model [3,4] were used to perform MD simulations in the present study. The particle mesh Ewald method [5] was employed for calculating long-range electrostatics. The short-range non-bonded interactions are calculated with a cutoff of 12 Å with a switch function turned on at 10 Å. Before production runs, each simulation

model is energy minimized (100,000 steps), gradually heated to 313 K (+4K/ps over 100 ps via velocity reassignment), and equilibrated for 4 ns. During minimization, heating, and the first 3 ns of equilibration, the heavy atoms of the protein were restrained to their initial positions via harmonic potentials with a force constant of 1 kcal/mol/Å^2, after which all restrained potentials on the protein are removed. If a glucan chain in the binding cleft of Cel7B CD is present, the heavy atoms of glucans are restrained and released the same way as protein atoms.

X-ray crystal structure by Kleywegt et al. (PDB ID: 1EG1) [1] is used as the initial structure of Cel7B CD. The 11 glucose units chain for the substrate was constructed from X-ray crystal structure of CBH I from *T. reesei* (PDB ID: 8CEL) and theoretical model (PDB ID: 6CEL) by Divne *et al* [6,7]. The Cel7B structure was first best fit to that of CBHI in 6CEL according to matching the positions of substrate-binding sites and active sites, and the structure of the glucan chain in 6CEL is then docked into the binding cleft of Cel7B CD. The model of Cel7B CD with the bound glucan chain, which is referred to as the bound-in-bulk simulation, was in turn energy minimized, heated and equilibrated following the same procedure as described earlier. Both of bound-in-bulk and apo-in-bulk simulations were performed for 110 ns with the last 100 ns used for analyses.

An equilibrated 36 glucan-chain microfibril [8] based on the I_α form of cellulose crystal was employed as the model microfibril. The initial structure for the complex-on-microfibril simulation was constructed based on the initial structure of the bound-in-bulk simulation. The position of an 11-residue glucan chain at the most exposed corner of the microfibril was replaced with that of the glucan chain bound in Cel7B with the non-reducing end linked to the microfibril by best fitting the structure of the bound chain to the chain on the microfibril. The coordinates Cel7B CD were also displaced and rotated along with those of the glucan chain in the structural best fit for exchanging glucan coordinates. Cel7B CD and the complexed chain were then lifted away from the microfibril via rigid body rotation to ensure no overlapping atoms, which is used as the initial structure for the complexed-on-microfibril simulation. For the adsorbed-on-microfibril simulation, the coordinates of the glucan chain in the binding cleft of Cel7B CD in the complexed-on-microfibril simulation are replaced with those of the intact microfibril with Cel7B CD retaining the same initial structure and orientation with respect to the microfibril. In the initial structure of the Cel7B CD, there are no-overlapping atoms with the microfibril and the loops extended from the binding cleft do not have close contacts with the surface of the microfibril.

The all-atom MD simulations of the Cel7B-microfibril complexes in explicit water were running at 313 K and 1 atm. During simulations, atoms at the two ends of each glucan chain were restrained to their initial positions via harmonic potentials with a force constant of 2000 kcal/mol/Å^2 to retain the crystalline structures of the microfibril with a finite size. 140 ns production-stage trajectories were performed for both on-microfibril simulations. Since the orientation respect to the microfibril reached the plateau value after 40 ns in both simulations, only the last 100ns data were collected for analyses and computing mechanical coupling strengths.

A time step of 2 fs was used to propagate dynamics simulations during which all covalent bonds associated hydrogen atoms were constrained at their equilibrium values defined in the CHARMM parameter. To neutralize the simulation models of Cel7B and to mimic the biological ionic strength, 38 Na^+ and 36 Cl^- ions were added to the two bulk simulations, while 64 Na^+ and 62 Cl^- ions were added to the two Cel7B-microfibril simulations. A total of 19,057 and 18,969 water molecules were used to solvate Cel7B in a cubic unit cell for the bound-in-bulk and apo-in-bulk simulations, respectively. Each of the two Cel7B-microfibril models contains 26195 water molecules. The models are further extended with periodic boundary conditions.

References

(1) Kleywegt, G.; Zou, J.; Divne, C.; Davies, G.; Sinning, I.; Stahlberg, J.; Reinikainen, T.; Srisodsuk, M.; Teeri, T.; Jones, T. *J. Mol. Biol.* **1997**, *272*, 383.

