PCCP

PAPER

View Article Online
View Journal | View Issue

Calculating binding free energies of host–guest systems using the AMOEBA polarizable force field†

Cite this: *Phys. Chem. Chem. Phys.*, 2016, **18**, 30261

David R. Bell,‡[a] Rui Qi,‡[a] Zhifeng Jing,‡[a] Jin Yu Xiang,[b] Christopher Mejias,[c] Michael J. Schnieders,[d] Jay W. Ponder[c] and Pengyu Ren*[a]

Received 14th April 2016,
Accepted 19th May 2016

DOI: 10.1039/c6cp02509a

www.rsc.org/pccp

Molecular recognition is of paramount interest in many applications. Here we investigate a series of host–guest systems previously used in the SAMPL4 blind challenge by using molecular simulations and the AMOEBA polarizable force field. The free energy results computed by Bennett's acceptance ratio (BAR) method using the AMOEBA polarizable force field ranked favorably among the entries submitted to the SAMPL4 host–guest competition [Muddana, *et al.*, *J. Comput.-Aided Mol. Des.*, 2014, **28**, 305–317]. In this work we conduct an in-depth analysis of the AMOEBA force field host–guest binding thermodynamics by using both BAR and the orthogonal space random walk (OSRW) methods. The binding entropy–enthalpy contributions are analyzed for each host–guest system. For systems of inordinate binding entropy–enthalpy values, we further examine the hydrogen bonding patterns and configurational entropy contribution. The binding mechanism of this series of host–guest systems varies from ligand to ligand, driven by enthalpy and/or entropy changes. Convergence of BAR and OSRW binding free energy methods is discussed. Ultimately, this work illustrates the value of molecular modelling and advanced force fields for the exploration and interpretation of binding thermodynamics.

Introduction

Molecular recognition is fundamental to biological processes and is utilized in applications ranging from therapeutics to chemical sensing.[1] Understanding the importance of molecular recognition, the interactions involved are exceedingly complex and dependent upon a high degree of order between the solutes and the solvent for binding. Computer prediction of binding affinity holds potential to accurately capture thermodynamic information from different states as well as allows for the design of novel ligands.

Molecular modelling and simulation can be a powerful tool for the quantitative understanding of the driving forces underlying molecular recognition.[2,3] However, accurate computation of binding free energy *via* molecular modelling faces two major challenges. First, the energetic description of binding requires high accuracy potential energy that is also transferable between different chemical and physical environments. Second, the flexibility of guest, water and host molecules results in many degrees of freedom making it difficult to adequately explore the configuration space using molecular dynamics. With increasing complexity up to protein–ligand systems, sufficient sampling of binding by traditional methods becomes limited by computational cost.

Numerous potential energy methods have been proposed to compute binding free energy, increasing in complexity from empirical docking methods to quantum mechanics (QM) calculations.[4] Empirical docking methods[5] are frequently used for library screening and though they allow for fast calculation, they do not maintain high accuracy of the potential energy function nor do they allow for sufficient sampling of binding conformations. QM calculations of binding free energy[6–8] are limited to small, predetermined binding sites. Bridging the gap between docking methods and QM, semi-empirical force-field methods use Molecular Dynamics or Monte Carlo sampling schemes to generate many configurations and energies.[9,10] In semi-empirical force field methods, the potential energy of the system is computed from analytical functions of the atomic coordinates. Classical force fields such as AMBER,[11] CHARMM,[12] OPLS-AA,[13] or GROMOS[14] typically represent intermolecular interactions by

[a] *Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, USA. E-mail: pren@mail.utexas.edu*
[b] *Department of Biochemistry and Molecular Biophysics, Washington University in St. Louis, St. Louis, Missouri 63110, USA*
[c] *Department of Chemistry, Washington University in Saint Louis, Saint Louis, Missouri 63130, USA*
[d] *Department of Biomedical Engineering and Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242, USA*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6cp02509a

‡ These authors contributed equally to this work.

a van der Waals (vdW) term and point charge electrostatics. This representation is computationally efficient and sufficiently accurate for many applications. However, the potential energy is limited by not capturing the electrostatic response to environmental stimulus, referred to as the polarization effect.[15,16] Additionally, modelling electrostatics as point charges neglects the intricate yet substantial effect of charge distribution,[17] which can be properly captured by higher order multipole moments.[18] Therefore, tremendous efforts have been made to develop advanced representations of electrostatics ranging from fluctuating charges,[19] Drude oscillators,[20,21] up to fully polarizable force-fields such as AMOEBA.[22,23]

Methods for binding free energy calculation can be classified according to the depiction of either alchemical or physical pathways. The alchemical pathway uses alchemical, or non-physical intermediates to compute binding free energy, which is popular for its general applicability and efficiency. Physical pathways are preferable for large molecules and can give the binding mechanism and kinetics.[24,25] While traditional methods such as Bennett's acceptance ratio (BAR)[26] have been successful, an improvement in computational efficiency is desired for application to large systems and more sophisticated potential energy representations. To this end, many enhanced sampling methods have been developed.[27,28]

