Electronic Supplementary Material (ESI) for RSC Advances.

Supporting Information for "Calculations of the absolute binding free energy for Ralstonia Solanacearum Lectins bound with methyl-α-L-fucoside at molecular mechanical and quantum mechanical/molecular mechanical levels"

Wei Liu,[†] Xiangyu Jia,[*,†,‖] Meiting Wang,[†] Pengfei Li,[†] Xiaohui Wang,[†] Wenxin Hu,[‡] Jun Zheng,[‡] and Ye Mei[*,†,¶,§]

†*State Key Laboratory of Precision Spectroscopy, School of Physics and Materials Science, East China Normal University, Shanghai 200062, China*

‡*The Computer Center, School of Computer Science and Software Engineering, East China Normal University, Shanghai 200062, China*

¶*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

§*Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China*

‖*Current address: NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

E-mail: xj6@nyu.edu; ymei@phy.ecnu.edu.cn

Measure the Overlaps between Adjacent Windows

The free energy difference between adjacent states, of which the potential energy surfaces are denoted as U_0 and U_1 respectively, can be estimated by BAR as

$$\Delta A = \frac{1}{\beta} \ln \frac{\langle f(-U_1 + U_0 + C) \rangle_1}{\langle f(U_1 - U_0 - C) \rangle_0} + C - \frac{1}{\beta} \ln \frac{n_1}{n_0}, \tag{1}$$

where f denotes the Fermi function $1/\left(1 + \exp\left(\beta x\right)\right)$, n_0 and n_1 are the number of configurations sampled on the U_0 and U_1 surfaces respectively, and

$$C = \Delta A + \frac{1}{\beta} \ln \frac{n_1}{n_0}. \tag{2}$$

According to Crooks theorem,[1]

$$\frac{P_0(W)}{P_1(W)} = e^{\beta(W - \Delta A)}, \tag{3}$$

$\langle f(-W + C) \rangle_1$ or $\langle f(W - C) \rangle_0$ can be proved to be related to the overlap between the normalized probability densities $P_0(W)$ and $P_1(W)$ sampled on U_0 and U_1 respectively (with equal number of conformations sampled on two adjacent states)[2]

$$\int_\Omega \frac{P_0 P_1}{P_0 + P_1} dW. \tag{4}$$

Therefore, $\langle f(-W+C) \rangle_1$ and $\langle f(W-C) \rangle_0$ can be employed to measure the overlap between two adjacent states (for instance, for state i and j) in BAR calculations which can be denoted as

$$O_{ij} = \langle f(-W + C) \rangle_{i/j}. \tag{5}$$

Increasing the overlap, O_{ij} approaches 0.5.

The convergence of free energy calculations at the MM level

The intermediate states defined by λ used in the alchemical process are listed in Table S1 . The overlaps of adjacent windows in three decoupling calculations are collected in Table S2.

Table S1: λ values for the decoupling calculations in $b \rightarrow c$ and $e \rightarrow g$ processes.

window	$-\Delta G_{int}^{bulk}$		window	$\Delta G_{int}^{site} - \Delta G_{t+r}^{site}$	
	GAFF	GLYCAM		GAFF	GLYCAM
1	0.0	0.0	1	0.0	0.0
2	0.05	0.05	2	0.05	0.05
3	0.1	0.1	3	0.1	0.1
4	0.2	0.2	4	0.15	0.15
5	0.3	0.3	5	0.2	0.2
6	0.4	0.4	6	0.25	0.25
7	0.5	0.5	7	0.3	0.3
8	0.6	0.6	8	0.35	0.35
9	0.7	0.7	9	0.4	0.4
10	0.8	0.8	10	0.45	0.45
11	0.9	0.9	11	0.5	0.5
12	1.0	1.0	12	0.55	0.55
			13	0.6	0.6
			14	0.65	0.65
			15	0.7	0.7
			16	0.75	0.75
			17	0.8	0.8
			18	0.85	0.85
			19	0.9	0.9
			20	0.95	0.95
			21	1.0	1.0

By using 21 windows for the decoupling calculation in the binding site, the overlaps are well achieved. For the decoupling calculation in bulk water, however, the overlap between $\lambda = 0$ and $\lambda = 0.1$ is quite poor (0.04 for GAFF and 0.05 for GLYCAM), thus we inserted an extra window ($\lambda = 0.05$) in between. The convergence of free energy calculations with BAR method are reflected by the standard errors for each free energy component. As shown in Table 2 in the main text, for both GAFF and GLYCAM, the standard errors of the decoupling free energy in the binding site are 0.03 kcal/mol. While the standard errors for decoupling the ligand molecule in water solution are 0.05 and 0.07 kcal/mol for GAFF and GLYCAM respectively. The standard error values declare that the BAR calculations for the RSL-methyl-α-L-fucoside complex are statistically well converged.