(2) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.

(3) Mackerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584.

(4) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400.

(5) Darden, T.; York, D.; Pederson, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(6) Divne, C.; Stahlberg, J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J.; Teeri, T.; Jones, T. *Science* **1994**, *265*, 524.

(7) Divne, C.; Stahlberg, J.; Teeri, T.; Jones, T. *J. Mol. Biol.* **1998**, *275*, 309.

(8) Gross, A. S.; Chu, J.-W. *J. Phys. Chem. B* **2010**, *114*, 13333.

(9) Silvestre-Ryan, J.; Lin, Y.; Chu, J.-W. *Plos Comput Biol* **2011**, *7*, e1002023.

Figure. S1. The temporal evolution of the C_α RMSDs to the initial protein structure for all 4 trajectories.

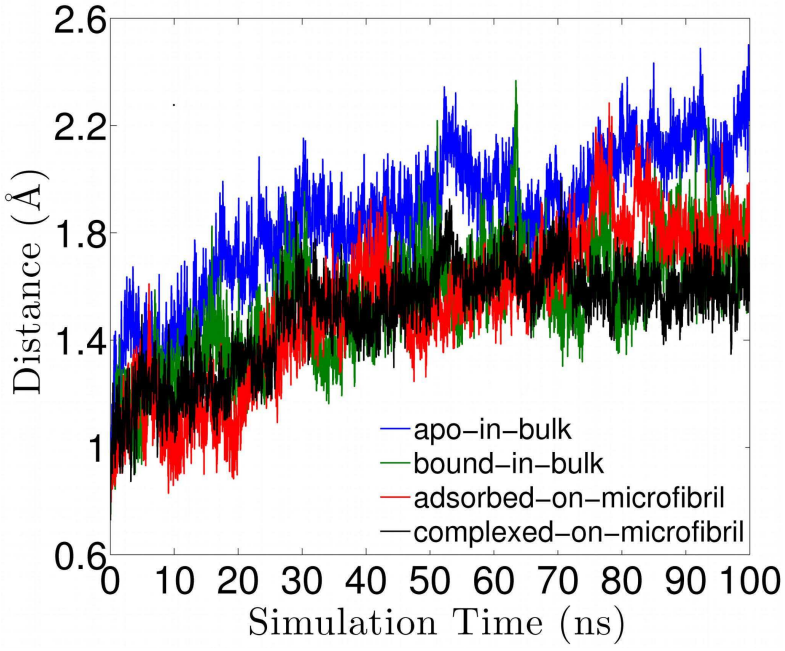


Figure. S2. The conformation change of L7 in the apo-in-bulk simulation. (A) The initial and final structures, L7 is colored in purple, L8 is colored in cyan, and L3 is colored in orange. (B) The conformational change is described by the temporal evolution of the C_α-C_α distances between Gln325 in L7 and Ser340 in L8 and between Gln325 and Gln174 in L3. The crossing in their values indicates that L7 is moving away from L8 toward L3 to partially close the binding cleft. The corresponding distance in (C) the bound-in-bulk simulation, (D) the complexed-on-microfibril simulation, and (E) the adsorbed-on-microfibril simulation.