Host–guest systems are often used as a model for binding affinity prediction because of their modest size and high specificity among guests. By computing the free energy behaviour of relatively small molecules, inadequacies can be better determined and remediated for the rigorous and strenuous computation of binding free energies for large proteins. In the SAMPL3[29] and SAMPL4[30] host–guest binding competitions, the cucurbit[7]uril macrocycle was used as the host molecule. The cucurbit[*n*]uril macrocycle (CB[*n*]) is composed of *n* conjoined glycoluril subunits forming a cylindrical molecule approximately 9.1 Å in height (for a thorough review of CB[*n*] chemistry, see ref. 31). As with many macrocycles, such as cyclodextrin, CB[*n*] has been explored as a molecular container for drug delivery.[32–34] The glycoluril subunits position a ring of carbonyl groups at the two faces of the cylinder, while the inner region of carbon–nitrogen chains remains hydrophobic. Hence, guests of hydrophobic cores with cationic end groups can bind with a high affinity for the CB[*n*] host.

In this work, we report the investigation of host–guest binding thermodynamics between a CB[7] host and a set of 14 small molecules. The guests range from linear hydrocarbons to cycloalkanes, species of norbornanes and adamantane. We use two free energy calculation methods and several thermodynamic inquiries to interpret experimental affinities. In particular, we dissect the roles of entropy and enthalpy in binding for each guest. For anomalous enthalpy/entropy values, the separate entropy contributions of water and the host–guest systems are investigated. We determine that binding affinities of the host–guest systems are both enthalpy- and entropy-driven. We further discuss the application and convergence of the OSRW and BAR binding free energy methods. Our results attest to the application of binding free energy simulation methods towards the understanding of experimental binding affinities.

Methods

AMOEBA polarizable force field

An implementation of the polarizable AMOEBA force field with the molecular modelling software package, TINKER,[35] is used in this study. Typical force fields treat charges as static entities, usually represented as fixed-atomic charges.[15,36] However, the actual charge distributions of atoms change in response to the environment's electric field.[15,36–38] As a physics-grounded force field, AMOEBA depicts molecular polarizability and electrostatic potential terms by using mutual atomic dipole–dipole induction along with permanent atomic monopole, dipole, and quadrupole moments.[22] This results in a more accurate description of molecular energetics in protein–ligand binding. Although the polarizability of the CB[7] cavity is posited to be low,[39] the authors anticipate that the large number of heavy atoms of the system as well as the differences in electronegativity of the guests make this host–guest system appropriate for the employment of a polarizable force field.

Bennet acceptance ratio (BAR)

The BAR[26] is a method to calculate the free energy difference between different thermodynamic states. It has been shown to be more efficient than other classic methods such as free energy perturbation.[40] Typically, simulations are conducted at multiple intermediate states that connect the two end states, and the free energy difference between neighbour states is calculated based on the energy difference.

Orthogonal space random walk (OSRW)

OSRW is an enhanced sampling scheme for free energy calculation, which performs a random walk in two orthogonal dimensions.[41–44] One dimension is along the order parameter λ representing an alchemical intermediate state that connects the two states of interest.[45,46] The other dimension is along the orthogonal generalized force ($F_\lambda = \partial U/\partial\lambda$), whose integral is the free energy (eqn (1)). Once a state is sampled, a Gaussian distributed bias is added to discourage the system from revisiting that state. The main advantage of OSRW over many other methods including the BAR is that it accelerates the sampling of the orthogonal generalized force. A complete explanation of the method as well as the requisite adjustments needed to employ a polarizable force field can be found in Abella *et al.*[47]

$$\Delta G = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda \tag{1}$$

Absolute binding free energy

The absolute binding free energy ΔG_{bind} of the host–guest system was calculated by using the double-decoupling method, which refers to "disappearing" the guest in both the solvent and solvated host–guest complex.[48] As illustrated in Fig. S1 (ESI†), and articulated in eqn (2), the binding free energy is the free energy difference between removing a guest from its water environment (ΔG_{hyd}) and decoupling a guest from its host–water environments (ΔG_{host}). At $\lambda = 0$ the guest is completely decoupled

 This journal is © the Owner Societies 2016

from its environment and can wander to other parts of the box, prolonging convergence. To solve this problem, a harmonic restraint[49] (k = 15 kcal mol^{-1}) is added between the centers of mass of the host and the guest, and a correction term[48–50] is needed to recover the true free energy difference (eqn (3)).

$$\Delta G_{\text{bind}} = \Delta G_{\text{host–guest}} - \Delta G_{\text{hydration}} + G_{\text{correction}} \quad (2)$$

$$G_{\text{correction}} = RT \ln\left[C^{\circ}\left(\frac{2\pi RT}{k}\right)^{\frac{3}{2}}\right] \quad (3)$$

where R is the gas constant, T is the temperature, C° is the standard concentration and k is the force constant of the restraint. Finite size effects on charging free energies[51] were not corrected since they are expected to be similar for $\Delta G_{\text{host–guest}}$ and $\Delta G_{\text{hydration}}$ and will cancel out.