Table S2: The overlaps between adjacent windows in the simulations for the calculations of ΔG_{int}^{bulk} and $\Delta G_{int}^{site} - \Delta G_{t+r}^{site}$ for the RSL-methyl-α-L-fucoside complex.

windows	$-\Delta G_{int}^{bulk}$		windows	$\Delta G_{int}^{site} - \Delta G_{t+r}^{site}$	
	GAFF	GLYCAM		GAFF	GLYCAM
$O_{0,0.05}$	0.16	0.19	$O_{0,0.05}$	0.26	0.28
$O_{0.05,0.1}$	0.33	0.32	$O_{0.05,0.1}$	0.34	0.35
$O_{0.1,0.2}$	0.26	0.26	$O_{0.1,0.15}$	0.40	0.41
$O_{0.2,0.3}$	0.34	0.34	$O_{0.15,0.2}$	0.41	0.43
$O_{0.3,0.4}$	0.35	0.35	$O_{0.2,0.25}$	0.43	0.42
$O_{0.4,0.5}$	0.34	0.34	$O_{0.25,0.3}$	0.42	0.43
$O_{0.5,0.6}$	0.30	0.30	$O_{0.3,0.35}$	0.44	0.44
$O_{0.6,0.7}$	0.24	0.22	$O_{0.35,0.4}$	0.45	0.35
$O_{0.7,0.8}$	0.15	0.13	$O_{0.4,0.45}$	0.41	0.42
$O_{0.8,0.9}$	0.26	0.28	$O_{0.45,0.5}$	0.29	0.37
$O_{0.9,1}$	0.40	0.41	$O_{0.5,0.55}$	0.40	0.37
			$O_{0.55,0.6}$	0.35	0.35
			$O_{0.6,0.65}$	0.34	0.34
			$O_{0.65,0.7}$	0.30	0.44
			$O_{0.7,0.75}$	0.32	0.21
			$O_{0.75,0.8}$	0.39	0.40
			$O_{0.8,0.85}$	0.43	0.43
			$O_{0.85,0.9}$	0.44	0.46
			$O_{0.9,0.95}$	0.48	0.47
			$O_{0.95,1}$	0.46	0.46

Force constants for the restraining potentials

In this work, the magnitudes of force constants for six restraints in GLYCAM and GAFF simulations were chosen in a way that the distributions of the restrained distance restraint and angle/dihedral at state e to be consistent with those distributions at state g (see Fig. 2 in the main text for the definition of states) so as to guarantee sufficient overlaps between them. The force constants utilized are shown in Table S3.

Table S3: The force constants of the restraints for GAFF and GLYCAM force field in the decoupling simulations of RSL/methyl-α-L-fucoside. The distance restraint is in a unit of kcal$\cdot$ mol$^{-1}\cdot$ Å^{-2} and the angle and dihedral restraints are in a unit of kcal$\cdot$ mol$^{-1}\cdot$ rad^{-2}.

FF	k_r	k_θ	k_ϕ	k_α	k_β	k_γ
GAFF	5.0	130.0	30.0	25.0	50.0	110.0
GLYCAM	5.0	100.0	25.0	30.0	45.0	110.0

The RMSDs of the heavy atoms of the protein (black) as well as the ligand (red) in the simulations are shown in Fig. S1, which imply that the simulated system in GAFF and GLYCAM appears to be fairly stable and methyl-α-L-fucoside possesses some flexibility in the pocket.

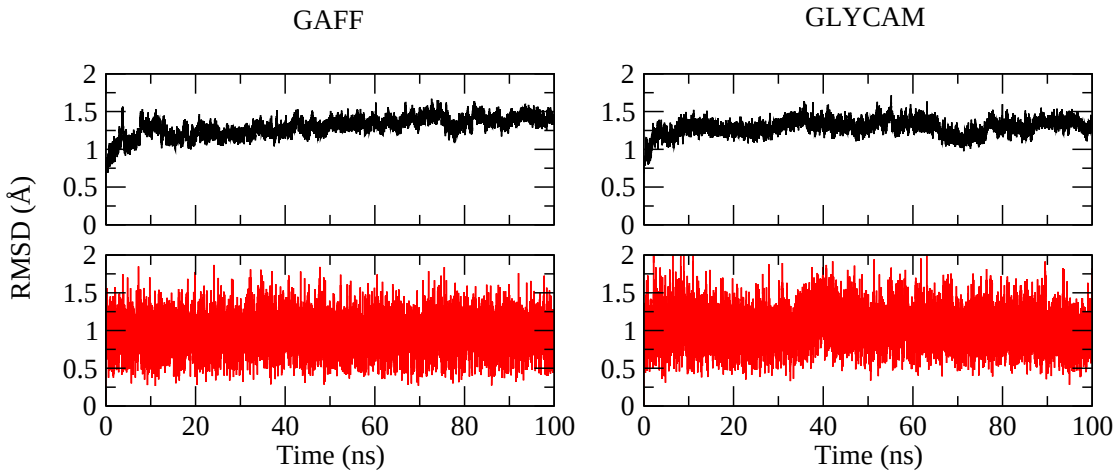

Figure S1: Root mean square deviations of the heavy atoms in the protein (black) and the ligand (red) during the simulations employing GAFF (left) and GLYCAM force field (right).

