


Cite this: *RSC Adv.*, 2017, 7, 38570

Calculations of the absolute binding free energies for *Ralstonia solanacearum* lectins bound with methyl- α -L-fucoside at molecular mechanical and quantum mechanical/molecular mechanical levels†

Wei Liu,^a Xiangyu Jia,^{‡*a} Meiting Wang,^a Pengfei Li,^a Xiaohui Wang,^a Wenxin Hu,^b Jun Zheng^b and Ye Mei^{*,acd}

A method that can reliably predict protein–ligand binding free energies is essential for rational drug design. Much effort has been devoted to this field, but it remains challenging especially for flexible ligands. In this work, both a molecular mechanical (MM) method and a hybrid quantum mechanical/molecular mechanical (QM/MM) method have been applied in the study of the binding affinities of methyl- α -L-fucoside to *Ralstonia solanacearum* lectins. The free energy at the MM level was calculated using the double-decoupling method (DDM) and the free energy change at each step was calculated via a series of intermediate states using the Bennett acceptance ratio (BAR). The binding free energy agrees well with the experimental measurement, no matter whether the general AMBER force field or GLYCAM06j was applied to the ligand. Nonetheless, slow convergence for some intermediate states has been observed, which requires substantially longer simulations than were used in many other studies. The QM/MM free energy was calculated by thermodynamic perturbation (TP) from the MM states. This strategy has been shown to yield minimal variance for the calculated free energy without direct sampling at the QM/MM level in a previous study. However, after this MM-to-QM/MM correction, the agreement with the experimental value decreased. This study serves as an implication of the demand for substantially longer simulations for the alchemical process than those that were used in many other studies and for further improvement of QM/MM methods, especially the description of interactions between the QM and MM regions.

Received 3rd June 2017

Accepted 17th July 2017

DOI: 10.1039/c7ra06215j

rsc.li/rsc-advances

Introduction

Carbohydrates, like other biomolecules such as proteins and nucleic acids, are essential components of life. Emerging evidence suggests a pivotal role for carbohydrate–protein interactions in participating in and regulating a broad range of biologically relevant processes such as cell–cell adhesion, cell differentiation and in-cell signaling.¹ Aside from this, such interactions are associated with a vast array of diseases such as

viral² and bacterial infections,^{3,4} lysosomal storage disorders,⁵ inflammatory processes,⁶ immune system response,⁷ and neurodegenerative disorders.⁸ Carbohydrates are sorted by size into four chemical groups, viz. monosaccharide, disaccharide, oligosaccharide, and polysaccharide. The open-chain and the closed-ring forms often coexist in a monosaccharide and the latter is more common. In cells, it is found that the majority of carbohydrates are attached to proteins or lipids.⁹ A variety of proteins are involved in carbohydrate interactions and recognitions, including enzymes, antibodies, sugar transporters and lectins.¹⁰ The elucidation of the mechanisms that modulate protein recognition in the carbohydrate-combining sites is clearly of immense importance.¹¹ Lectins hold the promise to bind mono- and oligosaccharides reversibly with high specificity.¹⁰ Each lectin protein conventionally possesses two or more carbohydrate-binding sites.¹⁰ Currently, carbohydrate-based drug design presents enormous opportunities for therapeutic development and has become one of the novel frontiers in pharmacology and medicinal chemistry.^{12,13} For instance, the inhibition of the binding of adhesin proteins to host glycans provides possibilities for anti-adhesion therapies.¹⁴ Developing

^aState Key Laboratory of Precision Spectroscopy, School of Physics and Materials Science, East China Normal University, Shanghai 200062, China. E-mail: xj6@nyu.edu.cn; ymei@phy.ecnu.edu.cn

^bThe Computer Center, School of Computer Science and Software Engineering, East China Normal University, Shanghai 200062, China

^cNYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China

^dCollaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7ra06215j

‡ Current address: NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China.



and improving theoretical methods to predict the binding affinity between carbohydrates and carbohydrate-binding proteins is currently a central issue of many studies.^{15,16} Numerous investigations^{17–22} have been focusing on the binding free energy calculations of lectin–carbohydrate complexes, with approaches at different levels of complexity and sophistication.

Alchemical free energy calculation methods are widely employed in the calculations of both relative and absolute free energies. For the calculation of absolute binding free energies, the most widely used method nowadays is the double decoupling method (DDM).^{23–25} The free energies of binding are achieved by switching off the interaction of the ligand with its surroundings progressively. Based on a non-physical thermodynamic cycle, the binding free energy is computed as the sum of multiple intermediate steps during the cycle. However, it is impractical to leverage this method directly as the decoupled ligand can drift away from the binding site and move around in the simulated system. Thus, the endpoint and the intermediate standard states of the ligand cannot be defined correctly, which is detrimental to convergence.^{23–27} To solve this problem, intermediate alchemical states with restraint potentials on the translational and rotational degrees of freedom relative to the protein are introduced. Then, the free energy contribution associated with these restraints can be subtracted analytically.^{23,24} It is evident that both geometrically unbinding methods using the potential of mean force (PMF) and alchemically unbinding methods can be used for the calculations of free energies associated with ligand–protein binding.²³ Generally, geometrically unbinding methods are employed with a superiority for charged ligands.²⁸ Whereas, alchemically unbinding methods are more suitable for ligands deeply buried in the interior of a receptor.²³ Recent reviews on alchemical free energy calculation methods can be found in ref. 26 and 29–33.

Over the past decade, a variety of carbohydrate force fields such as AMBER,³⁴ CHARMM,^{35,36} GLYCAM,^{37,38} GROMOS,^{39,40} OPLS,^{41,42} TRIPOS,⁴³ and MM4 (ref. 44–46) were developed with different functional forms and components, and have been applied to a wealth of studies related to glycans as well as carbohydrate–protein binding complexes.^{47–49} Most of the carbohydrate force fields (CHARMM, GLYCAM, GROMOS, OPLS, *etc.*) are developed to be compatible with simulations in the wider context of biomolecular (proteins, lipids, and nucleic acids) force fields in order to avoid the differences in force field parameterization protocols and/or functional forms between them. Therefore, these carbohydrate force fields adopt similar functional forms as biomolecular force fields including energy components of bonds, angles, torsion, Lennard-Jones interactions, electrostatic interactions and improper torsion.⁵⁰ However, the development of carbohydrate force field parameters is hampered by several difficulties. (1) The structural study of carbohydrates is complicated by virtue of their exceptionally high number of chiral centers.⁵⁰ (2) Subtle electronic effects occur during conformation changes such as the anomeric, the *exo*-anomeric, and the *gauche* effects.^{51,52} These electronic features make the prediction and modeling of the structural and electrostatic landscapes quite challenging.⁵⁰ (3) The monosaccharide units can be connected in a variety of ways.

Therefore, oligosaccharides and polysaccharides have the vast majority of achievable linear and branched conformations. Additionally, the glycosidic linkages and the ring conformations are highly flexible, resulting in an enormous conformational space and a huge number of possible low-energy conformations.⁵⁰ (4) It would be fairly difficult to mimic the situation by classic force fields where the open-chain form and the closed-ring form coexist in the monosaccharide.⁵³ Overall, carbohydrate force fields appear to be reaching a limit in terms of what is possible within its conventional functional form.^{50,53,54} This implies that carbohydrate force fields are in need of improvement in order to predict the ring conformation preferences, describe the electrostatic characteristics, and treat the flexible glycosidic linkages. However, force field refinement is a time-consuming process, and it normally takes years to validate the new parameters.