Computational details

In this study, the absolute binding free energy values of 14 guests in the SAMPL4 CB[7]–guest system were calculated using the polarizable AMOEBA force field. Parameters for the molecules were derived by following the procedure previously described in Ren *et al.*[23] All molecular dynamics simulations were run using TINKER with a RESPA integrator[52] with a 2.0 femtosecond time step and a Bussi thermostat.[53] The vdW calculations had a 12.0 Å cutoff while the electrostatics was calculated by particle mesh Ewald summation with a real-space cutoff of 7.0 Å. The Gaussian bias was deposited every 10 steps, with a height of 0.005 kcal mol^{-1} and widths of 4 kcal mol^{-1} for F_{λ} and 0.01 for λ. Additional simulations with a reduced height of 0.001 or 0.002 kcal mol^{-1} were also carried out for some guests. The production time of the OSRW is around 15–20 ns. All OSRW simulations were conducted on Texas Advanced Computing Center (TACC) Stampede as well as a local computer cluster. For the BAR simulations, first the electrostatics were gradually scaled off with vdW interactions kept at full strength, and then the vdW interactions were scaled off. The numbers of steps for these two stages were 11–12 and 10–13 respectively. The total simulation time for each step was 1 ns and coordinates were saved every 1 ps for analysis. The correction $G_{\text{correction}}$ was 6.245 kcal mol^{-1} and should be added to all binding free energy calculations for both BAR and OSRW. The uncertainties of the BAR results were estimated based on the distribution of uncorrelated samples, while the uncertainties of the OSRW results were obtained by comparing the final values of independent simulations and are imprecise due to the small numbers of simulations. The binding enthalpy was obtained from the difference between the average energies in the binding and free states. This method has comparable accuracy with that of the van't Hoff method[54] and that of the BAR method.[55]

Results

Fig. 1 and Table 1 both present binding free energy results from OSRW and BAR computations compared with experiment. In Table 1, structures and energies of the guest ligands studied here are presented.[30] The host for all ligands is CB[7] as stated previously. For each ligand in Table 1, the free energy values of the experiment, OSRW, and BAR are presented explicitly. Reported in the SAMPL4 results, the absolute uncertainty of all experimental free energy values is ±0.1 kcal mol^{-1}. The BAR results are those that were previously reported in the SAMPL4 contest.[30]

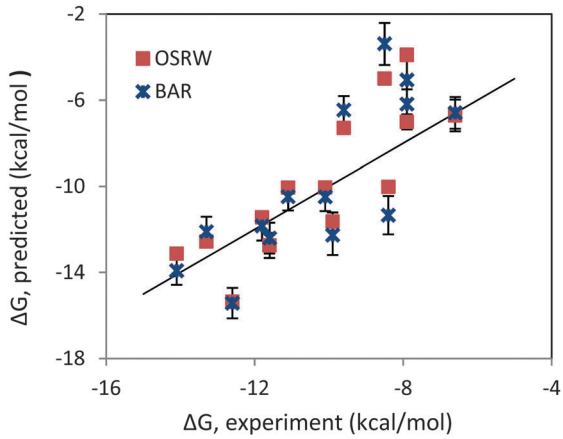


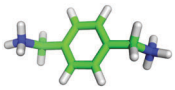
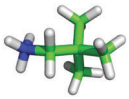
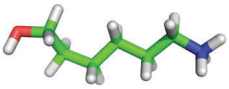
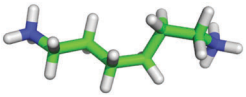
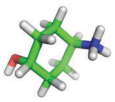
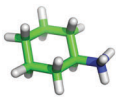
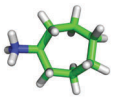
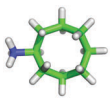
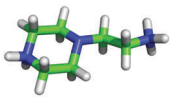
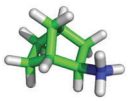
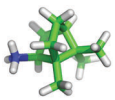
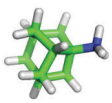
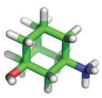
Fig. 1 Predicted binding free energy as a function of experimental binding free energy (in kcal mol^{-1}). Line is $y = x$.

In Fig. 1, both OSRW and BAR results are plotted against experimental free energy values taken from the SAMPL4 host–guest competition. The OSRW and BAR free energies establish good correlation with the experiment, having R^2 correlation values of 0.69 (OSRW) and 0.62 (BAR). The OSRW and BAR results also agree with each other within statistical uncertainties for most of the systems. The discrepancies for the other systems can be accounted for by the imprecise uncertainty estimate of the small numbers of OSRW simulations.

