

Supplementary Information for:

Interaction of Carbohydrate Binding Module 20 with starch substrates

Son Tung Ngo^{1,2*}, Phuong Duy Tran-Le³, Giap T. Ho³, Loan Q. Le⁴, Le Minh Bui³, Khanh B.

Vu,³ Huong Thi Thu Phung³, Dung H. Nguyen^{3,4}, Sang T. Vo³, Van V. Vu^{3*}

¹Laboratory of Theoretical and Computational Biophysics, Ton Duc Thang University, Ho Chi Minh City, Vietnam;

²Faculty of Applied Sciences, Ton Duc Thang University, Ho Chi Minh City, Vietnam;

³NTT Hi-Tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

⁴ Institute of Tropical Biology, Vietnam Academy of Science and Technology, Ho Chi Minh City, Vietnam

*Email: ngosontung@tdtu.edu.vn; vanvu@ntt.edu.vn

Free energy perturbation method. The binding free energy between amylose and CBM20 protein was calculated using free energy perturbation (FEP) method.¹ The way to estimate the Gibbs free energy difference of binding between protein to substrate was described in the previous study.² In particular, the coupling parameter λ was employed to turn the non-bonded potential from full-interaction state to non-interaction state. CBM20 protein was “destroyed” from two system including CBM20+substrate and isolated CBM20 systems. The difference of works of two processes is Gibbs free energy various of binding,³ which is evaluated utilizing Bennett acceptance ratio method.⁴ The ten values of coupling parameter λ were picked as 0.00, 0.10, 0.25, 0.40, 0.55, 0.65, 0.75, 0.85, 0.95, and 1.00 in order to change electrostatic potential. Consequently, ten values of λ were used to alter van der Waals (vdW) potential that are 0.00, 0.10, 0.20, 0.25, 0.30, 0.40, 0.55, 0.70, 0.85, and 1.00. Every λ -changed simulation was carried out with the length of 2 ns. The binding free energy terms were mentioned in Figure S13. The electrostatics binding free energy was found to be dominate over the vdW free interaction energy in the binding of the CBM20 protein to A3L substrate.

In addition, the dependence of binding affinities according to system size was estimated and shown in Figure S14. The docking energy of the CBM20 protein to substrate reaches equilibrium state when the size of the substrate is larger than three.

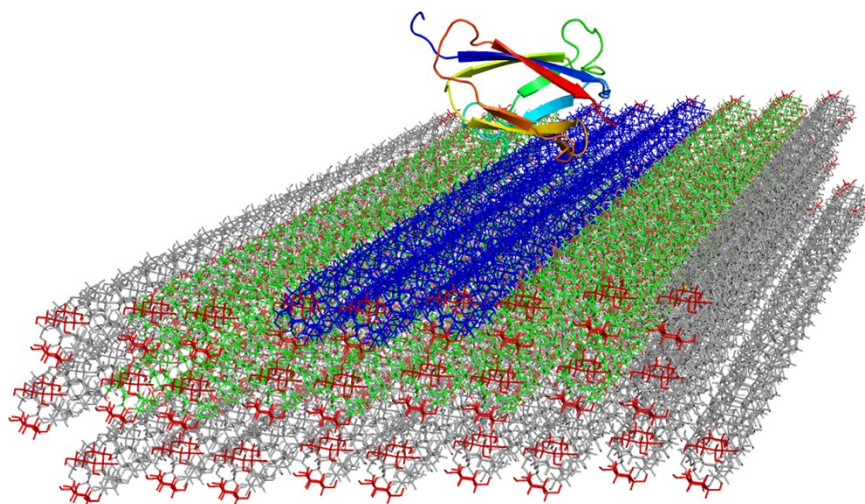


Figure S1. The structure of A3L in the complex with CBM20 at **BdS1** derived with MD simulations. Three amylose double helices forming non-covalent contacts with CBM20 are shown in blue. The RMSD of the combination of CBM20 and this domain is shown in Figure S2. The C1 and C4 atoms of the gray double helices were restrained during MD simulation using a small harmonic force with a value of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Heavy atoms of red residues were also

restrained during MD simulations. The double helices shown in blue and green were not restrained except for the atoms at the two termini (highlighted in red).