The hybrid quantum mechanical/molecular mechanical (QM/MM) method has become a sought-after tool, which has been extensively applied in the studies of enzyme catalysis, drug design, and protein engineering.⁵⁵ QM/MM calculations circumvent the need for parameterization and explicitly consider polarization, charge transfer, and many-body effects.^{56,57} QM/MM approaches have been applied to the calculations of protein–ligand interaction frequently by combination with associated free energy estimators.^{58–60} However, in order to calculate ensemble averaged properties such as free energy, uncorrelated samples are necessary. For biological systems in the condensed phase, the energy correlation time is usually on a scale of picoseconds. In order to obtain a reliable ensemble average, the simulation time should not be shorter than one nanosecond, which requires 10^6 propagations with a normal time step on the femtosecond scale. For most cases, direct QM/MM simulations are not feasible for free energy calculations because of the enormous computational expense. Therefore, reference-potential methods were developed,⁶¹ in which energy and probability corrections are applied to a restricted number of structures from a low level simulation, and the free energy difference between the low level and the high level Hamiltonians is obtained by reweighting the snapshots from low level simulations with the energy difference between these two Hamiltonians.⁵⁶ This process can be achieved using various methods, namely, non-Boltzmann thermodynamics perturbation (NBTP),^{62,63} non-Boltzmann Bennett acceptance ratio (NBB),^{61,64} MBAR⁶⁵ and BAR + TP.^{65,66} It has been proved both analytically and numerically that the BAR + TP method provides more precise results than NBB with only half of the expense of QM/MM calculations.⁶⁶ BAR + TP is a special case of MBAR.⁶⁵

In this work, the absolute binding free energy of *Ralstonia solanacearum* lectins (RSL) in complex with methyl- α -L-fucoside was calculated using the double-decoupling method. The structure is shown in Fig. 1. Both the general AMBER force field and GLYCAM were applied to methyl- α -L-fucoside and the results were compared. At each step in the double-decoupling process at MM level, the free energy change was calculated using BAR. Finally, the binding free energy at the QM/MM level was estimated from the MM level *via* thermodynamic



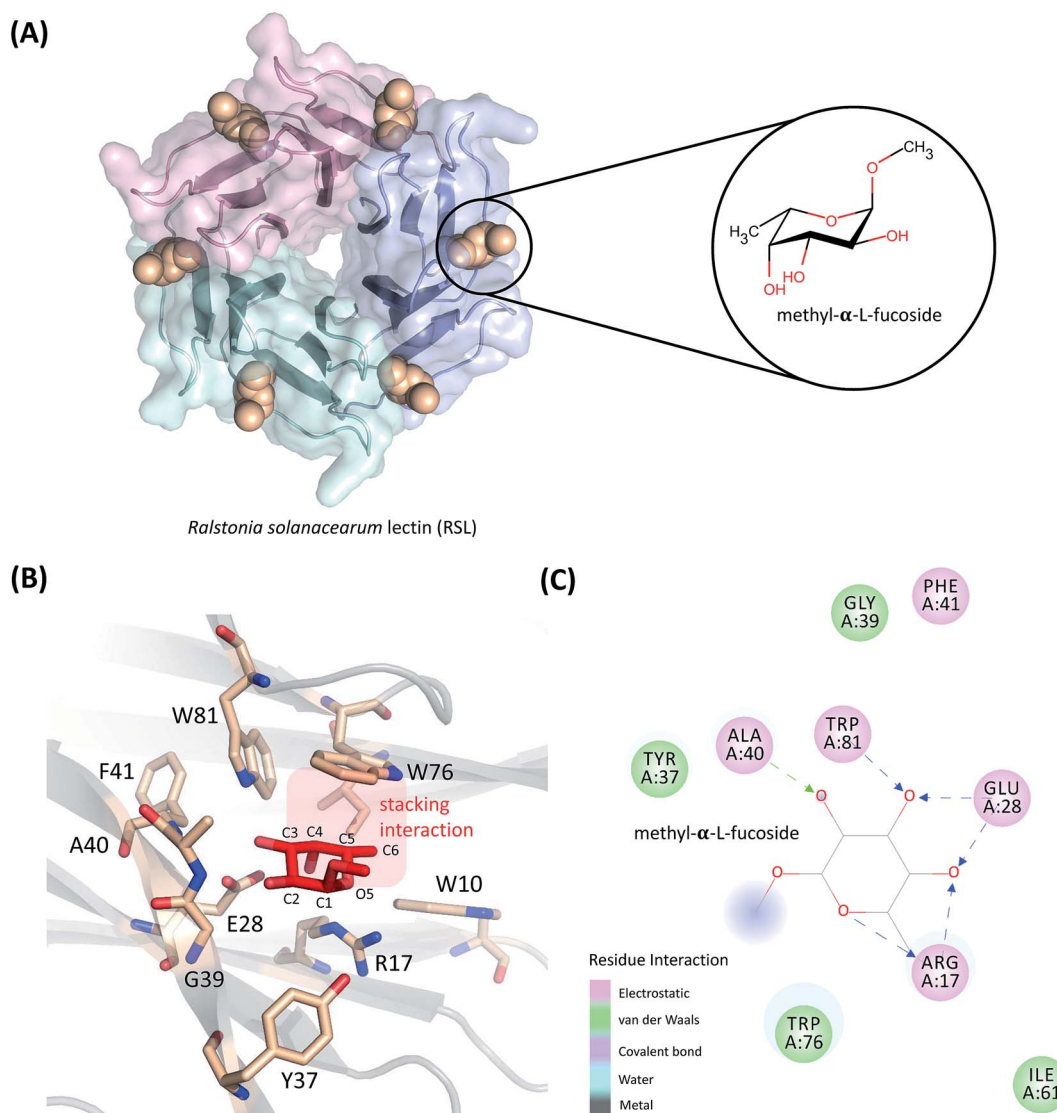


Fig. 1 (A) Structures of RSL bound with methyl- α -L-fucoside. (B) Residues within 5 Å of methyl- α -L-fucoside are highlighted. The aromatic side-chain of Trp76 is parallel to the two apolar faces of methyl- α -L-fucoside, and is stabilized by the CH- π interaction between them. The larger apolar face is defined by the C3, C4, C5, and C6 atoms on the sugar ring, and the smaller apolar face is created by the C1, C2, and O5 atoms. (C) Methyl- α -L-fucoside forms six hydrogen bonds with the receptor.

perturbation, without running formidably expensive QM/MM dynamics.

Method

Free energy calculation

For protein–ligand binding, the equilibrium constant can be expressed as²³

$$K_b = \frac{\int_{\text{site}} d(L) \int d(X) e^{-\beta U}}{\int_{\text{bulk}} d(L) \delta(r_{\text{com}} - r^*) \int d(X) e^{-\beta U}}, \quad (1)$$

where L and X are the coordinates of the ligand and the remaining molecules (protein and solvent), respectively. U is the energy of the whole system. β is the inverse of $k_B T$, k_B is the

Boltzmann constant and T is the temperature. r_{com} is the position of the center of mass (COM) of the ligand and r^* is an arbitrary location in the bulk region. The denominator is the partition function of one state where the COM of the ligand is fixed at r^* , far away from the binding site, and the numerator is the partition function of the other state where the ligand resides in the binding pocket of the protein. Calculating the free energy difference between these two states directly is impractical. One efficient strategy is the DDM and the corresponding non-physical thermodynamic cycle is shown in Fig. 2. In this thermodynamic cycle, the interaction between the ligand and the bulk solvent is turned off in the first step, of which the free energy cost equals to the solvation free energy. Then, extra potentials are added to restrain the translational and rotational degrees of freedom of the ligand in the gas phase with the