In Table S1 (ESI†), a decomposition of binding free energies is given. For ligand C5, positive binding free energies calculated from the BAR led to the exploration of multiple ligand protonation states, denoted as C5 and C5b (see Fig. S7, ESI† for structure comparison). In five ligand cases (C1, C3, C5b, C9, and C10), the OSRW computation displayed large fluctuations in free energy. Since the fluctuation is proportional to the bias deposition rate, additional OSRW simulations were conducted with decreased Gaussian-height biases for each of these ligands. In theory, lowering the height of the Gaussian distribution will suppress fluctuations at the expense of slowing down dynamics. However, in this work, the OSRW computations with a lowered Gaussian height (LGH) bias converged at roughly the same simulation time as the original computations. Lastly, in Table S1 (ESI†), ligands C3 and C10 were duplicated in the OSRW computation due to poor convergence of the original simulations.

For Fig. 1 and Table 1, the final OSRW ligand binding free energy value is taken to be the average over all of the OSRW computations for that ligand, with some values excluded (explained below). Multiple independent OSRW simulations were run for each ligand. As mentioned above, for two ligands, OSRW computations were repeated with the original parameters. The averaging of the free energies includes the LGH and repeated

Table 1 Host–guest binding free energies. The OSRW column presents the average of results from the full length of simulations, while the OSRW (10 ns) column presents values cut off at 10 ns. The host molecule for all structures is cucurbit[7]uril. All free energies are given in kcal mol^{-1}. The experimental free energies hold an uncertainty of ±0.1 kcal mol^{-1}

		ΔG_{bind}			
Guest	Guest structure	Experimental	BAR	OSRW	OSRW (10 ns)
C1		−9.9	−12.27 ± 0.92	−11.64 ± 0.42	−10.43 ± 1.01
C2		−9.6	−6.46 ± 0.65	−7.29 ± 0.20	−5.50 ± 0.05
C3		−6.6	−6.59 ± 0.74	−6.71 ± 0.74	−5.17 ± 1.03
C4		−8.4	−11.34 ± 0.89	−10.02 ± 0.25	−8.31 ± 0.75
C5b		−8.5	−3.39 ± 0.97	−5.00 ± 0.11	−4.46 ± 0.26
C6		−7.9	−6.18 ± 0.69	−7.01 ± 0.35	−6.64 ± 0.07
C7		−10.1	−10.49 ± 0.66	−10.05 ± 0.09	−10.96 ± 0.78
C8		−11.8	−11.84 ± 0.68	−11.44 ± 0.09	−11.15 ± 0.76
C9		−12.6	−15.42 ± 0.71	−15.35 ± 0.29	−15.44 ± 0.65
C10		−7.9	−5.06 ± 0.91	−3.90 ± 0.26	−3.69 ± 0.68
C11		−11.1	−10.48 ± 0.64	−10.06 ± 0.25	−9.82 ± 0.34
C12		−13.3	−12.11 ± 0.70	−12.57 ± 0.03	−12.33 ± 0.16
C13		−14.1	−13.92 ± 0.65	−13.13 ± 0.13	−12.63 ± 0.52
C14		−11.6	−12.41 ± 0.72	−12.75 ± 0.59	−12.05 ± 0.58

computations with the original pair of simulations. Exceptions to this average method are ligands C5 and ligands C10. The binding free energy value for ligand C5 was taken to be the average of the ligand C5b_LGH computation. The protonation state for ligand C5 reported by the BAR computations was similarly C5b. The 2.5 kcal mol^{-1} disagreement between original OSRW simulations for ligand C5b, as well as the nice agreement between the C5b_LGH simulations (within 0.3 kcal mol^{-1}),

 This journal is © the Owner Societies 2016

Table 2 Model deviation from the experiment. RMS energy difference, and AUE (average unsigned error) are in kcal mol^{-1}

Method	RMS error	AUE	R^2	Kendall τ
OSRW	1.92	1.51	0.69	0.74
BAR	2.26	1.73	0.62	0.58
OSRW (10 ns)	2.22	1.73	0.73	0.75

supported our use of the C5b_LGH data. For ligand C10, all of the OSRW free energy values were used in the average except the −0.76 value as it was in disagreement with all of the other five values by 2.5 kcal mol^{-1}. We suspect that this low free energy value is an artefact of a slow-convergence binding energy computation.

Table 2 presents errors and correlation metrics between OSRW/BAR and experimental values. Despite the duplicate runs and Gaussian height decrease necessary for the OSRW computations, the Kendall τ coefficient for OSRW supports a strong agreement between OSRW and experiment. For further validation, in Fig. S2–S4 (ESI†), we have computed correlation metrics for all of the possible answer combinations from the host–guest and hydration free energy values in Table S2 (ESI†). These three figures show the distribution of possible answer choices as well as our reported value and mean of possible answer choices. The OSRW computation times needed for all ligands range from 13.76–23.95 ns, further elucidated in the discussion section. The computational expense was likewise heavy, with several weeks required using the Texas Advanced Computing Centre, as well as several weeks on a local cluster.