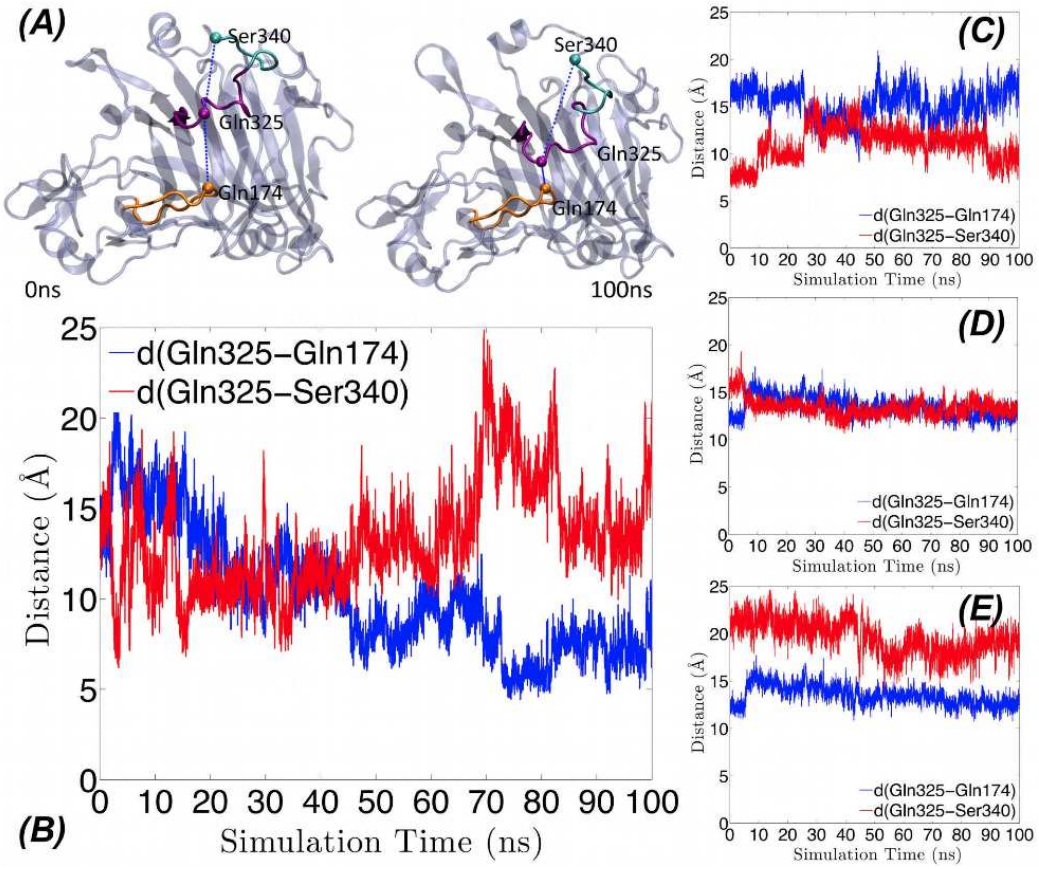


Figure. S3. The sequence correlation matrix of Cel7B CD and the annotation of sectors.