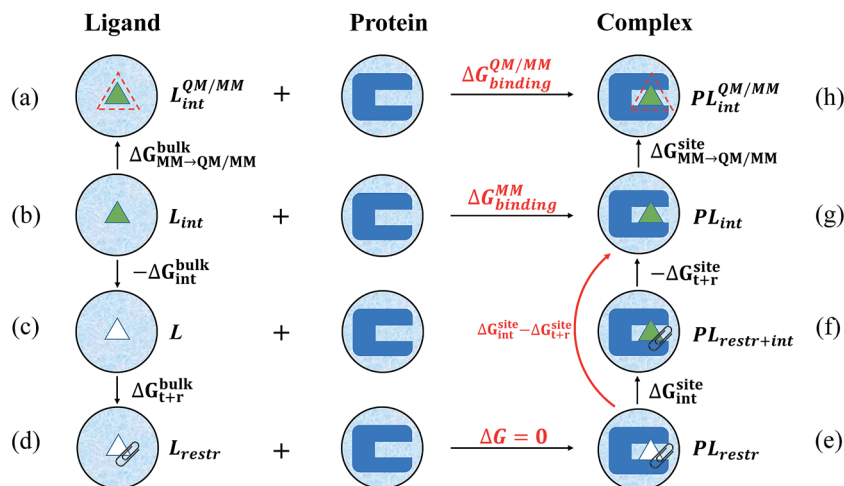


Fig. 2 Alchemical thermodynamic cycle utilized in the calculation of the absolute binding free energies. In the first row are the states under the QM/MM Hamiltonians, with the ligand in either bulk water (a) or in the binding pocket (h). Other states are under the MM Hamiltonians. The MM to QM/MM free energy changes are carried out using one-step thermodynamic perturbation in both bulk water ($\Delta G_{MM \rightarrow QM/MM}^{\text{bulk}}$) and the binding pocket ($\Delta G_{MM \rightarrow QM/MM}^{\text{site}}$). In state c, the ligand is decoupled from the water solvent and the free energy for decoupling ($\Delta G_{\text{int}}^{\text{bulk}}$) is calculated using BAR. In state d, the ligand is clipped in space by adding restraints on six degrees of freedom, and the free energy penalty for applying these restraints ($\Delta G_{\text{t+r}}^{\text{bulk}}$) is calculated analytically. In state e, the ligand is decoupled from its environment in the binding pocket but is clipped in space by restraints of the same strengths. At state g, the ligand is fully coupled to its environment without any restraints applied. The free energy difference ($\Delta G_{\text{int}}^{\text{site}} - \Delta G_{\text{t+r}}^{\text{site}}$) between state e and state g was calculated directly, bypassing state f where the ligand is fully coupled to its environment and the restraints are also applied.

restrained coordinates defined as shown in Fig. 3A. In the binding pocket, the interaction between the ligand and its surrounding environment is turned on gradually. Finally, the added restraints on the ligand are removed in the binding site. Therefore, the absolute binding free energy can be expressed as²³

$$\Delta G_{\text{bind}}^{\circ} = -\frac{1}{\beta} \ln[K_b C^{\circ}] = \Delta G_{\text{int}}^{\text{site}} - \Delta G_{\text{int}}^{\text{bulk}} - \Delta G_{\text{t+r}}^{\text{site}} + \Delta G_{\text{t+r}}^{\text{bulk}}, \quad (2)$$

where $C^{\circ} = 1/1660 \text{ \AA}^3$. When the ligand has a symmetry number (which is denoted as σ) larger than 1, a correction term should be added to fix the error caused by the restraint. The standard concentration C° and symmetry number σ have been included

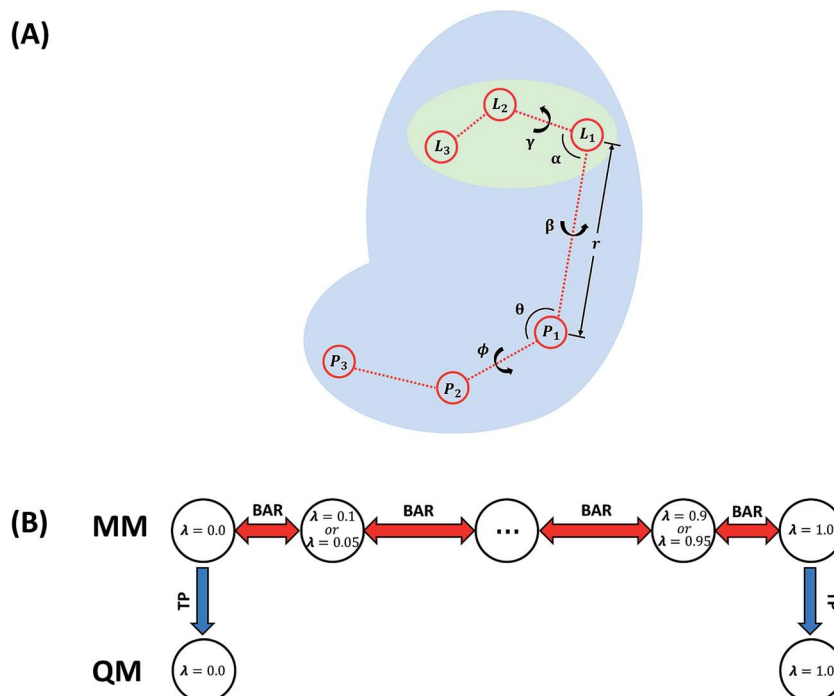


Fig. 3 (A) Restraints on the ligand. L_1 , L_2 , and L_3 are ligand atoms. The atom which is the closest to the center of mass of the ligand was chosen as L_1 . P_1 , P_2 , and P_3 are protein atoms which are picked randomly. (B) Illustration of the employed BAR + TP method.



into the calculation of $\Delta G_{t+r}^{\text{bulk}}$,²⁷ which can be estimated analytically. $\Delta G_{\text{int}}^{\text{site}}$, $\Delta G_{\text{int}}^{\text{bulk}}$ and $\Delta G_{t+r}^{\text{site}}$ can be calculated *via* a series of intermediate states and post-processed by BAR.

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

$$\Delta A = \frac{1}{\beta} \ln \frac{\langle f(-U_1 + U_0 + C) \rangle_1}{\langle f(U_1 - U_0 - C) \rangle_0} + C, \quad (3)$$

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

$$C = \Delta A + \frac{1}{\beta} \ln \frac{n_1}{n_0}. \quad (4)$$

The absolute binding free energy obtained on a MM potential can be corrected to a QM/MM Hamiltonian by the TP method. The correction term can be estimated by the following equation

$$\Delta G^{\text{corr}} = -\frac{1}{\beta} \ln \langle e^{-\beta V^{\text{bias}}} \rangle_{\text{MM}}, \quad (5)$$

where $\langle \dots \rangle_{\text{MM}}$ denotes the average over the conformations harvested under the MM potential. V^{bias} is the biasing potential defined as the energy difference between the QM Hamiltonian and MM Hamiltonian. When the total energy difference is used as the biasing potential, it is difficult to obtain convergent results. Although not theoretically rigorous, using the interaction energy difference as the biasing potential (denoted as $V_{\text{inter}}^{\text{bias}}$) accelerates the convergence of results and the polarization effect in the quantum mechanical treatment of the solute molecule has already been taken into consideration explicitly.⁶⁶ In Fig. 3B, the final corrected absolute binding free energy can be expressed as

$$\Delta G_{\text{binding}}^{\text{QM/MM}} = \Delta G_{\text{binding}}^{\text{MM}} - \Delta G_{\text{MM} \rightarrow \text{QM/MM}}^{\text{bulk}} + \Delta G_{\text{MM} \rightarrow \text{QM/MM}}^{\text{site}}. \quad (6)$$

Molecular dynamics simulations

Simulation setup. The initial structure of RSL complexed with methyl- α -L-fucoside was obtained from the protein data bank (entry 2BT9). The structures of RSL bound systems are shown in Fig. 1. Six methyl- α -L-fucoside molecules bind to three monomers of RSL in this crystal structure, all of which were kept in the simulations. Methyl- α -L-fucoside was modeled with both GAFF and GLYCAM06j force fields.³⁷ The protein was kept as a trimer and was parameterized with the Amber14SB force field. The complex was solvated in a TIP3P⁶⁷ water box with the minimum distance between the solute and the boundary of the box no less than 15 Å. All molecular simulations were carried out with AMBER 14 (ref. 68) suites of the program.

In order to remove close-contact in the system, an energy relaxation with 3000 steps was carried out using a steepest descent algorithm. Then, the entire system was heated up from 0 to 300 K in 100 ps, followed by an equilibration simulation for

Table 1 The length of MD simulations (in a unit of ns) in the DDM calculations. Shown in parentheses are the number of snapshots utilized in the TP calculations

Force field	$\Delta G_{\text{int}}^{\text{bulk}}$	$\Delta G_{\text{int}}^{\text{site}} - \Delta G_{t+r}^{\text{site}}$	$\Delta G_{\text{MM} \rightarrow \text{QM/MM}}^{\text{bulk}}$	$\Delta G_{\text{MM} \rightarrow \text{QM/MM}}^{\text{site}}$
GAFF	50	>50	200 (10 000)	300 (6000)
GLYCAM	50	>50	200 (10 000)	300 (6000)

400 ps. The production run was carried out in an NPT ensemble with isotropic periodic boundary conditions at 300 K. The temperature was regulated using the Langevin dynamics with a collision frequency of 1 ps^{-1} . The integration time step was set to 1.0 fs and SHAKE⁶⁹ algorithm was applied. The particle mesh Ewald^{70,71} method was used to calculate the long-range electrostatic interactions, and the cutoff on non-bonded interactions in real space was set to 12 Å. The trajectories were saved every 10 ps. Varying simulation lengths, which are collected in Table 1, were utilized for each step of the transformation in the thermodynamic cycle determined by the convergence rate.