Discussion

Analysis of computed binding affinities from the SAMPL challenge allows for the elucidation of binding thermodynamics as well as the examination of computational predictions. In the official SAMPL4 host–guest paper, free energy values from BAR simulations using the AMOEBA polarizable force field were noted for good performance.[30] Our OSRW-computed free energies correlate with experimental values slightly better than the BAR results. Note that both methods use the exact same parameter sets and simulation parameters. However, long computational times needed for the convergence of OSRW free energy were observed. Upwards of 20 ns of computation time in binding was required for some ligands, while in our previous work, the hydration free energy was able to converge in less than 4 ns.[47] For comparison, the BAR computations were performed for 1 ns for each vdW and electrostatic window. One possible reason that the OSRW method applied here may be slow to converge is due to the underlying metadynamics procedure. Recently, the orthogonal space tempering[43] method has been proposed to address this problem.

We also investigated the OSRW results if the free energy computations were carried out for only 10 ns rather than continued to 15+ ns. In Tables 1 and 2 the 10 ns OSRW binding free energies are presented in comparison to the experimental values. Surprisingly, the R^2 correlation value and the Kendall τ correlation coefficient are high, supporting strong correlation between OSRW and experiment after just 10 ns simulations (Table 2). Despite this strong correlation, the individual ligand errors and the RMSE between the experiment and OSRW are slightly higher than the final results.

To gain insights into the molecular driving forces for binding, we have examined the enthalpy and entropy contributions of the binding free energy. Table 3 lists the calculated binding enthalpy and entropy for each guest ligand. Although the binding free energies for different ligands are close, ranging from −15 to −5 kcal mol^{-1}, the binding enthalpies are vastly different. This is a good demonstration of the enthalpy–entropy compensation in host–guest binding. Due to the relatively short simulation time (1 ns), the uncertainties are on the order of 10 kcal mol^{-1}. Nevertheless, it can be seen that some of the recognitions are driven by enthalpy while others by entropy. Ligands 9, 12, and 13 have both favourable binding enthalpy and entropy. Extreme examples are ligand C10 for entropy-driven binding, and ligands C7 and C8 for enthalpy-driven binding. However, there appears to be no simple relationship between the binding thermodynamics of the ligand and its charge or geometry. Comparing C5 with C5b and C3 with C4, we find that the binding enthalpy does not correlate with the net charge. Ligands C7, C8 and C9 have the same functional groups and their binding affinities increase with the ring size, but their entropy values differ. Enthalpy values of ligands C7 and C8 clearly indicate a dominant contribution of enthalpy, while for ligand C9 the enthalpy value is competitive with entropy.

Further analyses were carried out to look into the binding mechanisms. To explain why guest ligands C7 and C8 are enthalpy-driven, we investigated the ligand hydrogen bonding formation in water and complexes. Table 4 lists hydrogen bond (H-bond) numbers for ligands C7, C8 and C10 between guest–water in solution and between guest–host/water in the complex. The data are averaged over 1000 frames in 1 ns. Compared to ligands C7 and C8, ligand C10 formed more H-bonds when free in water (5.7) and bound to the complex (5.4). Furthermore, we analysed the portion of H-bonds formed between guest–host and guest–water. The three ligands formed similar numbers of H-bonds with the host while ligand C10 has twice the H-bonds

Table 3 Host–guest binding enthalpies and entropies (kcal mol^{-1}). STD(ΔH) are the uncertainty of enthalpy

Guest	ΔH	STD(ΔH)	$-T\Delta S$
C1	−14.91	13.87	2.64
C2	−17.39	13.54	10.93
C3	−18.58	14.79	12.00
C4	−6.62	14.02	−4.72
C5	3.03	15.24	−6.41
C5b	−5.08	13.96	1.53
C6	−12.48	14.60	6.30
C7	−26.99	13.47	16.56
C8	−28.56	14.30	16.72
C9	−3.69	13.41	−11.73
C10	45.20	13.71	−49.57
C11	1.33	13.37	−11.94
C12	−5.58	12.68	−6.67
C13	−7.53	14.19	−6.18
C14	4.28	12.63	−16.90

Table 4 Analysis of hydrogen bond numbers for guests C7, C8 and C10. The number of hydrogen bonds between guest–water in solution and between guest–host/water in host–guest complex are listed as $N_{solution}$ and $N_{complex}$ respectively. Further decompositions of hydrogen bond numbers between guest–host, and between guest–water in host–guest complexes are given in $N^{g-h}_{complex}$ and $N^{w-h}_{complex}$. The presenting hydrogen bond numbers are averaged by 1000 frames over 1 ns