Shown in Figure S2 are the distributions of the restrained degrees of freedom (namely, r, θ, ϕ, α, β and γ) from the simulations using GAFF (top) and GLYCAM force field (bottom), with their means and variances collected in Table S4. It can be seen that the given magnitudes of the force constants for GAFF and GLYCAM force field at state e gave rise to similar distributions for six restraining potentials comparing with those at state g.

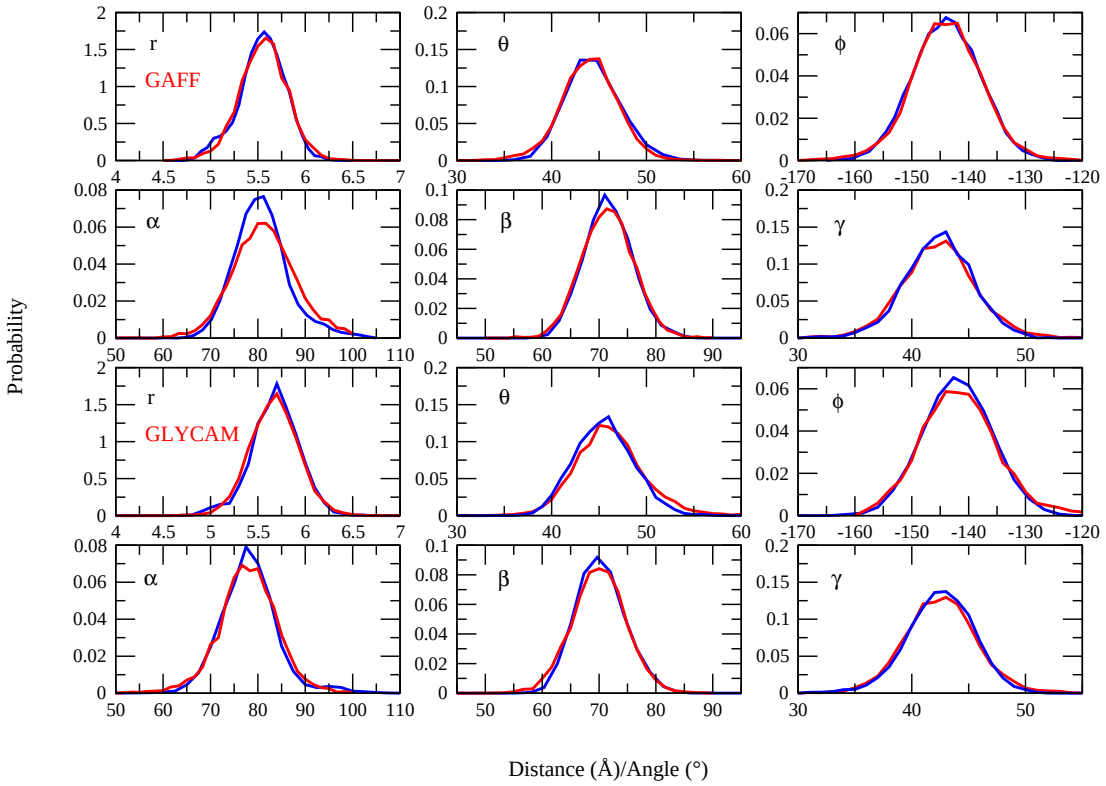

Figure S2: Distributions of the restrained translational and rotational degrees of freedom $(r, \theta, \phi, \alpha, \beta, \gamma)$ at state g (blue) and state e (red) using GAFF (two rows on the top) and GLYCAM force field (two rows on the bottom) respectively.

