Cellulose-specific Type B Carbohydrate Binding Modules : Understanding Substrate Recognition Mechanisms Through Molecular Simulation

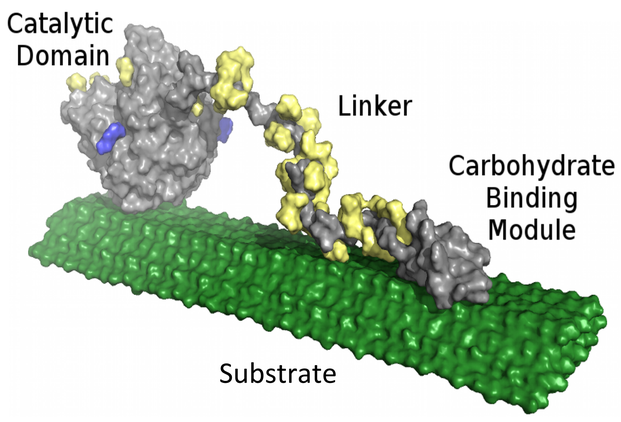
Abhishek A. Kognole[1] and Christina M. Payne[1]

1. Department of Chemical and Materials Engineering, University of Kentucky, Lexington, KY, USA.

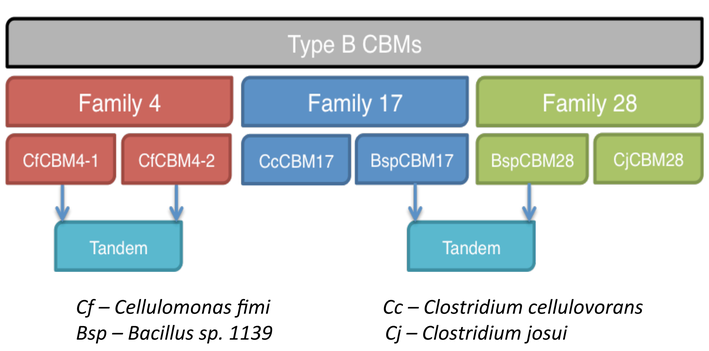
University of Kentucky

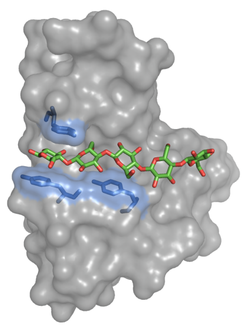
Background

Multi-modular glycoside hydrolases generally consist catalytic domains (CD) appended by linker peptides to **carbohydrate binding modules (CBM)**. The CD is responsible for cleaving the glycosidic linkages of cellulose. The non-catalytic CBM assists the CD in targeting the substrate and serves as the primary biological means of carbohydrate recognition by enzyme.

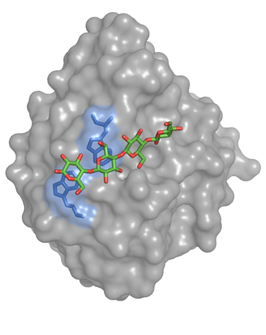


- Family 4, 17, and 28 Type B CBMs exhibit a beta-sandwich fold and have groove- or cleft-shaped binding sites. The representatives we examine here are cellulose specific and are capable of binding both single carbohydrate chains and non-crystalline/amorphous cellulose.





CfCBM4-1
Sandwich Platform

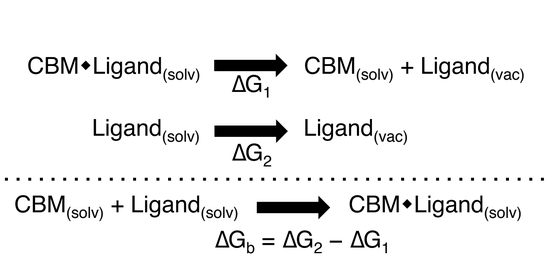


CcCBM17
Twisted Platform

- Using molecular simulations and free energy calculations, we investigate the molecular-level structural and dynamical features contributing to oligomeric and non-crystaline carbohydrate recognition in three families of Type B CBMs, providing critical details necessary for development of biomass conversion biotechnology.

Methods

- Molecular dynamics simulations were constructed from PDBs in CHARMM.
  - Force fields: CHARMM36 w/ CMAP correction for proteins; CHARMM 36 carbohydrates for cellopentaose, and modified TIP3P for water
  - Minimization, heating to 300 K, and 0.1 ns equilibration in the *NPT* ensemble
  - Data collection for 250 ns in the canonical ensemble in NAMD (~30000 atoms)
- Binding Free energy calculated using free energy perturbation with Hamiltonian replica exchange molecular dynamics in NAMD (3)
  - The Potential energy expressed independently as repulsion, dispersion, electrostatics, and restraints – scaled by thermodynamic coupling parameters.
  - Multistate Bennett Acceptance Ratio used to determine free energy and statistical uncertainty of energy components.