Guest	$N_{solution}$	$N_{complex}$	$N^{g-h}_{complex}$	$N^{g-w}_{complex}$
C7	2.690	3.554	1.617	1.937
C8	2.792	3.389	1.201	2.188
C10	5.709	5.452	1.325	4.127

formed with the surrounding water than other ligands. This may be attributed to the structural differences: ligand C10 has 3 polar amine groups with two of them exposed to the surroundings, attracting water and other polar groups. In contrast, ligands C7 and C8 have only one amine group each, leading to less intermolecular interaction. Noticeably, an increase of H-bonds in ligands C7 and C8 is found when moved from solution to the host–guest complex. On the other hand, the number of H-bonds formed by C10 decreases upon binding. The changes in H-bonds may explain why the binding of ligands C7 and C8 was found to be enthalpy-driven.

The rotation of guest ligands C7, C8, and C10 inside the CB[7] host was measured to explore the entropic aspects of these ligands. Three atoms from each ligand's aromatic ring were chosen to represent one plane, while three atoms from the host were chosen to produce a plane that bisects the host equally. The rotation of the guest plane with respect to the host plane was measured over the coordinates of 5 ns trajectories. The potential of mean force (PMF) was also computed for the rotation angles. Similar to a study of an octa-acid host–guest system,[7] the guest ligands here were determined to rotate almost freely with only small free energy barriers ($\sim$0.5 kcal mol^{-1}). Likewise, computation of the entropy using $S = -k_B \times \Sigma p \ln(p)$ resulted in minute contributions ($S^{rot}_{lig(complex)} - S^{rot}_{lig(free)} \approx 0.05$ kcal mol^{-1} at $T = 300$ K), shown in Table S2 (ESI†).

The configurational entropies of host–guest complexes C7, C8, and C10 were computed using quasiharmonic analysis.[56,57] In the quasiharmonic analysis method, the mass weighted covariance matrix of atomic fluctuations is computed. Eigenvalues λ_i of this covariance matrix are then expounded to frequencies of collective motions, $\omega_i = (RT/\lambda_i)^{1/2}$. The estimated entropy S of the molecule is determined by eqn (5) where R is the gas constant, $\hbar$ is Planck's constant, and T is temperature.

$$S = R\sum_{i=1}^{3N-6}\frac{\hbar\omega_i/RT}{e^{\hbar\omega_i/RT}-1} - \ln\left(1-e^{-\hbar\omega_i/RT}\right) \quad (5)$$

The quasiharmonic entropy was computed using AMBER14.[58] For each molecule, all heavy atoms (C, N, O) were included in the covariance matrix. Table 5 shows the quasiharmonic entropy values for the host–guest systems C7, C8 and C10. These values include entropies of the host–guest complex S_{hg}, the guest only in complex $S_{g(complex)}$, the host only in complex $S_{h(complex)}$, the guest in solution $S_{g(solution)}$, the host in solution $S_{h(solution)}$, and the entropic contribution of binding $-T\Delta S_{conf}$ where

Table 5 Configurational entropy computed from quasiharmonic analysis.[a] $S_{h(solution)}$ is 495.61 cal mol^{-1} K^{-1}

Guest	S_{hg}	$S_{g(complex)}$	$S_{h(complex)}$	$S_{g(solution)}$	$-T\Delta S_{conf}$
C7	364.38	74.53	302.73	80.99	5.69
C8	379.41	91.88	289.87	96.34	3.12
C10	366.28	87.67	294.41	94.59	6.61

[a] Entropy values are given for S_{hg} the host–guest complex, $S_{g(complex)}$ the guest only in complex, $S_{h(complex)}$ the host only in complex, and $S_{g(solution)}$ the guest in solution. S_{hg}, $S_{g(complex)}$, $S_{h(complex)}$, and $S_{h(solution)}$ are computed from 5 ns simulations while $S_{g(solution)}$ values are computed from 1 ns simulations. All entropy values (except where marked) in cal mol^{-1} K^{-1}. $\Delta S_{conf} = S_{hg} - S_{h(solution)} - S_{g(solution)}$. $T\Delta S_{conf}$ computed at 300 K, with units of kcal mol^{-1}.

$\Delta S_{conf} = S_{hg} - S_{h(solution)} - S_{g(solution)}$. The quasiharmonic approximation maintains limitations involving the use of Cartesian coordinates and the presence of multiple steep energy wells.[59] Further, the quasiharmonic approximation is known to present an upper-bound to entropy primarily due to correlations between modes.[60,61] However, several trends may be observed from the computed values. The CB[7]–C10 complex has the highest entropic cost ($-T\Delta S_{conf}$) out of the three complexes computed.