At state g, no restraints are applied, while at state e the interaction between the protein and the ligand was turned off, but the relative position and the orientation of the ligand in the binding pocket are restrained by six harmonic potentials. At state g, the RMSF in the distance restraint (r) is 0.25 Å in GLYCAM force field, while those of the angles and dihedrals are 3.08° (θ), 5.97° (α), 6.01° (ϕ), 4.25° (β) and 2.95° (γ). While state e stands for a state without any interaction between the protein and the ligand, but concurrently

Table S4: Means and RMSF of the degree-of-freedoms used for the translational and rotational restraints. The units for r are Å, and those for the rest of the DOF are °.

states	mean						RMSF					
	r	θ	ϕ	α	β	γ	r	θ	ϕ	α	β	γ
Coupled, w/t restraint[a]	5.51	43.44	-144.76	79.88	70.71	41.83	0.24	2.82	5.90	5.76	4.17	2.96
Decoupled, w/ restraint[a]	5.51	43.46	-144.85	80.74	70.84	41.90	0.25	3.02	6.27	7.08	4.54	3.32
Coupled, w/t restraint[b]	5.62	45.12	-143.05	77.10	69.05	41.98	0.25	3.08	6.01	5.97	4.25	2.95
Decoupled, w/ restraint[b]	5.62	45.63	-142.62	77.11	69.17	42.05	0.25	3.65	7.31	6.21	4.74	3.23

[a]The trajectory employing GAFF force field.
[b]The trajectory employing GLYCAM force field.

contains full strengths of the six restraining potentials. Improper magnitudes of restraining potentials lead to increased distance between state e and g. By using the force constants of 5 kcal$\cdot$ mol$^{-1}\cdot$ Å^{-2}, 130 kcal$\cdot$ mol$^{-1}\cdot$ rad^{-2}, 30 kcal$\cdot$ mol$^{-1}\cdot$ rad^{-2}, 25 kcal$\cdot$ mol$^{-1}\cdot$ rad^{-2}, 50 kcal$\cdot$ mol$^{-1}\cdot$ rad^{-2} and 110 kcal$\cdot$ mol$^{-1}\cdot$ rad^{-2} for k_1, k_2, k_3, k_4, k_5 and k_6 respectively, the RMSFs of six restraints at state e are remarkably consistent with those at state g. Likewise, six averaged values for Gaussian-like restraining potential distributions at state e and state g match very well, indicating that the overlap between state e and g is good enough to carry out the decoupling of the ligand from the binding pocket. Similar results are presented as well for GAFF.

Long autocorrelation time for the weakly binding intermediates

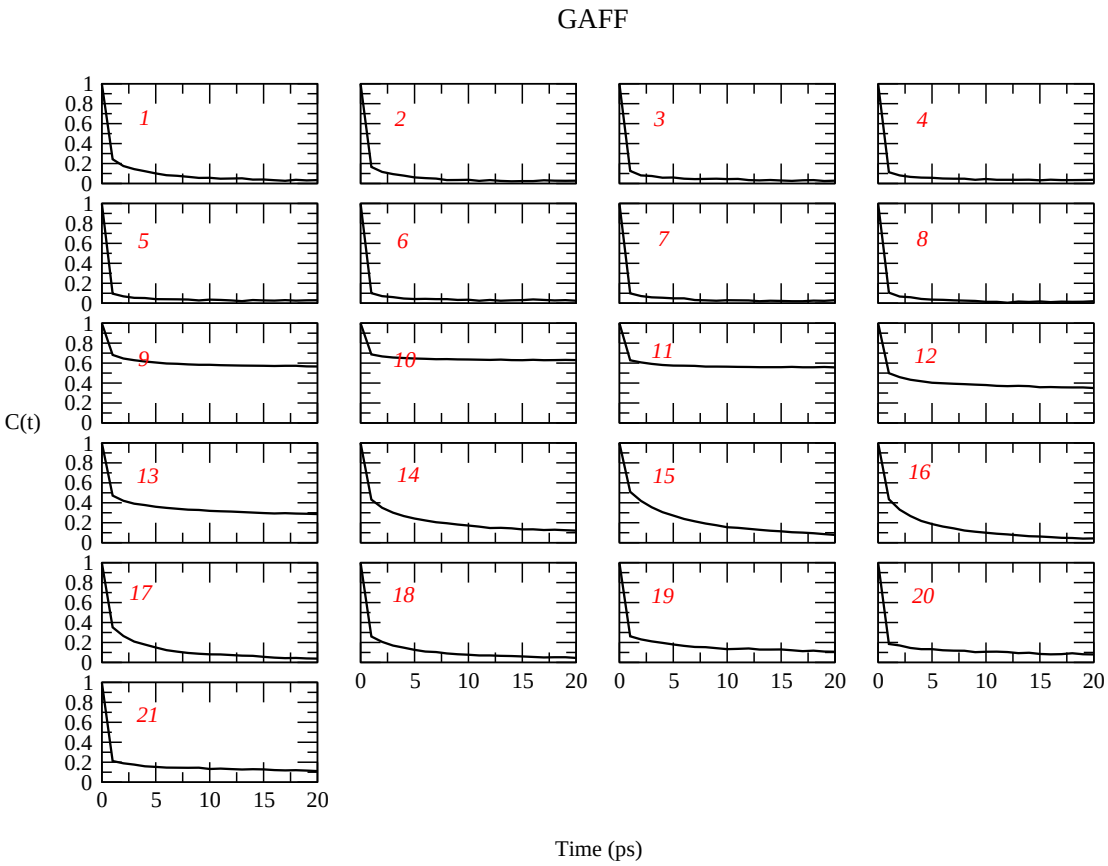

Figure S3: Time dependence of the autocorrelation function for the 21 windows from state e to state g under GAFF.