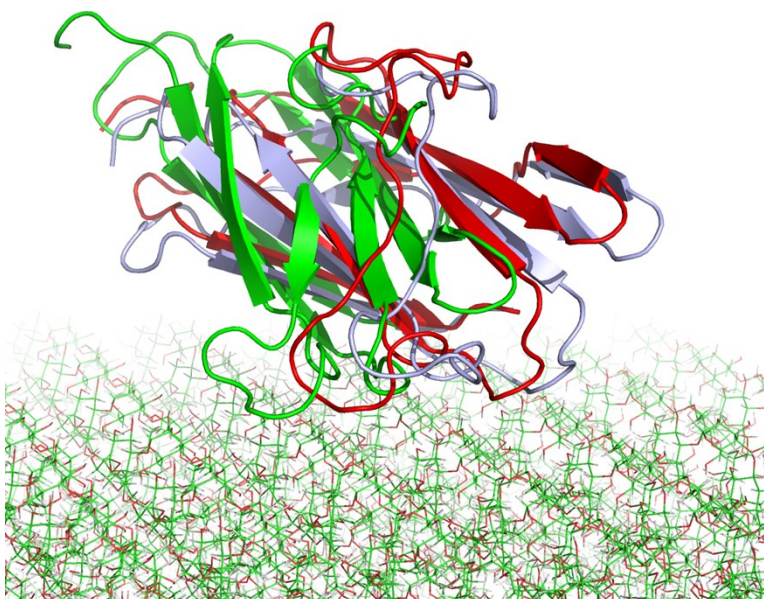


Figure S2. The CBM20-A3L complex at **BdS2** at the beginning (green), 50 ns (red), and 100 ns (gray) of the first MD trajectory.

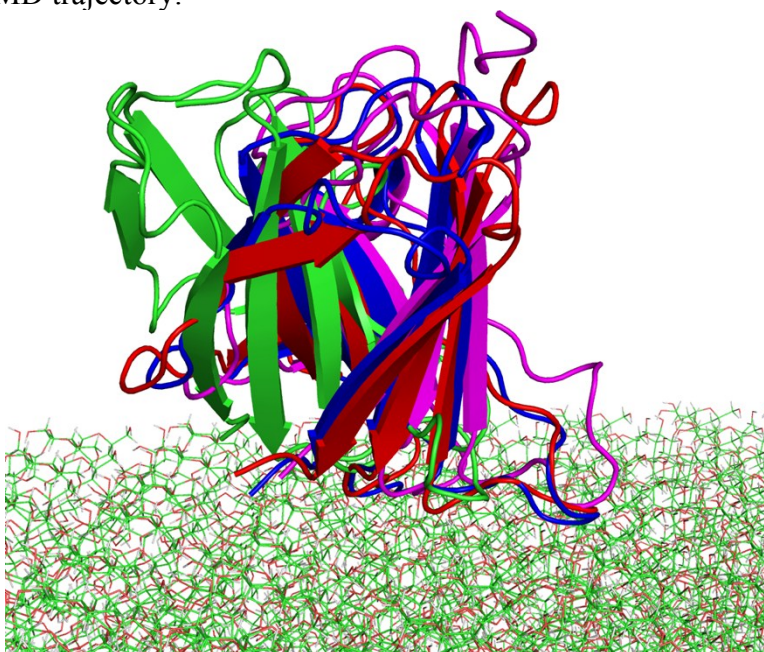


Figure S3. The CBM20-A3L complex at **BdS2** at the beginning (green), 23.5 ns (red), 48 ns (blue), and 70 ns (purple) of the second MD trajectory.