Given that the enthalpy/entropy decomposition analysis suggested binding of guest C10 to be entropically favorable ($-T\Delta S_{tot} < 0$), the positive configurational entropy change computed here (Table 5) indicates that the favourable binding entropy of ligand C10 is likely water driven. Binding of guests C7 and C8 resulted in approximately the same entropic cost. Although the values of S_{hg} and $S_{g(solution)}$ differ for guests C7 and C8, when combined, the values largely offset the differences. From the analysis of guests C7 and C8, intramolecular atomic fluctuations of the aromatic carbon atoms inside the host are greater for C8 than for C7. This is consistent with intuition: the larger aromatic molecule of ligand 8 is slightly pressed inward by the host. This effect is evident in the $S_{h(complex)}$ values, where the host in guest C7 complex has roughly 4 kcal mol^{-1} greater entropy (TS) than the host in guest C8 complex, which is strained due to the ligand size. Similar to the C10 complex, guest C7 and C8 complexes have a positive (unfavourable) entropy contribution, and additional unfavourable entropic interactions from water likely increase the binding entropy to the values shown in Table 3.

As noted above, there are discrepancies between OSRW and BAR results as well as between independent OSRW simulations for some ligands. To explain this, we observe that for an unbiased estimator, the uncertainty of a measured quantity is related to the sample distribution and the autocorrelation time as[40]

$$\sigma(\bar{A}) = \sigma(A)\sqrt{\frac{2\tau}{t}} \quad (6)$$

where τ is the integrated autocorrelation time and t is the total sampling time. $t/2\tau$ is also interpreted as the effective number of independent samples. Eqn (6) is valid for BAR. As for OSRW, since the underlying metadynamics does not converge asymptotically,[62] eqn (6) should provide a lower bound for its error. The sample distribution depends on the hybrid Hamiltonian,

 This journal is © the Owner Societies 2016

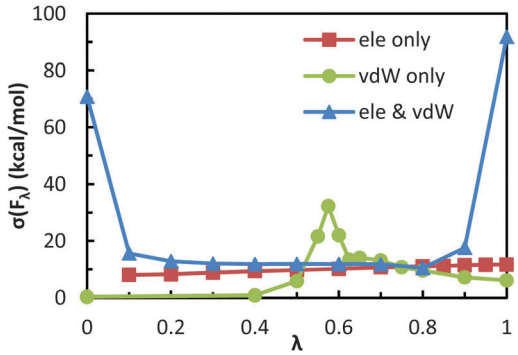


Fig. 2 Standard deviation of F_λ as a function of λ for different coupling schemes. All analyses are based on the decoupling of guest C10 from its host–guest complex state. "vdW only" means that the vdW interaction is decoupled when there is no electrostatics. "ele only" means that the electrostatics is decoupled while vdW interaction is modelled at full strength. "ele & vdW" means that both electrostatics and vdW interactions are decoupled simultaneously as in the current OSRW implementation.

i.e. the decoupling scheme for the alchemical transition, which is different in the OSRW and BAR simulations. The correlation time varies with the simulation method. Generally, the correlation time in metadynamics should be shorter than that of a classical molecular dynamics simulation on the same Hamiltonian. However, it is difficult to compare the correlation time between OSRW and BAR because OSRW has an additional degree of freedom λ. So here we focus on the effect of the decoupling scheme on the convergence. Fig. 2 shows the standard deviation of F_λ for decoupling of guest C10 from its host–guest complex state in different decoupling schemes. When only the vdW interaction is decoupled (scaled down), the distribution F_λ is very narrow up to $\lambda = 0.5$. $\sigma(F_\lambda)$ increases sharply and then falls to roughly 10 kcal mol^{-1} when λ goes from 0.5 to 0.6. When the electrostatic interaction is decoupled in the presence of vdW interaction, $\sigma(F_\lambda)$ is nearly constant around 10 kcal mol^{-1}, which means that there is no dramatic change in phase space and that the evenly spaced λ points perform very well in distributing the simulation time. When vdW and electrostatic interactions are decoupled simultaneously, $\sigma(F_\lambda)$ is significantly higher than when the two interactions are decoupled separately as λ approaches 0 and 1. In other words, decoupling the two interactions together enlarges the available phase space. As a result, more independent samples are needed for $\langle F_\lambda \rangle$ to converge at these two end states. In addition, we note that the autocorrelation time in the fixed λ OSRW simulations at $\lambda = 0$ is $\sim$30 ps, much longer than in the BAR simulations when $\lambda = 0$. Based on eqn (6), the uncertainty of the fixed λ OSRW simulations with a total simulation time of 20 ns was estimated to be $\sim$1 kcal mol^{-1}. Although the dynamics are different from those of the OSRW simulations reported in Table 1, this result manifests that decoupling both interactions will create a rough energy landscape that makes sampling difficult. Therefore, the poor convergence of some of the OSRW simulations can be largely attributed to the decoupling scheme.

There is a positive correlation between the uncertainties of the OSRW simulations and the net charge of the system.