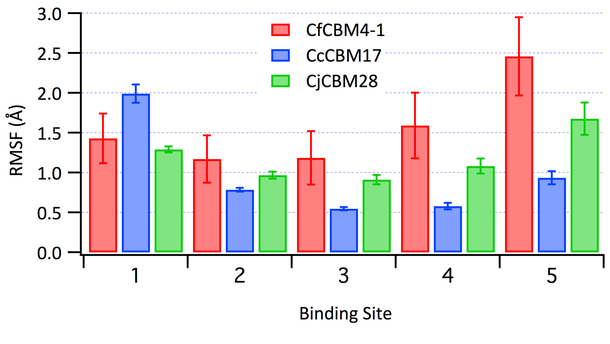
Thermodynamic cycle used to determine ligand binding free energy from FEP/HREMD. "solv" and "vac" refer to solvated and vacuum (or decoupled) systems, respectively.

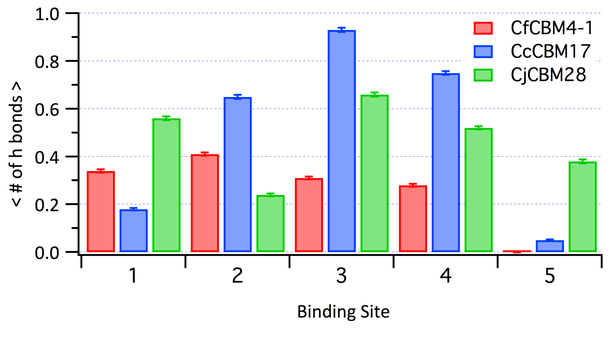
Cello-oligomer binding in different Type B CBM families

- The higher affinities of Family 17 & 28 CBMs for cellopentaose, relative to Family 4, suggest that their 'twisted' binding cleft configuration necessitates tighter ligand binding than the 'sandwich' platform of Family 4 CBM binding site.

ΔG of binding of Cellopentaose to	***Cf*CBM4-1** (kcal/mol)	***Cc*CBM17** (kcal/mol)	***Cj*CBM28** (kcal/mol)
Experimental (ITC)	- 5.24 ± 0.9 (1)	- 5.8 ± 0.03 (2)	- 7.7 ± 0.6 (3)
Computational (FEP/HREMD)	- 4.5 ± 1.3	- 6.9 ± 0.9	- 6.3 ± 0.7

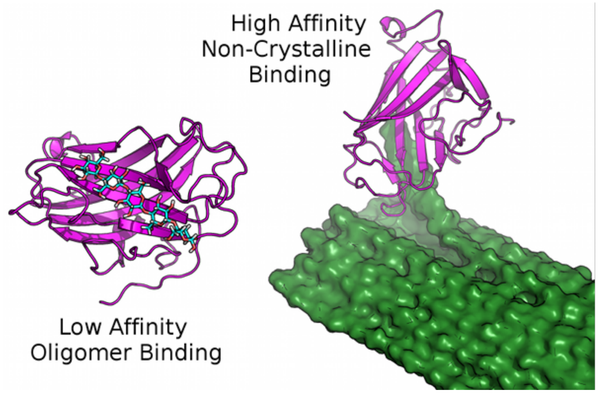
- MD simulations support the binding free energy calculations, revealing that *Cc*CBM17 and *Cj*CBM28 form a more stable non-covalent interaction with the cellopentaose ligand. Root mean square fluctuation (RMSF) of ligand and Hydrogen binding per binding site shown below.





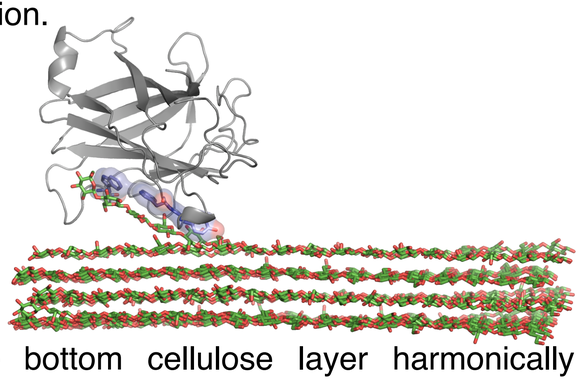
Non-crystalline cellulose binding in Type B CBMs

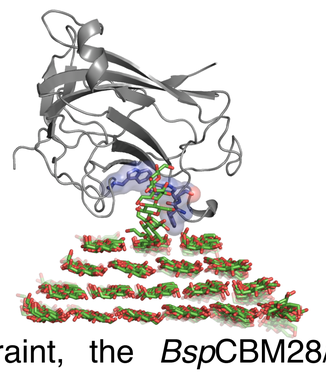
We investigate non-crystalline carbohydrate recognition in *Bsp*CBM28 and *Cc*CBM17, as experimental evidence suggests these two families demonstrated both low and high affinity binding sites on regenerated cellulose (4). We hypothesize this phenomenon possibly correlates to oligomeric and non-crystalline cellulose binding.