Double-decoupling free energy simulations. Alchemical double decoupling calculations were performed using the non-physical thermodynamic cycle illustrated in Fig. 2 (see states b to g). For the calculations from state b to c and those from e to f, the van der Waals and the Coulomb interactions of the ligand with its environment were turned off gradually using an alchemical pathway. A total of 21 windows were used in the transformation from state e to state g, in which the ligand is decoupled from the binding pocket in the receptor, and 12 windows were used in connecting state b to state c, where the ligand is decoupled from the bulk solvent (see Table S1† for the specific λ values). For these calculations, $\lambda = 0$ stands for the state where the ligand has full interaction with its surroundings, while $\lambda = 1$ represents the state with the ligand decoupled completely from its surroundings. In order to reduce the statistical error, state e was directly converted to state g, bypassing state f. The translational and rotational restraints (shown in Fig. 3A) were added using harmonic potentials. This translational restraint potential has the following form

$$U_t(r, \theta, \phi) = \frac{1}{2}k_1(r - r_0)^2 + \frac{1}{2}k_2(\theta - \theta_0)^2 + \frac{1}{2}k_3(\phi - \phi_0)^2, \quad (7)$$

where k_1 , k_2 and k_3 are the associated force constants. For rotational restraints, the restraint potential is defined as

$$U_r(\alpha, \beta, \gamma) = \frac{1}{2}k_4(\alpha - \alpha_0)^2 + \frac{1}{2}k_5(\beta - \beta_0)^2 + \frac{1}{2}k_6(\gamma - \gamma_0)^2, \quad (8)$$

The free energy changes in the processes (b to c, e to g) were estimated using the BAR method. The free energy in terms of the change of state c to d was estimated by the same formula used in ref. 23 analytically.

Quantum mechanics/molecular mechanics calculations. In the QM/MM calculations, the ligand was treated by QM methods, and the rest were studied with MM force fields. The differences in the interaction energies, instead of the total



energies,⁶⁶ between the QM/MM and the MM Hamiltonians were used as the biasing potential (V^{bias}), which is defined as

$$V^{\text{bias}} = \Delta E_{\text{ele}}^{\text{MM}} - (\Delta E_{\text{ele}}^{\text{QM/MM}} + \Delta E_{\text{wfd}}^{\text{QM/MM}}), \quad (9)$$

where $\Delta E_{\text{ele}}^{\text{MM}}$ and $\Delta E_{\text{ele}}^{\text{QM/MM}}$ are the electrostatic energies of the ligand interacting with its surroundings at the MM and QM/MM levels respectively, and $\Delta E_{\text{wfd}}^{\text{QM/MM}}$ stands for the wave function distortion energy. This approximation is based on the assumption that the MM Hamiltonian can well reproduce the internal energy of the molecule under the QM Hamiltonian. In order to calculate the interaction energy at the QM/MM level, single-point energy calculations were performed twice, *i.e.*, under vacuum and in the embedding potential. All the QM/MM single-point energy calculations were performed at the BLYP/6-31G* level^{72,73} using Gaussian 09,⁷⁴ which showed competitive accuracy against some more sophisticated functionals in a recent study, even though a part of the accuracy could have resulted from a cancellation of errors.⁷⁵ The simulation lengths of methyl- α -L-fucoside for the TP calculations in the bulk solution and in the binding site were 200 ns and 300 ns, respectively (see Table 1). The quantum mechanical calculations were performed for states b and g, which are the endpoints of the MM simulations. The illustration of the alchemical double decoupling method in conjunction with the TP method is presented in Fig. 3B. During the QM/MM calculations, the periodic boundary condition was not applied and the molecules in the central unit cell were considered. The ligand was placed at the center of the box and was described at the QM level. The protein and water molecules were described at the MM level. For the QM/MM corrections, 10 000 frames were used in bulk water and 6000 frames were used in the binding pocket.

Results and discussion

Force constants of the applied restraints

In this work, the strengths of the six restraints in GLYCAM and GAFF simulations were carefully chosen in order to make the distributions of the restrained distance and angles/dihedrals at state e consistent with the distributions of the degrees of freedom at state g (see Fig. 2 for the definition of states). This is critical to guarantee sufficient overlaps between these two states. The force constants utilized are listed in Table S3.† Because of the difference in the parameters of GAFF and GLYCAM, different force constants were used for the restraint potential. Fig. S1† shows the RMSD fluctuations of the heavy atoms of the protein (solid black line) as well as the ligand (solid red line) during the simulations, implying that the simulated system under both GAFF and GLYCAM force fields was fairly stable and methyl- α -L-fucoside possessed some flexibility in the pocket. The distributions of the six coordinates associated with the translational and rotational degrees of freedom (DOF) (namely, r , θ , ϕ , α , β and γ) in the simulations are shown in Fig. S2.† Their means and root mean square fluctuations (RMSFs) are collected in Table S4.† Employing GAFF for the ligand, in the decoupled but restrained state e, the means of the six DOF can well reproduce those in the unrestrained but

coupled state g. The RMSFs show small deviations, which are acceptable. The largest deviation is seen for α (1.32°). Similar results can be observed in the simulations using GLYCAM, with the largest deviation seen for ϕ (1.30°). Sufficient overlap in these distributions is critical for the free energy calculation between state e and state g.

Long autocorrelation time for the weak binding intermediates

The decay of auto-correlation functions over time for several intermediate states in the binding site is quite slow under both GAFF and GLYCAM force fields. The plots of time correlation functions from the simulations employing GAFF and GLYCAM force fields are presented in Fig. S3 and S4† respectively. Correlation may reduce the effective number of the samples, because when two samples are not independent of each other, the second sample can only be taken as a fractional one. Therefore, the number of uncorrelated samples is smaller than the actual number of samples, which may increase the variance of the calculated means. The number of uncorrelated samples is defined as N/g , where g is defined in eqn (20) in ref. 76. The number of uncorrelated samples for windows 9 ($\lambda = 0.4$), 10 ($\lambda = 0.45$), 11 ($\lambda = 0.5$), and 12 ($\lambda = 0.55$) was 9, 8, 8, and 91 when the ligand was parameterized by GAFF. While employing the GLYCAM force field, the number of uncorrelated samples for windows 9 ($\lambda = 0.4$), 10 ($\lambda = 0.45$), and 15 ($\lambda = 0.7$) was 6, 13, and 8. Such small numbers of uncorrelated samples can be ascribed to the multiple binding modes of the ligand in the pocket when the interaction between the ligand and the receptor is weak. Evidence can be partly found in the displacement of the restrained distance (r) as shown in Fig. S5.† When $\lambda = 0.4, 0.45, 0.5$, and 0.55 , the restrained distance has remarkable fluctuations implying that when its interaction with the protein is weak, the ligand explores the phase space and searches for more favorable binding patterns within the pocket in order to lower the free energy. However, the seeking process requires a considerably long time to assess various binding positions and orientations, leading to a long correlation time (even reaching a time scale of ns), especially when the ligand is trapped in several favorable regions in the pocket. The GLYCAM force field has similar behavior as shown in Fig. S6.† It can also be observed that these long-time-correlation trajectories appear constantly for λ around 0.4 and 0.7, which is caused by the implementation of soft-core potentials and has been reported elsewhere.²⁷ The restrained angles and dihedrals presented similar trends for these intermediates (data not shown). The MD simulation time of these windows was lengthened to over 70 ns. Before calculating the free energies of $\Delta G_{\text{int}}^{\text{site}} - \Delta G_{\text{r+r}}^{\text{site}}$, we discarded the first 30 ns for $\lambda = 0.4$ and 35 ns for $\lambda = 0.5$ in the trajectories employing GAFF, and discarded the first 20 ns for $\lambda = 0.4$ and 25 ns for $\lambda = 0.45$ in the trajectories utilizing the GLYCAM force field, which were considered as the time for the ligand to search for more favorable binding modes.