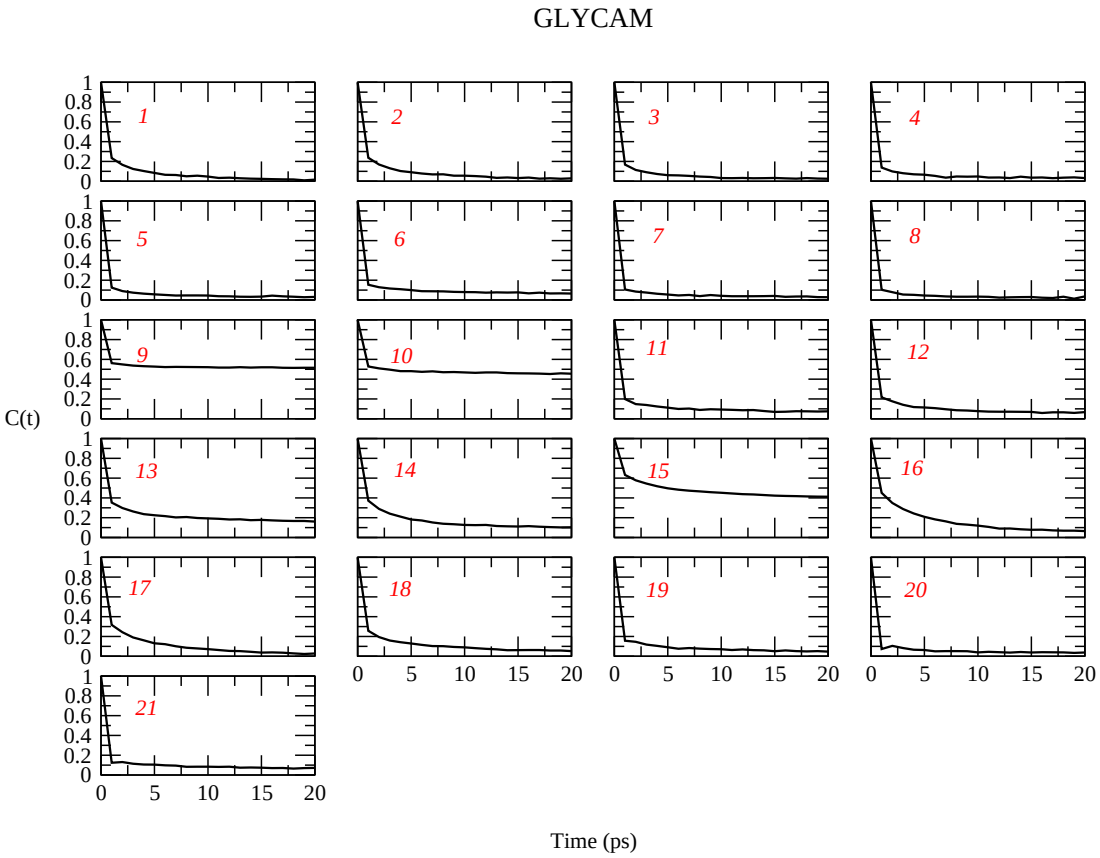

Figure S4: Time dependence of the autocorrelation function for the 21 windows from state e to state g under GLYCAM force field.