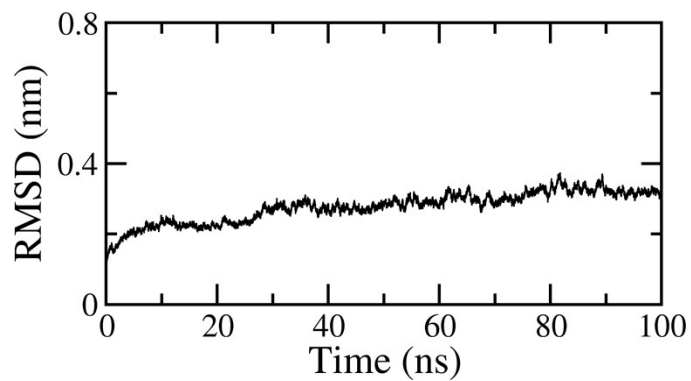


Figure S4. RMSD of CBM20+selected interaction domain of A3L (highlighted in blue in Figure S1) at **BdS1**.

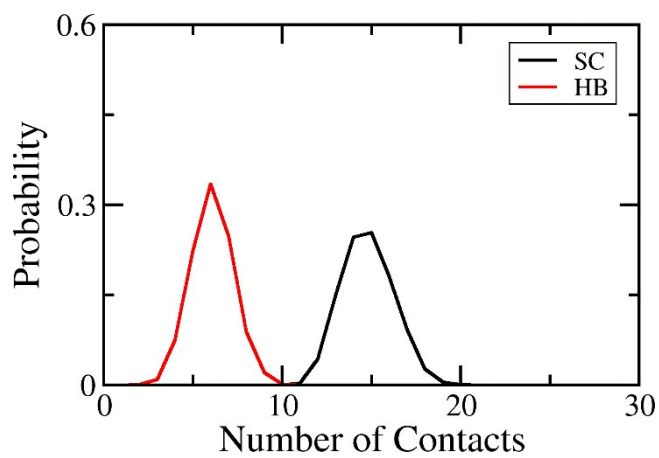


Figure S5. Intermolecular contacts between CBM20 and A3L at **BdS1** during the last 60 ns of MD simulation.

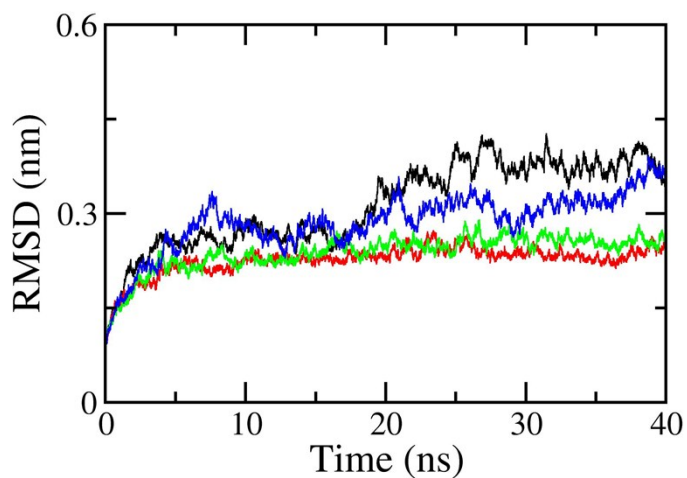


Figure S6. RMSD of CBM20+selected interaction domain of A3L (highlighted in blue in Figure S1) at **BdS1** obtained over 4 independent MD trajectories.