Alchemical binding free energies

Because long time simulations were conducted and the ensemble means were calculated over large samples, the



standard errors of $\Delta G_{\text{int}}^{\text{bulk}}$ and $\Delta G_{\text{int}}^{\text{site}} - \Delta G_{\text{t+r}}^{\text{site}}$ are very small with the largest one being $0.07 \text{ kcal mol}^{-1}$. With these negligible uncertainties, all the free energy components are unambiguously determined. $\Delta G_{\text{int}}^{\text{bulk}}$ is the energy penalty in turning off the interaction between the solute and the solvent in aqueous solution, which is the solvation free energy of the solute molecule. Employing GAFF, the solvation free energy of methyl- α -L-fucoside is $-15.92 \text{ kcal mol}^{-1}$. While employing GLYCAM, the solvation free energy of this ligand molecule is slightly larger (more negative), which is $-16.52 \text{ kcal mol}^{-1}$. The difference is only $0.6 \text{ kcal mol}^{-1}$ and is within 5 percent of the total solvation free energy. For homogeneous environments such as water and vacuum, $\Delta G_{\text{t+r}}^{\text{bulk}}$ can be calculated analytically.²³ Because restraints with different strengths were utilized, there is a small difference ($0.4 \text{ kcal mol}^{-1}$) in $\Delta G_{\text{t+r}}^{\text{bulk}}$ under the two force fields, which are $11.84 \text{ kcal mol}^{-1}$ for GAFF and $11.41 \text{ kcal mol}^{-1}$ for GLYCAM. These two force fields yielded very close results for $\Delta G_{\text{int}}^{\text{site}} - \Delta G_{\text{t+r}}^{\text{site}}$, which are $-37.32 \text{ kcal mol}^{-1}$ and $-37.27 \text{ kcal mol}^{-1}$ for GAFF and GLYCAM respectively. The difference is only $0.05 \text{ kcal mol}^{-1}$. With these free energy components and the thermodynamic cycle in Fig. 2, the binding free energy at the molecular mechanical level can be calculated, and is $-9.56 \text{ kcal mol}^{-1}$ and $-9.34 \text{ kcal mol}^{-1}$. These calculated binding affinities are slightly stronger than the experimentally measured binding affinity ($-8.4 \pm 0.3 \text{ kcal mol}^{-1}$ (ref. 77)). This difference may have originated from the many approximations adopted during the calculations, one of which is the representation of methyl- α -L-fucoside.

The free energies (see Table S5†) calculated from short simulations may deviate from those from the long simulations. When only 5 ns trajectories were used in the calculations of $\Delta G_{\text{int}}^{\text{bulk}}$, the results differ by nearly 2 kcal mol^{-1} for GAFF from that based on the 50 ns trajectories. But for the GLYCAM force field, these 5 ns trajectories already provided converged results for $\Delta G_{\text{int}}^{\text{bulk}}$. The different behaviors of the GAFF and GLYCAM force fields may come from the different ruggedness of their corresponding potential energy surfaces (PES) when the ligand is in the binding pocket. This can be attributed to the difference between the GAFF and GLYCAM force fields, especially the difference in atomic charges. The partial charges of the GLYCAM06 force field are calculated by restrained fitting to the electrostatic potential (ESP) from QM calculations but are scaled to reproduce crystal unit-cell geometries. Unique charge sets are also provided for α -anomers. Besides, in order to reproduce molecular ESPs properly, GLYCAM06 adopts ensemble-averaged partial charge sets which are specific for each anomer. Moreover, restraints are employed to ensure that the charges on aliphatic hydrogen atoms are zero.¹⁵ The van der

Waals parameters in GLYCAM06 are taken from AMBER PARM94. The parameters for torsion terms are obtained by fitting to QM rotational energy curves.⁵³ As shown in Fig. S7,† the atomic charges from the GAFF and GLYCAM force fields have evident differences. The methyl hydrogen atoms have no partial charge in the GLYCAM force field, which may smooth the PES. These differences have a great influence on sampling, which further leads to the differences in the calculated free energies. GLYCAM serves as a specialized force field for carbohydrates, showing its superiority while conducting the simulations involving carbohydrates. GAFF is a general force field for most organic molecules but is not specialized for carbohydrates. For the solvation free energy, however, it is relatively easier to reach convergence. Thus, long-time simulations and short-time simulations have only negligible discrepancies. For instance, under the GLYCAM force field, the solvation free energy was $-16.66 \text{ kcal mol}^{-1}$ in a 10 ns simulation. A more lengthy simulation (50 ns) provided a slightly smaller solubility ($-16.52 \text{ kcal mol}^{-1}$). Such a small difference suggests that the calculation of the solvation free energy of methyl- α -L-fucoside has reached convergence. Similarly, GAFF achieves convergence quickly with the evidence that the solvation free energy has only a slight change ($0.06 \text{ kcal mol}^{-1}$) while increasing the MD simulation time from 10 to 50 ns.

In our QM/MM calculations, methyl- α -L-fucoside is described using quantum mechanics at the BLYP/6-31G(d) level, in which the discrete point charges are replaced by continuous electron densities and its Coulomb interaction with the environment is calculated at the QM level. However, in order to avoid expensive QM/MM molecular dynamics simulations, the QM/MM free energies were calculated using thermodynamic perturbations from the MM Hamiltonian to the QM/MM Hamiltonian. The results are listed in Table 2. The corrected free energies from GAFF to QM/MM are smaller than those from GLYCAM to QM/MM, which may be because atomic charges in GAFF are fitted from electrostatic potentials at the QM level and are more compatible with QM methods. For the solvation free energy in bulk water, the correction from GAFF to QM/MM ($\Delta G_{\text{MM} \rightarrow \text{QM/MM}}^{\text{bulk}}$) is only $0.44 \text{ kcal mol}^{-1}$, while that from GLYCAM to QM/MM is $-1.21 \text{ kcal mol}^{-1}$. After the MM to QM/MM correction, the solvation free energies at the QM/MM level are $-15.48 \text{ kcal mol}^{-1}$ and $-17.73 \text{ kcal mol}^{-1}$. Although both of these solvation free energies are at the same QM/MM level, this difference is not unexpected. The van der Waals interactions between the QM and MM regions in the QM/MM Hamiltonian were borrowed from the force fields in this work. Besides, the biasing potential was not rigorously calculated. Instead, the biasing potential was calculated as the difference in the

Table 2 The binding free energies of RSL when in a complex with methyl- α -L-fucoside at the GAFF and GLYCAM force fields. All data are in units of kcal mol^{-1} . The experimental binding free energy is $-8.4 \pm 0.3 \text{ kcal mol}^{-1}$

Ligand FF	$\Delta G_{\text{int}}^{\text{bulk}}$	$\Delta G_{\text{t+r}}^{\text{bulk}}$	$\Delta G_{\text{int}}^{\text{site}} - \Delta G_{\text{t+r}}^{\text{site}}$	$\Delta G_{\text{binding}}^{\text{MM}}$	$\Delta G_{\text{MM} \rightarrow \text{QM/MM}}^{\text{bulk}}$	$\Delta G_{\text{QM/MM}}^{\text{bulk}}$	$\Delta G_{\text{MM} \rightarrow \text{QM/MM}}^{\text{site}}$	$\Delta G_{\text{binding}}^{\text{QM/MM}}$
GAFF	-15.92 ± 0.05	11.84	-37.32 ± 0.03	-9.56 ± 0.06	0.44 ± 0.06	-15.48 ± 0.09	-1.04 ± 0.13	-11.04 ± 0.15
GLYCAM	-16.52 ± 0.07	11.41	-37.27 ± 0.03	-9.34 ± 0.08	-1.21 ± 0.06	-17.73 ± 0.12	-4.10 ± 0.07	-12.23 ± 0.12



interaction energies by assuming that the potential energy surfaces of the isolated QM region were the same under the QM and MM Hamiltonians. Therefore, the exact Hamiltonians corresponding to these two QM/MM solvation free energies were not the same. In the binding pocket, the correction free energy from GAFF to QM/MM is $-1.04 \text{ kcal mol}^{-1}$, and that from GLYCAM to QM/MM is $-4.10 \text{ kcal mol}^{-1}$. The difference is over 3 kcal mol^{-1} . The final binding free energies at the QM/MM level are $-11.04 \text{ kcal mol}^{-1}$ and $-12.23 \text{ kcal mol}^{-1}$, which deviated further from the experimental measurement.

