Cello-oligomer Binding Dynamics and Directionality in *Cellulomonas fimi* Family 4 Carbohydrate Binding Modules (CBMs)

Abhishek A. Kognole[1] and Christina M. Payne[1,2]

1. Department of Chemical and Materials Engineering, University of Kentucky, Lexington, KY
2. Center for Computational Sciences, University of Kentucky, Lexington, KY

University of Kentucky

Background

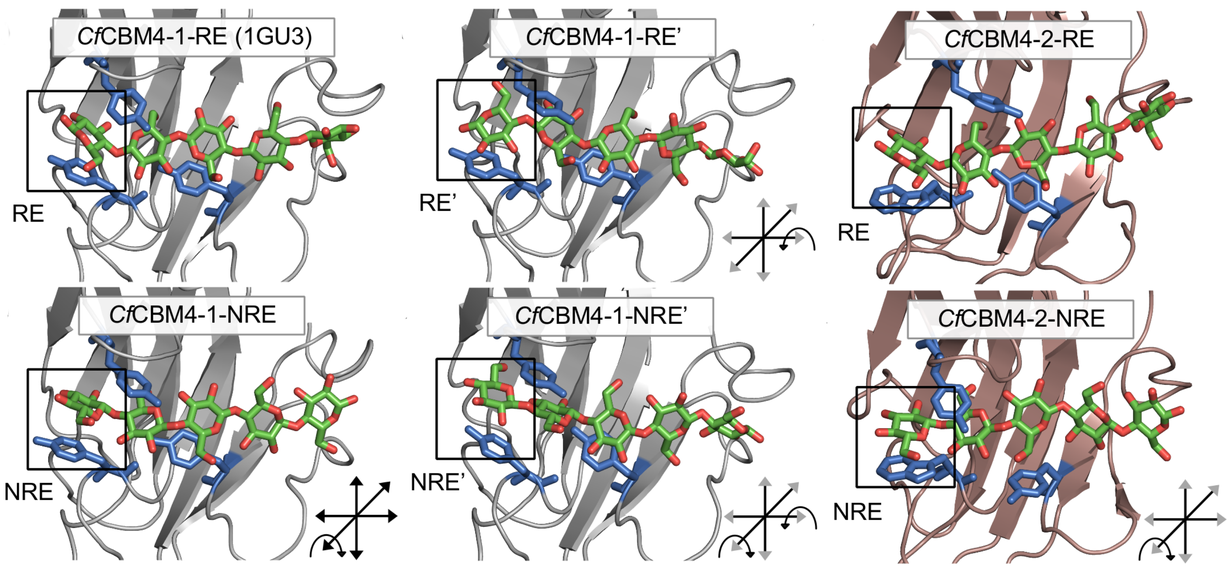
- β-1,4-glucanase CenC from *Cellulomonas fimi* has two tandem Type B CBMs, *Cf*CBM4-1 and *Cf*CBM4-2. Each exhibits the β-jelly roll fold, forming a cleft that binds oligosaccharides and amorphous cellulose.
- Binding studies of *Cf*CBM4-1 and *Cf*CBM4-2 have not conclusively determined the orientation of the bound cello-oligomers in the cleft.
  - NMR spectroscopy suggests cellopentaose binds bi-directionally (1).
  - X-ray crystallography has captured only one ligand orientation (2).
- Our objective is to understand how the orientation of the ligand affects the binding properties and determine which orientations are preferred; at the same time, these results provide general insight into the mechanisms of protein-carbohydrate recognition mechanisms.

Molecular Dynamics and Free Energy Calculations

The cellopentaose may bind to the CBM4s in four possible orientations. These four orientations differ from each other based on:

- The position of reducing end (RE) of the ligand in the binding cleft
- The orientation of hydrophilic pyranose side chains in a given binding site

All four were considered in the case of *Cf*CBM4-1 and two for *Cf*CBM4-2.



- 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*
  - Data collection for 250 ns in the canonical ensemble in NAMD (~27000 atoms)
- Free energy calculated using free energy perturbation with Hamiltonian replica exchange molecular dynamics in NAMD (3)
  - System 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.

$$\text{CBM}\bullet\text{Ligand}_{(solv)} \xrightarrow[\Delta G_1]{} \text{CBM}_{(solv)} + \text{Ligand}_{(vac)}$$

$$\text{Ligand}_{(solv)} \xrightarrow[\Delta G_2]{} \text{Ligand}_{(vac)}$$

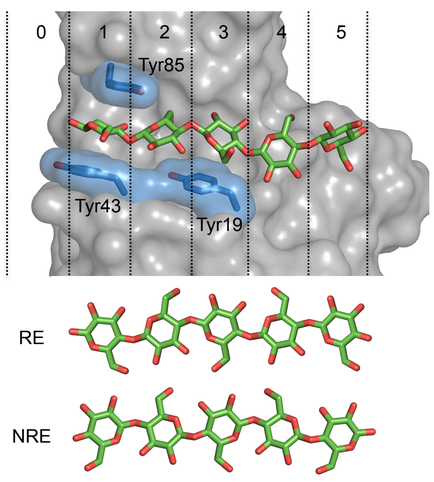
$$\text{CBM}_{(solv)} + \text{Ligand}_{(solv)} \longrightarrow \text{CBM}\bullet\text{Ligand}_{(solv)}$$