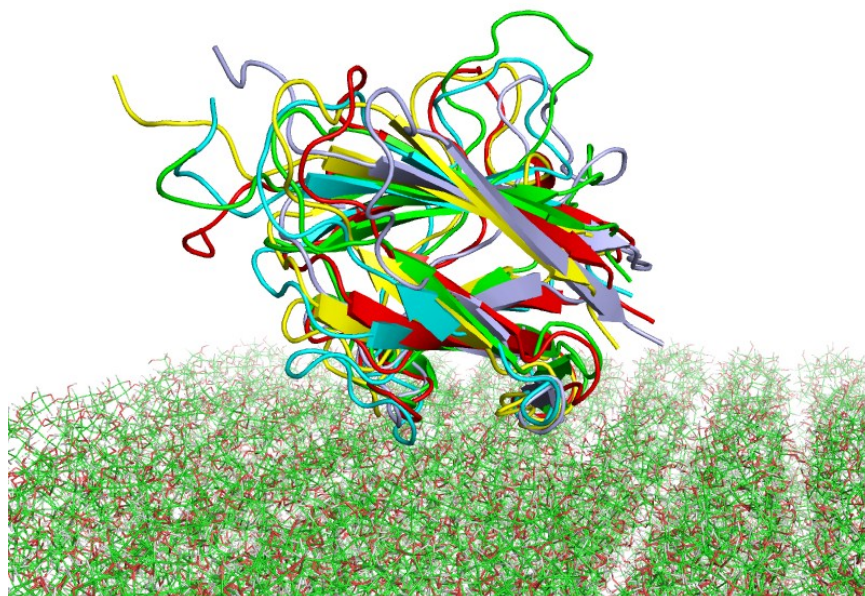


Figure S7. Superposition of CBM20-A3L complex at **BdS1** in different independent MD trajectories. The CBM20 molecule of the stable structure of CBM20-A3L complex shown in Figure 5 is shown here in green. The CBM20 molecules of the optimized structures of CBM20-A3L obtained from 4 other 40-ns-long independent MD simulations are shown in cyan, yellow, red, and gray.

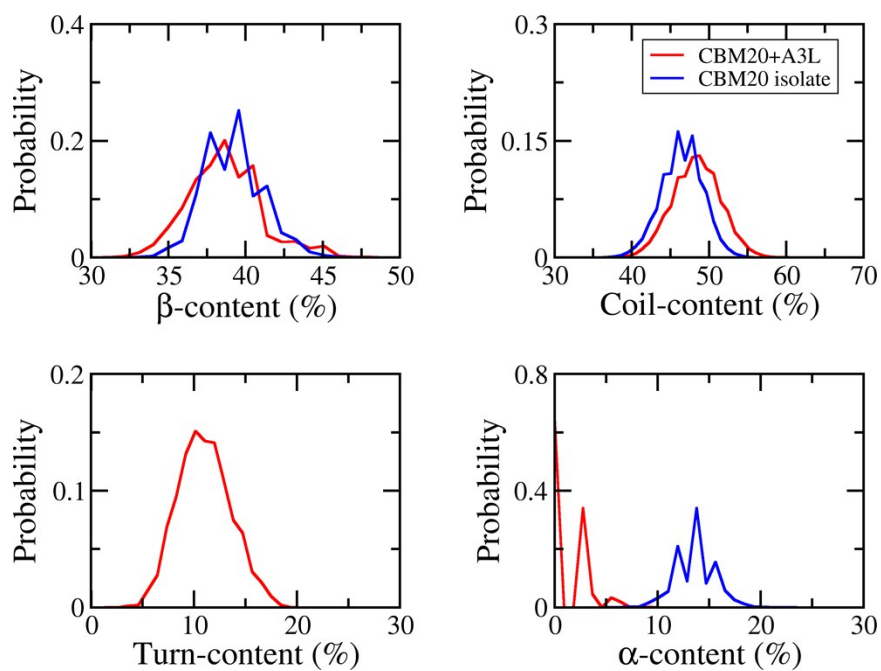


Figure S8. The secondary structure parameters of of CBM20 (blue) and CBM20-A3L at **BdS1** (red).

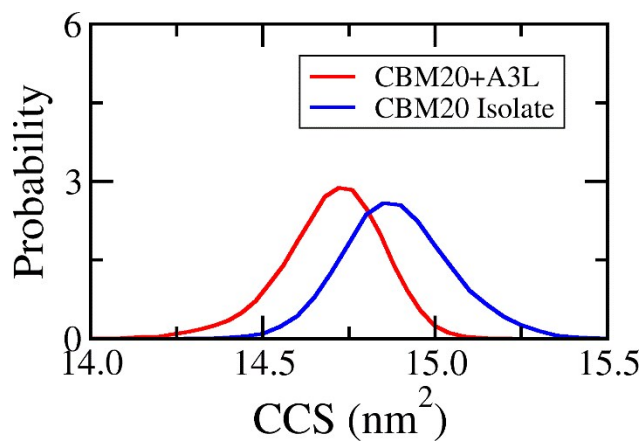


Figure S9. The CCS distribution difference between CBM20 (blue) and CBM20-A3L complex at **BdS1** (red).