It might be surprising that this QM/MM method yielded worse results than the MM method, although the former one is far more demanding than the latter one. It can be explained that the MM method used in this study has been well parameterized but the QM/MM method is yet to be calibrated. Albeit there is rigorous treatment of the intramolecular interaction for the QM region at the density functional theory level, the intermolecular interaction between the QM and MM regions has not been well parameterized. In this QM/MM study, like in most of the QM/MM studies, the interaction between the QM and MM regions included only electrostatic and vdW interactions. The QM region was embedded in the point charges of the MM atoms and the electrostatic interaction between the MM and QM atoms was calculated as the interaction between the point charges and the diffusive electron density. Since no Pauli repulsion is considered, there are no forces that can prevent electron density from spilling out into the MM region.^{78,79} It is well known that this undamped Coulomb interaction overestimates the interaction because a point charge is not a good approximation for the atoms in the vicinity of the QM region.^{80–86} This Coulomb interaction can be corrected by introducing a damping factor or by smearing the charge distribution on the MM atoms. Moreover, the vdW interaction between the QM and MM regions is not calibrated consistently with the Coulomb interaction. Instead, the vdW interactions were calculated at the MM level, which does not depend on the electron density of the QM region. Although this vdW scheme is inexpensive, it can introduce considerable errors into the results.^{87–91}

Conclusion

In this work, the binding affinity of methyl- α -L-fucoside to RSL was studied at both molecular mechanical and quantum mechanical/molecular mechanical levels and further with TP correction at the QM/MM level. The free energy at the MM level was calculated using the double-decoupling method and the free energy change at each step was calculated via a series of intermediate states using Bennett's acceptance ratio (BAR). The QM/MM free energy was calculated by thermodynamic perturbation from the MM Hamiltonians. The main observations of this study are:

- Molecular mechanical methods with fine calibration can well describe the binding affinity of methyl- α -L-fucoside to RSL with errors around $1.0 \text{ kcal mol}^{-1}$. However, the convergence of the free energies is very slow. Normally very long simulations are required for the real states as well as some intermediate

states in the alchemical path. Careful checking of the correlation of the potential energy along the trajectory is critical.

- Simulations using GLYCAM converge much faster than those using the general AMBER force field, due to the ensemble-averaged atomic charges and the null charges for the aliphatic hydrogen atoms in GLYCAM.

- Although the hybrid quantum mechanical/molecular mechanical methods provide more accurate presentation of the charge distribution and the response of the molecule to external fields, the accuracy of these methods is limited by the approximations in the description of the interaction between the quantum mechanical (core) region and the molecular mechanical region (environment). These limitations include the over-polarization by the point charges nearby and the van der Waals interaction being incompatible with the wavefunctions of the quantum mechanical region, which have been observed in some other studies.^{80,81,86–91}

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 21173082), Shanghai Sailing Program (Grant No. 17YF1413200) and the Large Instruments Open Foundation of East China Normal University. CPU time was supported by the Supercomputer Center of East China Normal University. WL and YM thank Dr Yu Kang for many helpful discussions.

References

- 1 R. S. Haltiwanger and J. B. Lowe, Role of Glycosylation In Development, *Annu. Rev. Biochem.*, 2004, **73**, 491–537.
- 2 M. M. Alen, S. J. Kaptein, T. D. Burghgraeve, J. Balzarini, J. Neyts and D. Schols, Antiviral Activity of Carbohydrate-binding Agents and the Role of DC-SIGN in Dengue Virus Infection, *Virology*, 2009, **387**, 67–75.
- 3 P. H. Seeberger, Chemical Glycobiology: Why Now?, *Nat. Chem. Biol.*, 2009, **5**, 368–372.
- 4 Y. Kang, U. Gohlke, O. Engstrom, C. Hamark, T. Scheidt, S. Kunstmann, U. Heinemann, M. Santer and S. Barbirz, Bacteriophage Tailspikes and Bacterial O-antigens as a Model System to Study Weak-Affinity Protein-Polysaccharide Interactions, *J. Am. Chem. Soc.*, 2016, **138**, 9109–9118.
- 5 J.-C. Michalskia and A. Klein, Glycoprotein lysosomal storage disorders: α - and β -mannosidosis, fucosidosis and α -N-acetylgalactosaminidase deficiency, *Biochim. Biophys. Acta, Mol. Basis Dis.*, 1999, **1455**, 69–84.
- 6 B. J. Campbell, L.-G. Yu and J. M. Rhodes, Altered Glycosylation in Inflammatory Bowel Disease: A Possible Role in Cancer Development, *Glycoconjugate J.*, 2001, **18**, 851–858.
- 7 B. A. Cobb and D. L. Kasper, Coming of Age: Carbohydrates and Immunity, *Eur. J. Immunol.*, 2005, **35**, 352–356.
- 8 H. E. Murrey and L. C. Hsieh-Wilson, The Chemical Neurobiology of Carbohydrates, *Chem. Rev.*, 2008, **108**, 1708–1731.