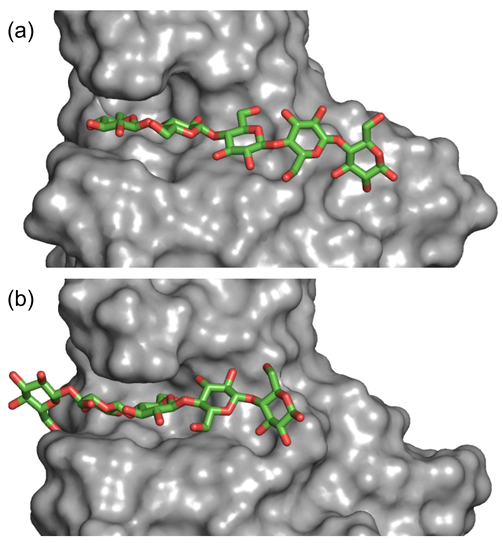
$$\Delta G_b = \Delta G_2 - \Delta G_1$$

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

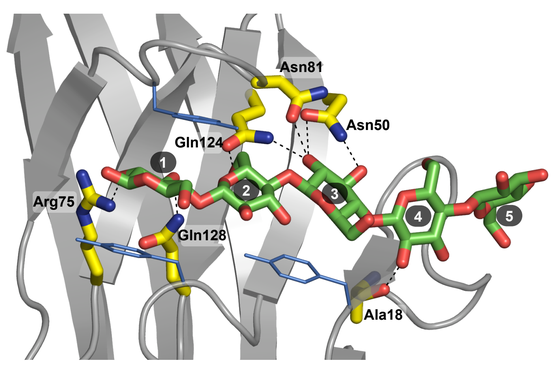
Effect of Cellopentaose Symmetry on Binding

- The approximate structural symmetry of oligo-saccharides accounts for the ability of the protein to bind the oligomer regardless of directionality.
- Reversing the direction of cellopentaose (*Cf*CBM4-1-NRE) does not change the structural symmetry, while rotation of pyranose ring along C1-C4 axis puts the hydroxymethyl groups on the opposite side of chain than that of the structural orientation disrupting symmetry.





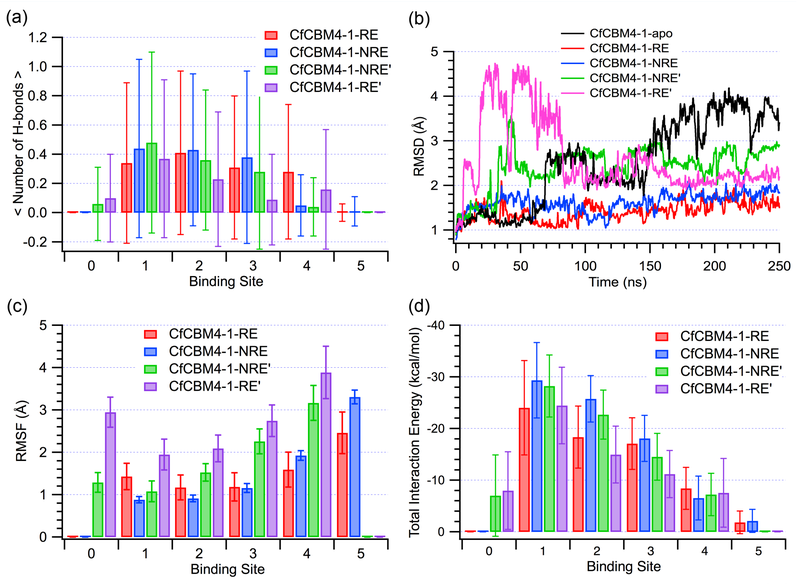
- The *Cf*CBM4-1 binding groove will not accept the hydroxymethyl groups in arbitrary sites.
- Hydrogen bonding in sites 1 to 3 determines oligomeric acceptance.



Snapshots of CfCBM4-1-NRE' at (a) 0 ns and (b) 2 ns of MD simulation.

Carbohydrate-Protein Binding Dynamics

- Same number of hydrogen bonds formed between ligand and binding site regardless of orientation (after RE' and NRE' shift)
- Protein unaffected by either RE or NRE orientation
- Flexibility of the RE and NRE ligands equal within error; RE' and NRE' affected by solvent exposed '0' site
- Equivalent site interactions suggest dynamics are same irrespective of orientation.



MD results: (a) average hydrogen bonds with ligand per site, (b) root mean square deviation (RMSD) of the protein backbone over time, (c) root mean square fluctuation (RMSF) of ligand per site, and (d) interaction energy of the ligand with each site.