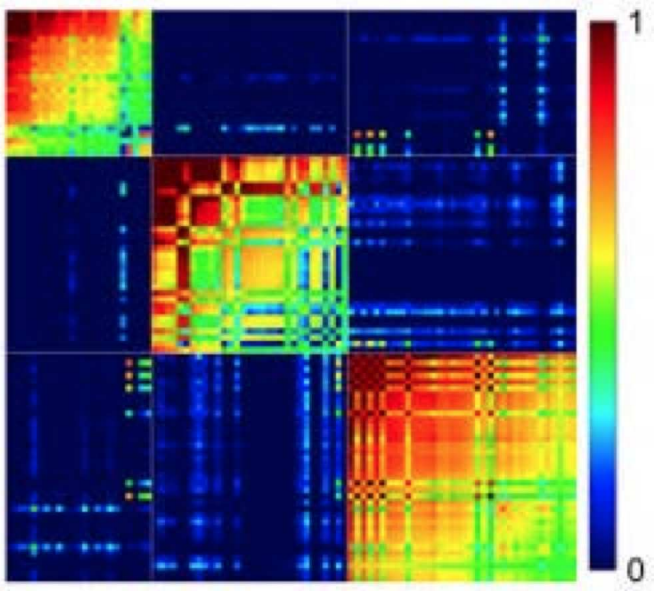
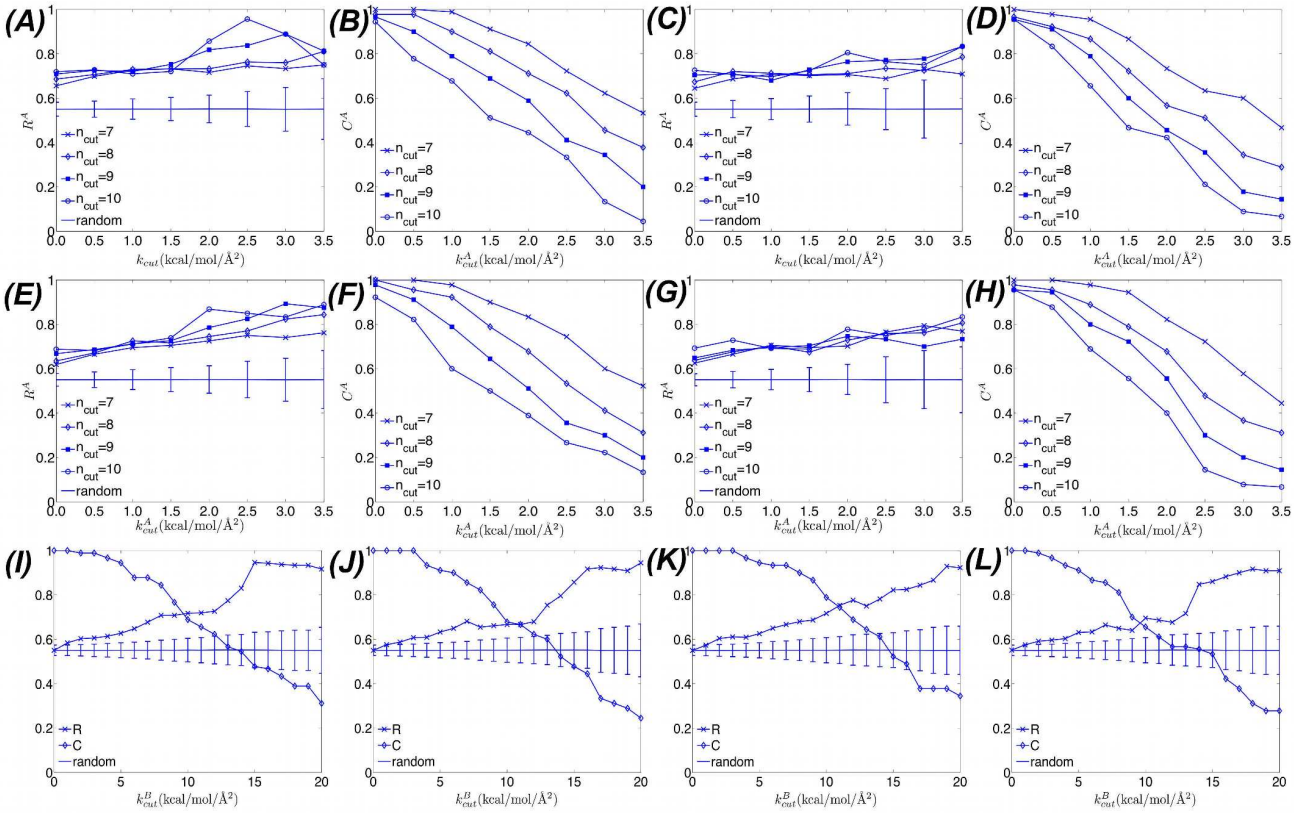


Figure. S4. The hit rates *(R)* and coverages *(C)* of using criterion-A and criterion-B described in the main Text to capture the highly conserved and co-evolved residues shown in Fig. 4. The hit rates ((number of predicted residues hitting those in Fig.4 to within ±1 in residue number)/(number of predictions)) and coverages ((number of residues in Fig. 4 got hit by the predicted residues)/(number of residues in Fig. 4)) are defined in the same way as in [9]. Both hit rate and coverage can have a maximum value of one and a minimum value of zero. Criterion-A screens for protein residues with the largest force constant observed in the fluctuogram larger than a coupling-strength cutoff and with the number of neighbors satisfying this requirement larger than a neighbor-number cutoff. Criterion-B screens for protein residues with the averaged coupling strengths of the fluctuogram larger than a coupling-strength cutoff. Hit rates and coverages of using criterion-A in bound-in-bulk simulation (A and B); hit rates and coverages of using criterion-A in apo-in-bulk simulation (C and D); hit rates and coverages of using criterion-A in complexed-on-microfibril simulation (E and F); hit rates and coverages of using criterion-A in adsorbed-on-microfibril simulation (G and H). Results of using criterion-B are shown in (I) for bound-in-bulk, (J) for apo-in-bulk, (K) for complexed-on-microfibril, and (L) for adsorbed-on-microfibril simulations.



Complete Ref. 47 and Ref. 59.

(47) Beckham, G. T.; Bomble, Y. J.; Matthews, J. F.; Taylor, C. B.; Resch, M. G.; Yarbrough, J. M.; Decker, S. R.; Bu, L.; Zhao, X.; McCabe, C.; Wohlert, J.; Bergenstrahle, M.; Brady, J. W.; Adney, W. S.; Himmel, M. E.; Crowley, M. F. *Biophys. J.* **2010**, *99*, 3773.

(59) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.