- 9 R. A. Dwek, Glycobiology: Toward Understanding the Function of Sugars, *Chem. Rev.*, 1996, **96**, 683–720.
- 10 H. Lis and N. Sharon, Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition, *Chem. Rev.*, 1998, **98**, 637–674.
- 11 S. Tsuzuki, T. Uchimarui and M. Mikami, Magnitude and Nature of Carbohydrate-Aromatic Interactions: Ab Initio Calculations of Fucose-Benzene Complex, *J. Phys. Chem. B*, 2009, **113**, 5617–5621.
- 12 C. H. Wong, *Carbohydrate-Based Drug Discovery*, Wiley-VCH Verlag GmbH & Co. KGaA, 2003.
- 13 M. M. Fuster and J. D. Esko, The Sweet and Sour of Cancer: Glycan as Novel Therapeutic Targets, *Nat. Rev. Cancer*, 2005, **5**, 526–542.
- 14 N. Sharon, Carbohydrates as Future Anti-Adhesion Drugs for Infectious Disease, *Biochim. Biophys. Acta*, 2006, **1760**, 527–537.
- 15 E. Fadda and R. J. Woods, Molecular Simulations of Carbohydrates and Protein-Carbohydrate Interactions: Motivation, Issues and Prospects, *Drug Discovery Today*, 2010, **15**, 596–609.
- 16 O. C. Grant and R. J. Woods, Recent Advances in Employing Molecular Modelling to Determine the Specificity of Glycan-Binding Proteins, *Curr. Opin. Struct. Biol.*, 2014, **28**, 47–55.
- 17 M. Wimmerová, S. Kozmon, I. Nečasová, S. K. Mishra, J. Komárek and J. Koča, Stacking Interactions between Carbohydrate and Protein Quantified by Combination of Theoretical and Experimental Methods, *PLoS One*, 2012, **7**, e46032.
- 18 R. Sommer, T. E. Exner and A. Titz, A Biophysical Study with Carbohydrate Derivatives Explains the Molecular Basis of Monosaccharide Selectivity of the *Pseudomonas aeruginosa* Lectin LecB, *PLoS One*, 2014, **9**, e112822.
- 19 N. K. Mishra, Z. Kříž, M. Wimmerová and J. Koča, Recognition of Selected Monosaccharides by *Pseudomonas aeruginosa* Lectin II Analyzed by Molecular Dynamics and Free Energy Calculations, *PLoS One*, 2014, **9**, e112822.
- 20 S. K. Mishra, G. Calabró, H. H. Loeffler, J. Michel and J. Koča, Evaluation of Selected Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate Complexes, *J. Chem. Theory Comput.*, 2015, **11**, 3333–3345.
- 21 C. Clarke, R. J. Woods, J. Gluska, A. Cooper, M. A. Nutley and G.-J. Boons, Involvement of Water in Carbohydrate-Protein Binding, *J. Am. Chem. Soc.*, 2001, **123**, 12238–12247.
- 22 R. Kadirvelraj, B. L. Foley, J. D. Dyekjær and R. J. Woods, Involvement of Water in Carbohydrate-Protein Binding: Concanavalin A Revisited, *J. Am. Chem. Soc.*, 2008, **130**, 16933–16942.
- 23 Y. Deng and B. Roux, Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant, *J. Chem. Theory Comput.*, 2006, **2**, 1255–1273.
- 24 S. Boresch, F. Tettinger and M. Leitgeb, Absolute Binding Free Energies: A Quantitative Approach for Their Calculation, *J. Phys. Chem. B*, 2003, **107**, 9535–9551.
- 25 J. Hermans and L. Wang, Inclusion of Loss of Translational and Rotational Freedom in Theoretical Estimates of Free Energies of Binding. Application to a Complex of Benzene and Mutant T4 Lysozyme, *J. Am. Chem. Soc.*, 1997, **119**, 2707–2714.
- 26 N. Hansen and W. F. van Gunsteren, Practical Aspects of Free-Energy Calculations: A Review, *J. Chem. Theory Comput.*, 2014, **10**, 2632–2647.
- 27 D. L. Mobley, J. D. Chodera and K. A. Dill, On the Use of Orientational Restraints and Symmetry Corrections in Chemical Free Energy Calculations, *J. Chem. Phys.*, 2006, **125**, 084902.
- 28 H.-J. Woo and B. Roux, Calculation of Absolute Protein-Ligand Binding Free Energy from Computer Simulations, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6825–6830.
- 29 J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson and V. S. Pande, Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges, *Curr. Opin. Struct. Biol.*, 2011, **21**, 150–160.
- 30 C. D. Christ, A. E. Mark and W. F. van Gunsteren, Basic Ingredients of Free Energy Calculations: A Review, *J. Comput. Chem.*, 2010, **31**, 1569–1582.
- 31 X. Du, Y. Li, Y.-L. Xia, S.-M. Ai, J. Liang, P. Sang, X.-L. Ji and S.-Q. Liu, Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods, *Int. J. Mol. Sci.*, 2016, **17**, 144.
- 32 D. L. Mobley and P. V. Klimovich, Perspective: Alchemical Free Energy Calculations for Drug Discovery, *J. Chem. Phys.*, 2012, **137**, 230901.
- 33 M. R. Shirts, D. L. Mobley and J. D. Chodera, in *Annual Reports in Computational Chemistry*, ed. D. C. Spellmeyer and R. Wheeler, Elsevier, 2007, vol. 3, pp. 41–59.
- 34 H. Senderowitz, C. Parish and W. C. Still, Carbohydrates: United Atom AMBER* Parameterization of Pyranoses and Simulations Yielding Anomeric Free Energies, *J. Am. Chem. Soc.*, 1996, **118**, 2078–2086.
- 35 E. P. Raman, O. Guvench and A. D. MacKerell Jr, CHARMM Additive All-Atom Force Field for Glycosidic Linkages in Carbohydrates Involving Furanoses, *J. Phys. Chem. B*, 2010, **14**, 12981–12994.
- 36 O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison and A. D. Mackerell, CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate-Protein Modeling, *J. Chem. Theory Comput.*, 2011, **7**, 3162–3180.
- 37 K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates, *J. Comput. Chem.*, 2008, **29**, 622–655.
- 38 M. Basma, S. Sundara, T. Vernali and R. J. Woods, Solvated Ensemble Averaging in the Calculation of Partial Atomic Charges, *J. Comput. Chem.*, 2001, **22**, 1125–1137.
- 39 K.-H. Ott and B. Meyer, Parametrization of GROMOS Force Field for Oligosaccharides and Assessment of Efficiency of Molecular Dynamics Simulations, *J. Comput. Chem.*, 1996, **17**, 1068–1084.
- 40 S. A. H. Spieser, J. A. Van Kuik, L. M. J. Kroonbatenburg and J. Kroon, Improved Carbohydrate Force Field for Gromos:



- Ring and Hydroxymethyl Group Conformations and Exo-Anomeric Effect, *Carbohydr. Res.*, 1999, **322**, 264–273.
- 41 W. Damm, A. Frontera, J. T. Rives and W. L. Jorgensen, OPLS All-atom Force Field for Carbohydrates, *J. Comput. Chem.*, 1997, **18**, 1955–1970.
 - 42 D. Kony and W. F. Van Gunsteren, An Improved OPLS-AA Force Field for Carbohydrates, *J. Comput. Chem.*, 2002, **23**, 1416–1429.
 - 43 A. Imberty, E. Bettler, M. Karababa, K. Mazeau, P. Petrova and S. Perez, in *Perspectives in Structural Biology*, ed. M. Vijayan, N. Yathindra and A. S. Kolaskar, Indian Academy of Sciences and Universities Press, Hyderabad, 1999, pp. 392–408.
 - 44 N. L. Allinger, K.-H. Chen, J.-H. Lii and K. A. Durkin, Alcohols, Ethers, Carbohydrates, and Related Compounds. I. The MM4 Force Field for Simple Compounds, *J. Comput. Chem.*, 2003, **24**, 1447–1472.
 - 45 J. H. Lii, K. H. Chen, K. A. Durkin and N. L. Allinger, Alcohols, Ethers, Carbohydrates, and Related Compounds. II. The Anomeric Effect, *J. Comput. Chem.*, 2003, **24**, 1473–1489.
 - 46 J. H. Lii, K. H. Chen and N. L. Allinger, Alcohols, Ethers, Carbohydrates, and Related Compounds. IV. Carbohydrates, *J. Comput. Chem.*, 2003, **24**, 1504–1513.
 - 47 Y. Kang, S. Barbirz, R. Lipowsky and M. Santer, Conformational Diversity of O-Antigen Polysaccharides of the Gram-negative Bacterium *Shigella Flexneri* Serotype Y, *J. Phys. Chem. B*, 2014, **118**, 2523–2534.
 - 48 S. Andre, T. Velasco-Torrijos, R. Leyden, S. Gouin, M. Tosin, P. V. Murphy and H. J. Gabius, Phenylenediamine-Based Bivalent Glycocyclophanes: Synthesis and Analysis of the Influence of Scaffold Rigidity and Ligand Spacing on Lectin Binding in Cell Systems with Different Glycomic Profiles, *Org. Biomol. Chem.*, 2009, **7**, 4715–4725.
 - 49 A. Singh, M. B. Tessier, K. Pederson, X. Wang, A. P. Venot, G. Boons, J. H. Prestegard and R. J. Woods, Extension and Validation of the GLYCAM Force Field Parameters for Modeling Glycosaminoglycans, *Can. J. Chem.*, 2016, **94**, 1–9.
 - 50 B. L. Foley, M. B. Tessier and R. J. Woods, Carbohydrate Force Fields, *WIREs Computational Molecular Science*, 2012, **2**, 652–697.
 - 51 F. V. Toukach and V. P. Ananikov, Recent Advances in Computational Predictions of NMR Parameters for the Structure Elucidation of Carbohydrates: Methods and Limitations, *Chem. Soc. Rev.*, 2013, **42**, 8376–8415.
 - 52 A. J. Kirby and N. H. Williams, in *The Anomeric Effect and Associated Stereoelectronic Effect*, ed. G. R. J. Thatcher, ACS Symposium Series, 1993, vol. 539, pp. 6–25.
 - 53 X. Xiong, Z. Chen, B. P. Cossins, Q. Shao, K. Ding, W. Zhu and J. Shi, Force Fields and Scoring Functions for Carbohydrate Simulation, *Drug Discovery Today*, 2015, **401**, 73–81.
 - 54 J. F. Matthews, G. T. Beckham, M. Bergenstråhle-Wohlert, J. W. Brady, M. E. Himmel and M. F. Crowley, Comparison of Cellulose I β Simulations with Three Carbohydrate Force Fields, *J. Chem. Theory Comput.*, 2012, **8**, 735–748.
 - 55 X. Lu, D. Fang, S. Ito, Y. Okamoto, V. Ovchinnikov and Q. Cui, QM/MM Free Energy Simulations: Recent Progress and Challenges, *Mol. Simul.*, 2016, **42**, 1056–1078.
 - 56 U. Ryde and P. Söderhjelm, Ligand-Binding Affinity Estimates Supported by Quantum Mechanical Methods, *Chem. Rev.*, 2016, **116**, 5520–5566.
 - 57 A. J. Sodt, Y. Mei, G. König, P. Tao, R. P. Steele, B. R. Brooks and Y. Shao, Multiple Environment Single System Quantum Mechanical/Molecular Mechanical(MESS-QM/MM) Calculations. 1. Estimation of Polarization Energies, *J. Phys. Chem. A*, 2015, **119**, 1511–1523.
 - 58 M. R. Reddy and M. D. Erion, Relative Binding Affinities of Fructose-1,6-Bisphosphatase Inhibitors Calculated Using a Quantum Mechanics-Based Free Energy Perturbation Method, *J. Am. Chem. Soc.*, 2007, **129**, 9296–9297.
 - 59 R. S. Rathore, R. N. Reddy, A. K. Kondapi, P. Reddanna and M. R. Reddy, Use of Quantum Mechanics/molecular Mechanics-based FEP Method for Calculating Relative Binding Affinities of FBPase Inhibitors for Type-2 Diabetes, *Theor. Chem. Acc.*, 2012, **131**, 1096.
 - 60 K. Swiderek, S. Marti and V. Moliner, Theoretical Studies of HIV-1 Reverse Transcriptase Inhibition, *Phys. Chem. Chem. Phys.*, 2012, **14**, 12614–12624.
 - 61 G. König, P. S. Hudson, S. Boresch and H. L. Woodcock, Multiscale Free Energy Simulations: An Efficient Method for Connecting Classical MD Simulations to QM or QM/MM Free Energies Using Non-Boltzmann Bennet Reweighting Schemes, *J. Chem. Theory Comput.*, 2014, **10**, 1406–1419.
 - 62 T. P. Straatsma and J. A. McCammon, Treatment of Rotational Isomeric States. III. The Use of Biasing Potentials, *J. Chem. Phys.*, 1994, **101**, 5032–5039.
 - 63 M. Leigeb, C. Schroder and S. Boresch, Alchemical Free Energy Calculations and Multiple Conformational Substates, *J. Chem. Phys.*, 2005, **122**, 084109.
 - 64 G. König, I. F. C. Pickard, Y. Mei and B. R. Brooks, Predicting Hydration Energies with a Hybrid QM/MM Approach: An Evaluation of Implicit and Explicit Solvation Models in SAMPL4, *J. Comput.-Aided Mol. Des.*, 2014, **28**, 245–257.
 - 65 E. Dybeck, G. König, B. R. Brooks and M. R. Shirts, Comparison of Methods To Reweight from Classical Molecular Simulations to QM/MM Potentials, *J. Chem. Theory Comput.*, 2016, **12**, 1466–1480.
 - 66 X. Jia, M. Wang, Y. Shao, G. König, B. R. Brooks, J. Z. H. Zhang and Y. Mei, Calculations of Solvation Free Energy through Energy Reweighting from Molecular Mechanics to Quantum Mechanics, *J. Chem. Theory Comput.*, 2016, **12**, 499–511.
 - 67 W. L. Jorgensen, J. Chandross, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 1983, **79**, 926–935.
 - 68 D. A. Case, J. T. Berryman, R. M. Betz, D. S. Cerutti, T. E. I. Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke and A. W. Goetz, *et al.*, *AMBER 15*, University of California, San Francisco, 2015.