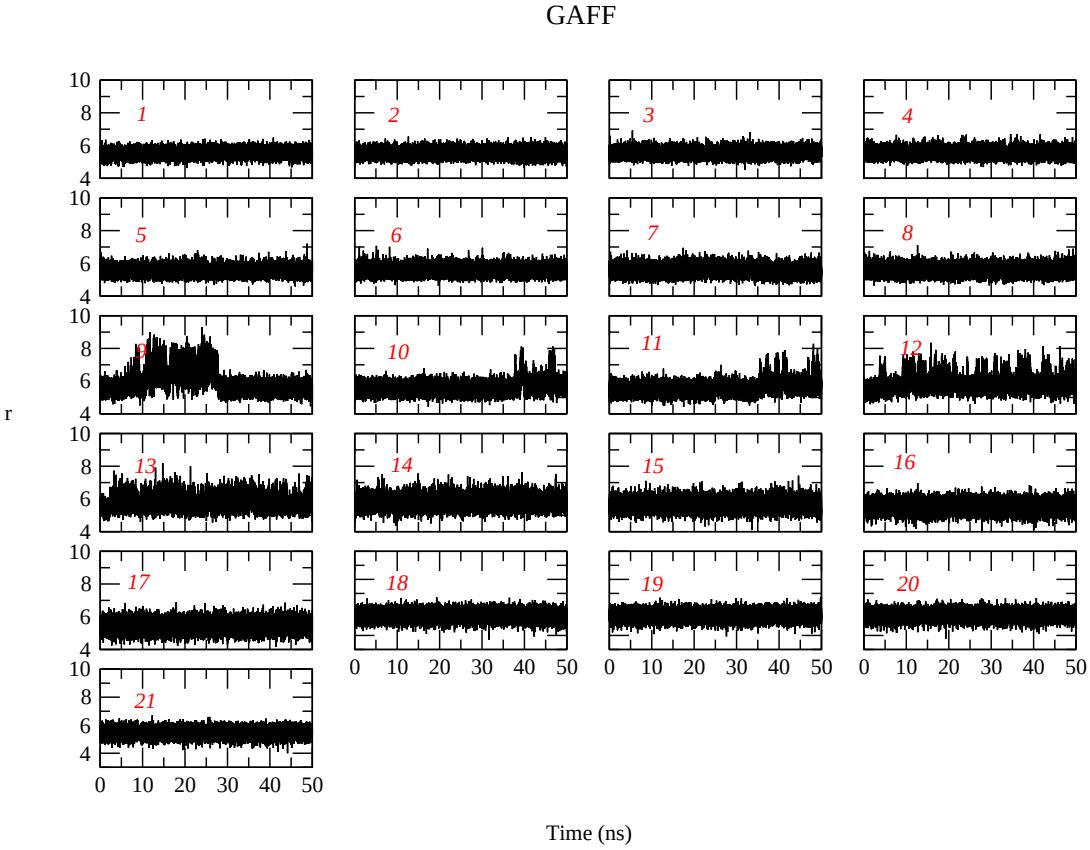

Figure S5: Distributions of distance restraining potential (r) for the 21 windows from state e to state g under GAFF.