Modeling of non-crystalline cellulose recognition by *Bsp*CBM28

- To model non-crystalline cellulose, we used structures of crystalline cellulose from previous studies. Using targeted MD simulation, we decrystallized a single middle chain from the top crystalline layer of the microfibril to create a free glycan chain.
- Based upon the *Bsp*CBM28-cellopentaose system from oligomeric simulations, we aligned the CBM over non-crystalline cellulose such that the glycan chain end bound in the cleft. This system was simulated for 100ns after solvation and equilibration.

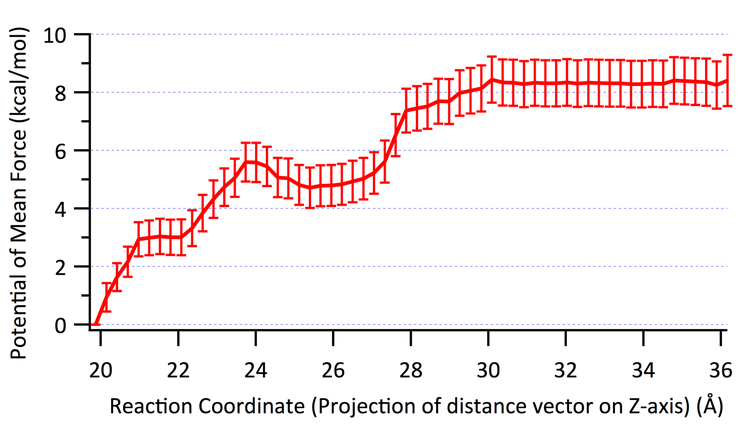




- With the bottom cellulose layer harmonically restraint, the *Bsp*CBM28/non-crystalline cellulose complex reaches local equilibrium after approximately 30 ns and maintains a stable interaction with the substrate for the remaining 70 ns.
- Interaction energy analysis reveals carbohydrate recognition is almost entirely mediated by two peptide loops; these loops encompass the aromatic residues forming the twisted platform as well as a key pair of acidic residues external to the binding cleft.

Thermodynamics of non-crystalline carbohydrate recognition

- Using umbrella sampling, we determined the work required to dissociate the CBM from non-crystalline cellulose..
- From the end points of the reaction coordinate, we obtained the free energy of binding of *Bsp*CBM28 to non-crystalline cellulose microfibril.



	Substrate	**ΔG for High affinity site** (kcal/mol)	**ΔG for Low affinity Site** (kcal/mol)
Experimental	Regenerated cellulose	- 8.28 ± 0.35	- 5.93 ± 0.38
Computational	Cellulose microfibril modeled as non-crystalline substrate and Cellopentaose	- 8.3 ± 0.8	- 5.0 ± 1.2

Conclusions

- Using FEP/HREMD and MD simulation, we elucidated a key difference in cello-oligomer binding across the three evaluated CBM families; the twisted, solvent exposed binding grooves of Family 17 and 28 CBMs necessitate tighter substrate binding than the sandwich-like Family 4 CBMs.
- We elucidated the mechanisms of non-crystalline carbohydrate recognition by modeling a Family 28 CBM complexed with a partially decrystallized cellulose substrate. Comparing both protein-carbohydrate interactions and ligand binding free energies, which were within error of experimental values, we have partially validated the correlation of high- and low-affinity binding sites with non-crystalline and oligomeric binding, respectively.
- In future, we will investigate oligomeric carbohydrate recognition in remaining Type B CBMs and will model binding of *Cc*CBM17 over non-crystalline cellulose to corroborate these results. We also intend to study the tandem CBMs to relate these recognition mechanisms to evolution of tandem systems.

References and Acknowledgements

1. Tomme P, Creagh AL, Kilburn DG, and Haynes CA (1996) *Biochemistry*, **35**, 13885-13894.
2. Notenboom V, Boraston AB, Chiu P, Freelove ACJ, Kilburn DG, Rose DR (2001) *J. Mol. Biol.*, **314,** 797-806.
3. Araki Y, Karita S, Tanaka A, Kondo M, and Goto M (2009) *Biosci. Biotechnol. Biochem.*, **73(5),** 1028-1032.
4. Boraston AB, Kwan E, Chiu P, Warren RAJ, Kilburn DG (2003) *J. Biol. Chem.*, **278(8),** 6120-6127.

This material is based upon work supported by the **National Science Foundation** under **Grant No. CHE-1404849.** Computational time was provided by the Extreme Science and Engineering Discovery Environment, which is supported by the National Science Foundation Grant No. ACI-1053575 under allocation MCB090159. Additional computing resources were made available by the Center for Computational Sciences at the University of Kentucky.