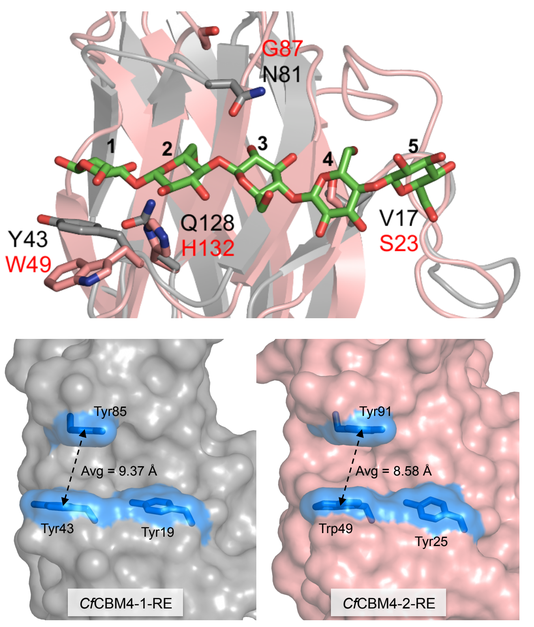
Thermodynamic Favorability

The binding free energies, ΔG_b°, of cellopentaose to *Cf*CBM4-1-RE and *Cf*CBM4-1-NRE are within error and are consistent with isothermal titration calorimetry (ITC) (4).

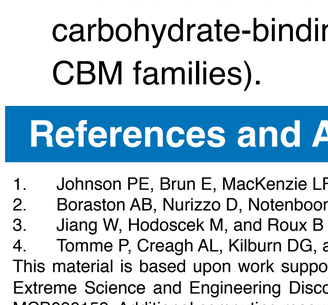
	ΔG_b° [kcal/mol]	ΔG_{repu} [kcal/mol]	ΔG_{disp} [kcal/mol]	ΔG_{elec} [kcal/mol]	ΔG_{rstr} [kcal/mol]
Cellopentaose	-	68.0 ± 0.4	-61.8 ± 0.1	-66.3 ± 0.3	-
*Cf*CBM4-1-RE	-4.5 ± 1.3	73.8 ± 1.1	-78.9 ± 0.2	-59.2 ± 0.5	- 0.3
*Cf*CBM4-1-NRE	-5.9 ± 1.5	74.2 ± 1.2	-78.9 ± 0.3	-61.3 ± 0.6	0.1
*Cf*CBM4-1 Experimental (4)	-5.2 ± 0.9	-	-	-	-

Similarity of *Cf*CBM4-2 to *Cf*CBM4-1

- NMR structure of apo *Cf*CBM4-2 (pink) suggests cleft is wider than *Cf*CBM4-1 (gray).
  - MD simulations illustrate the cleft width quickly tightens around the docked ligand, going from 15.3 Å across to 9 Å – a possible chain acquisition mechanism.
- Dynamic measurements from simulation reveal similar behavior as in *Cf*CBM4-1.
  - Hydrogen bonds per site, RMSF of the ligand per binding site, and total interaction energy of the ligand with each site is equivalent in both *Cf*CBM4-2-RE and *Cf*CBM4-2-NRE.
- *Cf*CBM4-2 is likely also capable of bi-directional oligomer binding.



Bi-directional binding beyond *C. fimi* CBM4s



P. cellulosa xylanase Xyn10C (purple) and CfCBM4-1-RE

- β-sandwich fold is common among CBMs (29 of 69 families) and noted for broad specificity.
  - Two binding sites – one on the face of β-sheets and one on the edge of β-sheets.
- Of deposited structures, 10 families have glycan bound at face of β-sheets (as in 1GU3) – 34 total structures
- Structural alignment with DALI
  - 22 structures with the ligand in *Cf*CBM4-1-RE orientation
  - 12 ligands in the opposite orientation (similar to *Cf*CBM4-1-NRE)

Conclusions

- Simulation supports the hypothesis that *C. fimi* CBM4s are capable of binding cello-oligomers with the pyranose reducing end at either end of the cleft.
- Free energy calculations are remarkably comparable to ITC measurements, suggesting ITC captures an average conformational ensemble of *Cf*CBM-4-1-RE and *Cf*CBM4-1-NRE.
- MD simulations of *Cf*CBM4-2 extend bi-directional binding observations to loosely related (36% sequence similarity) familial representatives.
- Bi-directional binding may not be limited to CBM4s, potentially including many carbohydrate-binding proteins bearing the β-sandwich fold (currently 29 additional CBM families).

References and Acknowledgements

1. Johnson PE, Brun E, MacKenzie LF, Withers SG, McIntosh LP (1999) *J. Mol. Biol.*, **287**, 609-625.
2. Boraston AB, Nurizzo D, Notenboom, V, Ducros, V, Rose DR, Kilburn, DG, and Davies, GJ (2002) *J. Mol. Biol.*, **319**, 1143-1156.
3. Jiang W, Hodoscek M, and Roux B (2009) *J. Chem. Theory Comput.* **5**, 2583-2588.
4. Tomme P, Creagh AL, Kilburn DG, and Haynes CA (1996) *Biochemistry*, **35**, 13885-13894.

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.