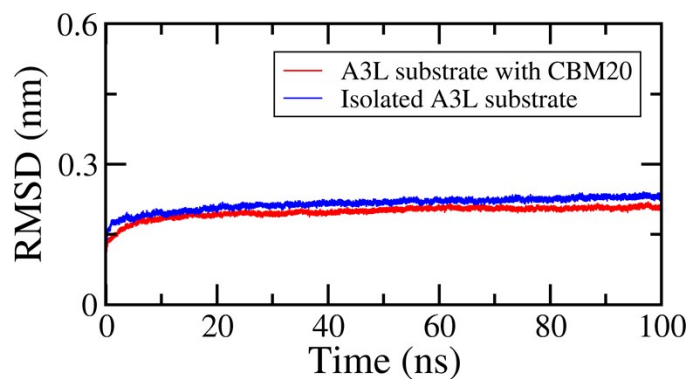


Figure S10. The RMSD of CBM20-binding helices of A3L (belonging to the helices highlighted in blue in Figure S1) during 100 ns of MD simulation.

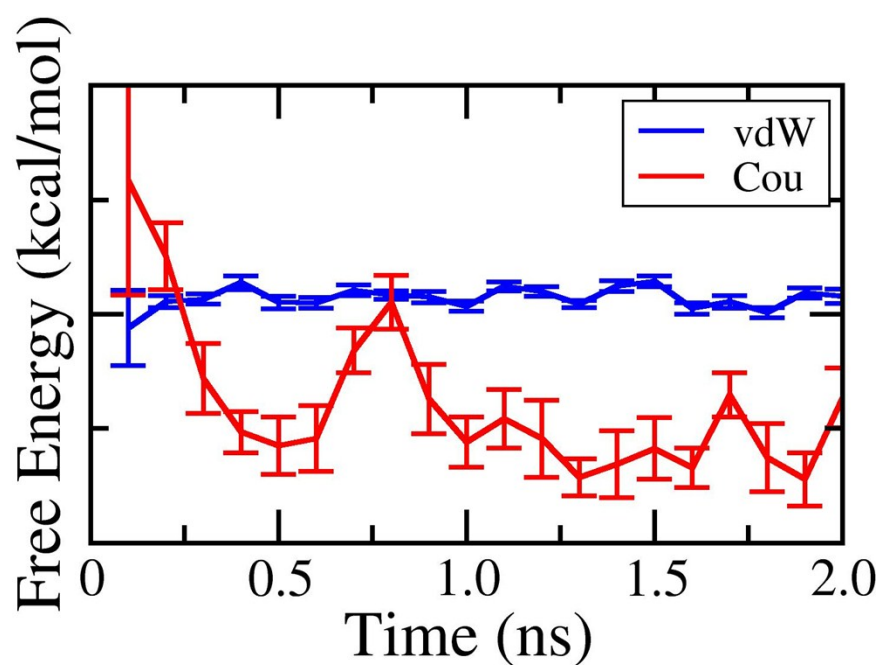


Figure S11. The free energy difference of binding of the CBM20 to A3L system estimated using the Bennett acceptance ratio method with a cycle of 100 ps each.⁴ The values composed in the equilibrium region ($t > 1$ ns) were used to predict the relative binding free energy.