- 69 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes, *J. Chem. Phys.*, 1977, **23**, 327–341.
- 70 T. Darden, D. York and L. Pedersen, Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 71 U. Essmann, L. Perera and M. L. Berkowitz, A Smooth Particle Mesh Ewald Method, *J. Chem. Phys.*, 1995, **103**, 8577.
- 72 A. D. Becke, Density-functional Exchange-energy Approximation with Correct Asymptotic Behavior, *Phys. Rev. A*, 1988, **38**, 3098–3100.
- 73 C. Lee, W. Yang and R. G. Parr, Development of the Colle-Salvetti Correlation-energy Formula into a Functional of the Electron Density, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
- 74 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci and G. A. Petersson, *et al.*, *Gaussian 09, Revision B.01*, Gaussian, Inc., Wallingford, CT, 2010.
- 75 G. König, Y. Mei, I. F. C. Pickard, A. C. Simmonett, B. T. Miller, J. M. Herbert, H. L. Woodcock, B. R. Brooks and Y. Shao, Computation of Hydration Free Energies Using the Multiple Environment Single System Quantum Mechanical/Molecular Mechanical Method, *J. Chem. Theory Comput.*, 2016, **12**, 332–344.
- 76 J. D. Chodera, W. C. Swope, J. W. Pitner, C. Seok and K. A. Dill, Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations, *J. Chem. Theory Comput.*, 2007, **3**, 26–41.
- 77 S. K. Mishra, J. Adam, M. Wimmerová and J. Koča, In Silico Mutagenesis and Docking Study of Ralstonia solanacearum RSL Lectin: Performance of Docking Software To Predict Saccharide Binding, *J. Chem. Inf. Model.*, 2012, **52**, 1250–1261.
- 78 L. J. Năbo, J. M. H. Olsen, N. H. List, L. M. Solanko, D. Wüstner and J. Kongsted, Embedding Beyond Electrostatics—The Role of Wave Function Confinement, *J. Chem. Phys.*, 2016, **145**, 104102.
- 79 G. Groenhof, in *Biomolecular Simulations: Methods and Protocols*, ed. L. Monticelli and E. Salonen, Humana Press, 2013, pp. 43–66.
- 80 D. D. Das, K. P. Eurenus, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodošček and B. R. Brooks, Optimization of Quantum Mechanical Molecular Mechanical Partitioning Schemes: Gaussian Delocalization of Molecular Mechanical Charges and the Double Link Atom Method, *J. Chem. Phys.*, 2002, **117**, 10534–10547.
- 81 G. A. Cisneros, S. N.-I. Tholander, O. Parisel, T. A. Darden, D. Elking, L. Perera and J.-P. Piquemal, Simple Formulas for Improved Point-charge Electrostatics in Classical Force Fields and Hybrid Quantum Mechanical/Molecular Mechanical Embedding, *Int. J. Quantum Chem.*, 2008, **108**, 1905–1912.
- 82 B. Wang and D. G. Truhlar, Geometry Optimization Using Tuned and Balanced Redistributed Charge Schemes for Combined Quantum Mechanical and Molecular Mechanical Calculations, *Phys. Chem. Chem. Phys.*, 2011, **13**, 10556–10564.
- 83 B. Wang and D. G. Truhlar, Partial Atomic Charges and Screened Charge Models of the Electrostatic Potential, *J. Chem. Theory Comput.*, 2012, **8**, 1989–1998.
- 84 G. Hou, X. Zhu, M. Elstner and Q. Cui, A Modified QM/MM Hamiltonian with the Self-Consistent-Charge Density-Functional-Tight-Binding Theory for Highly Charged QM Regions, *J. Chem. Theory Comput.*, 2012, **8**, 4293–4304.
- 85 B. Wang and D. G. Truhlar, Tuned and Balanced Redistributed Charge Scheme for Combined Quantum Mechanical and Molecular Mechanical (QM/MM) Methods and Fragment Methods: Tuning Based on the CM5 Charge Model, *J. Chem. Theory Comput.*, 2013, **9**, 1036–1042.
- 86 B. Wang and D. G. Truhlar, Screened Electrostatic Interactions in Molecular Mechanics, *J. Chem. Theory Comput.*, 2014, **10**, 4480–4487.
- 87 M. Freindorf and J. Gao, Optimization of the Lennard-Jones Parameters for a Combined ab initio Quantum Mechanical and Molecular Mechanical Potential Using the 3-21G Basis Set, *J. Comput. Chem.*, 1996, **17**, 386–395.
- 88 D. Riccardi, G. Li and Q. Cui, Importance of van der Waals Interactions in QM/MM Simulations, *J. Phys. Chem. B*, 2004, **108**, 6467–6478.
- 89 M. Freindorf, Y. Shao, T. R. Furlani and J. Kong, Lennard-Jones parameters for the combined QM/MM method using the B3LYP/6-31G*/AMBER potential, *J. Comput. Chem.*, 2005, **26**, 1270–1278.
- 90 T. J. Giese and D. M. York, Charge-dependent Model for Many-body Polarization, Exchange, and Dispersion Interactions in Hybrid Quantum Mechanical/Molecular Mechanical Calculations, *J. Chem. Phys.*, 2007, **127**, 194101.
- 91 E. R. Kuechler, T. J. Giese and D. M. York, Charge-dependent Many-body Exchange and Dispersion Interactions in Combined QM/MM Simulations, *J. Chem. Phys.*, 2015, **143**, 234111.