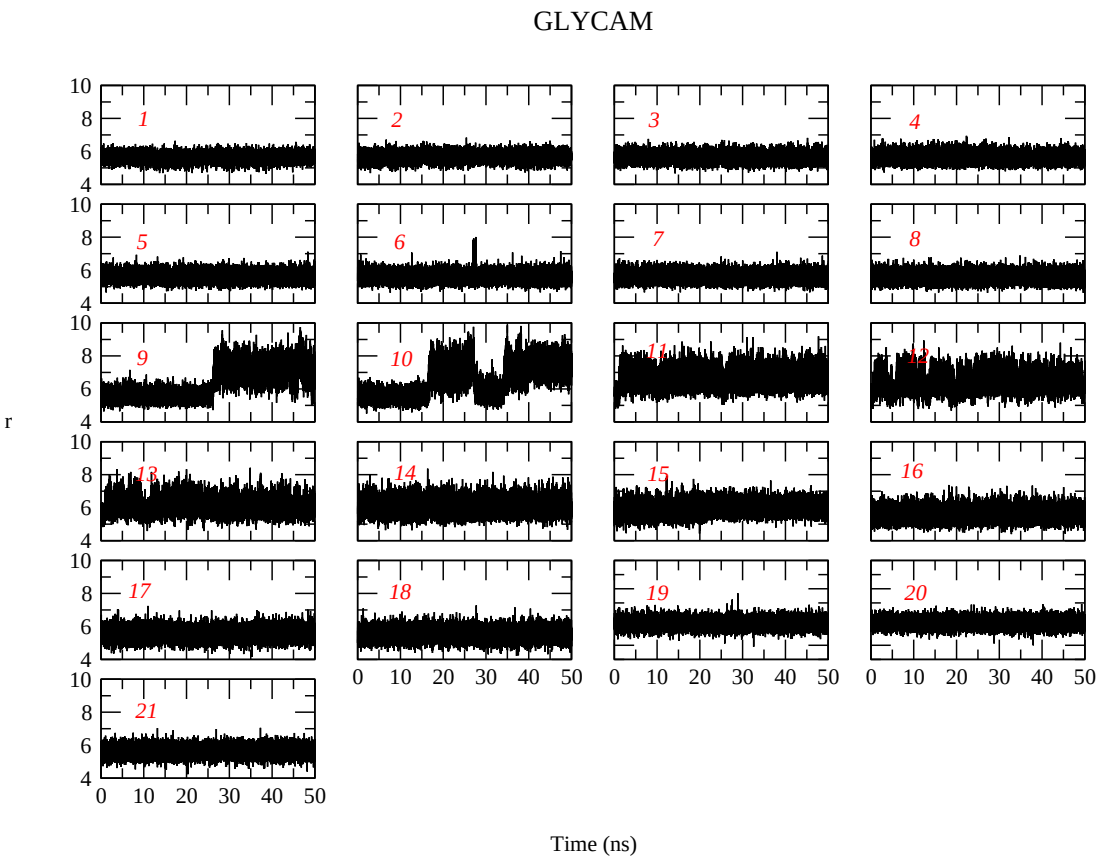

Figure S6: Distributions of distance restraining potential (r) for the 21 windows from state e to state g under GLYCAM force field.

Table S5: Binding free energies calculated from short simulations (10 ns for the calculations of ΔG_{int}^{bulk} and 5 ns for the calculations of ΔG_{int}^{site}-ΔG_{t+r}^{site}). All data are in a unit of kcal/mol. The experimental binding free energy is -8.4±0.3 kcal/mol.

Ligand FF	ΔG_{int}^{bulk}	ΔG_{t+r}^{bulk}	ΔG_{int}^{site}-ΔG_{t+r}^{site}	$\Delta G_{binding}^{MM}$
GAFF	-15.86±0.06	11.84	-39.30±0.10	-11.60±0.12
GLYCAM	-16.66±0.06	11.41	-37.75±0.11	-9.68±0.13