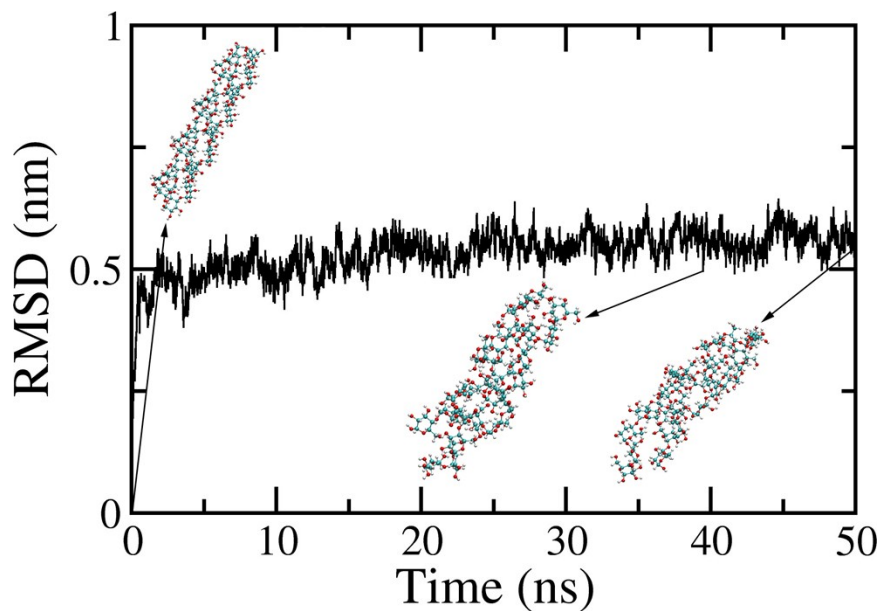


Figure S12. The RMSD of the isolated ADH in solution during 50 ns of MD simulation. The various snapshots taken during the trajectory indicated that the isolate ADH stabilized during the MD simulations.

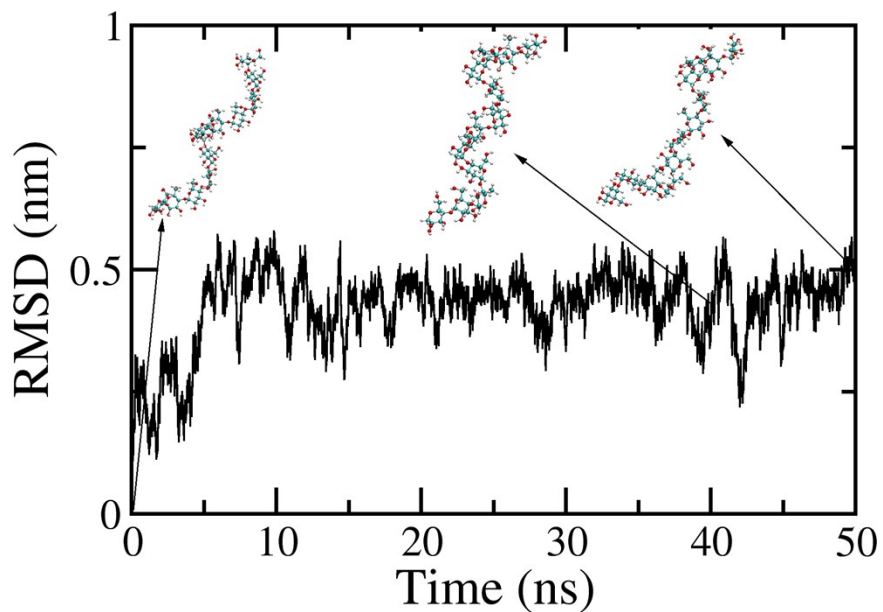


Figure S13. The RMSD of the isolated ASH in solution during 50 ns of MD simulation. The various snapshots taken during the trajectory indicated that the isolate ADH stabilized during the MD simulations.

Reference

1. R. W. Zwanzig, *J. Chem. Phys.*, 1954, **22**, 1420-1426.
2. S. T. Ngo, H. M. Hung, K. N. Tran and M. T. Nguyen, *RSC Adv.*, 2017, **7**, 7346-7357.
3. S. T. Ngo, B. K. Mai, D. M. Hiep and M. S. Li, *Chem. Biol. Dru. Des.*, 2015, **86**, 546-558.
4. C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245-268.