The differences between GAFF and GLYCAM force field

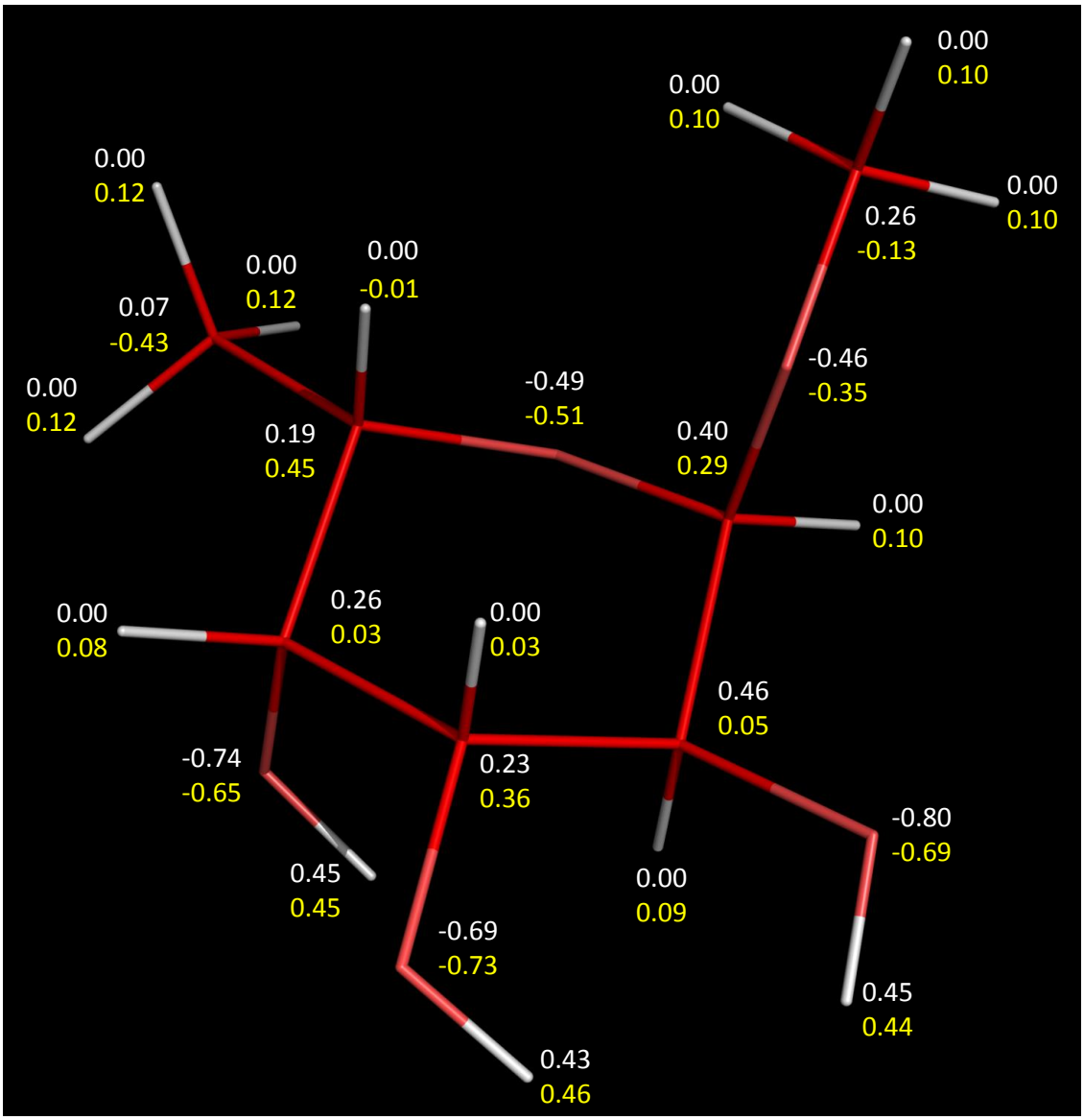

Figure S7: The atomic charge of methyl-α-L-fucoside taken from GLYCAM force field (white) and GAFF force field (yellow).

The convergence of thermodynamic perturbation from the MM level to the QM/MM level

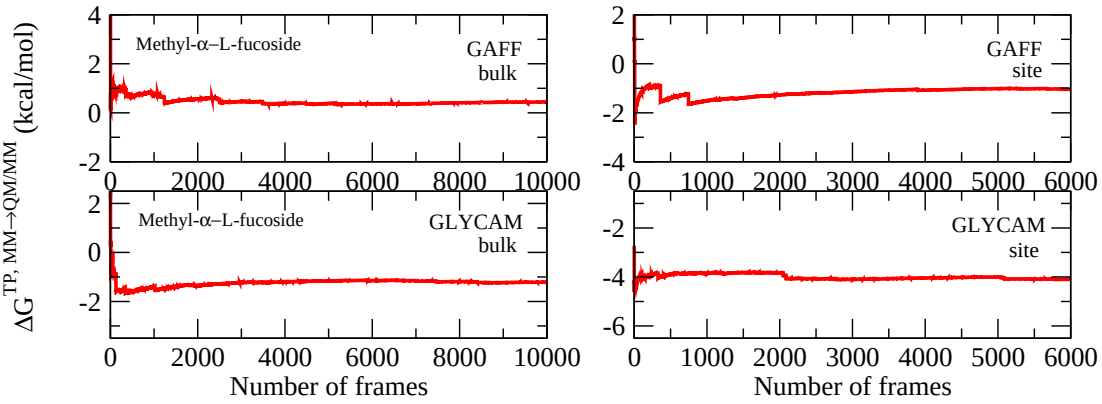

Figure S8: Variations of the TP correction free energies (ΔG^{TP}) from the MM Hamiltonian to the QM/MM Hamiltonian with the number of frames.

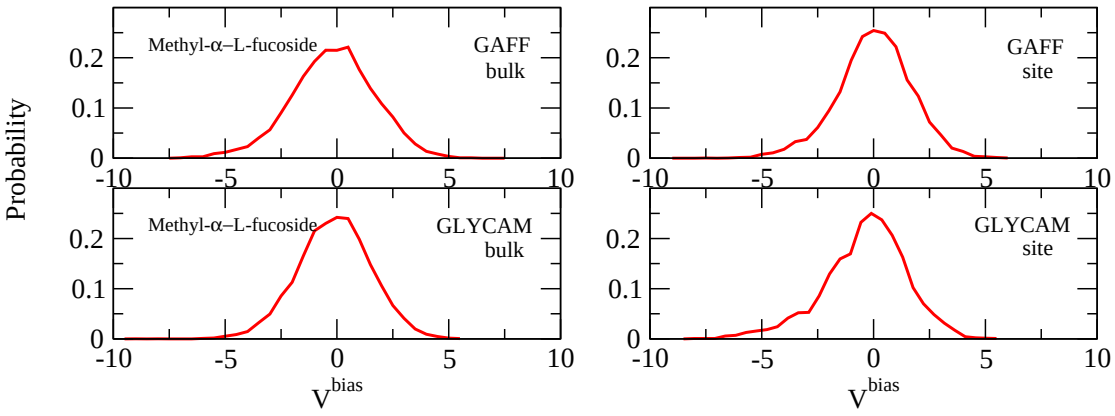

Figure S9: Distributions of the biasing potential between MM and QM/MM Hamiltonians from the trajectories of the methyl-α-L-fucoside binding system employing GAFF (top) and GLYCAM force field (bottom).

References

(1) Crooks, G. E. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *J. Stat. Phys.* **1998**, *90*, 1481–1487.

(2) Hahn, A. M.; Then, H. A Characteristic of Bennetts Acceptance Ratio Method. *Phys. Rev. E* **2009**, *80*, 031111.