Table 6 Correlation between uncertainties of binding free energies and net charge for each system. RMSE is the root mean square difference between OSRW results and the reference BAR results

Guest	RMSE (kcal mol^{-1})	Charge
C5b	1.35	2
C4	1.34	2
C10	1.28	2
C1	1.05	2
C5	0.98	1
C6	0.90	1
C2	0.85	1
C13	0.80	1
C3	0.80	1
C14	0.68	1
C11	0.49	1
C12	0.46	1
C7	0.45	1
C8	0.41	1
C9	0.30	1

Since the uncertainty estimate for OSRW results is limited in precision by the small number of simulations, here we use the differences between OSRW and BAR results to measure the uncertainties. Except for guest 3, all the OSRW results for systems with charge +1 agree well with those of BAR results, whereas large differences can be found for systems with charge +2 (see Table 6). This further supports our finding that decoupling vdW and electrostatics interactions together hinders the sampling. We expect that the problem will be less prominent for neutral systems.

Conclusions

In this work, binding free energies of the SAMPL4 host–guest system CB[7] with 14 guest molecules were computed with both BAR and OSRW methods and the AMOEBA polarizable force field. Overall, the AMOEBA binding free energy values computed using both BAR and OSRW are in good agreement with experimental results. The binding thermodynamics of this series of host–guest systems varies from ligand to ligand. Some are driven by enthalpy changes while others by entropy gains. We further examined guest ligands C7, C8 and C10, which display high enthalpy or entropy changes upon binding. The enthalpy–entropy decomposition suggests that the binding of guest C10 is entropy driven, while binding of guests C7 and C8 has large enthalpic contributions. Hydrogen bonding analysis showed that guest C10 formed several hydrogen bonding interactions with both water and host CB[7], largely due to the three hydrophilic amine groups. Guests C7 and C8 gain additional H-bonds upon binding while C10 loses H-bonds upon binding, consistent with the enthalpy–entropy decomposition results. Configurational entropy was computed for guests C7, C8, C10 and their complexes with the host using quasiharmonic analysis. The configurational binding entropy was determined to be relatively small for all guests, hinting at the substantial role of water molecules. Through the analysis of intramolecular atomic fluctuations of guests C7 and C8, cyclic carbon atoms inside the host were found to fluctuate more for guests C8 than C7, intuitively a

result of the larger ring of C8. Unlike ligand–protein binding, the guest molecules were observed to freely rotate inside the host ring. Convergence of the BAR and OSRW free energy calculation methods was compared. The current OSRW implementation encounters convergence problems at the low end of vdW and electrostatic decoupling. Possible improvements can be achieved by separating the vdW and electrostatic decoupling, well-tempered metadynamics[61] and employing metadynamic alternatives.[43] Nonetheless, here, both BAR and OSRW methods are found to be adequate to determine the binding affinities for the model host–guest systems.

Acknowledgements

We appreciate support from the Robert A. Welch Foundation (Grant F-1691 to P. R.), the National Institutes of Health (Grants GM106137 and GM114237 to J. W. P. and P. R.) and the National Science Foundation (Grant CHE1152823 to J. W. P.). The high performance computing resources were provided by TACC and XSEDE (Grant TG-MCB100057 to P. R.).

References

1 E. Persch, O. Dumele and F. Diederich, *Angew. Chem., Int. Ed.*, 2015, **54**, 3290–3327.
2 H. Gohlke and G. Klebe, *Angew. Chem.*, 2002, **41**, 2644–2676.
3 K. N. Houk, A. G. Leach, S. P. Kim and X. Y. Zhang, *Angew. Chem., Int. Ed.*, 2003, **42**, 4872–4897.
4 M. K. Gilson and H. X. Zhou, *Annual Review of Biophysics and Biomolecular Structure*, Annual Reviews, Palo Alto, 2007, vol. 36, pp. 21–42.
5 D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, *Nat. Rev. Drug Discovery*, 2004, **3**, 935–949.
6 F. Grater, S. M. Schwarzl, A. Dejaegere, S. Fischer and J. C. Smith, *J. Phys. Chem. B*, 2005, **109**, 10474–10483.
7 P. Mikulskis, D. Cioloboc, M. Andrejic, S. Khare, J. Brorsson, S. Genheden, R. A. Mata, P. Soderhjelm and U. Ryde, *J. Comput.-Aided Mol. Des.*, 2014, **28**, 375–400.
8 V. M. Anisimov and C. N. Cavasotto, *J. Comput. Chem.*, 2011, **32**, 2254–2263.
9 M. Lawrenz, J. Wereszczynski, J. M. Ortiz-Sanchez, S. E. Nichols and J. A. McCammon, *J. Comput.-Aided Mol. Des.*, 2012, **26**, 569–576.
10 J. I. Monroe and M. R. Shirts, *J. Comput.-Aided Mol. Des.*, 2014, **28**, 401–415.
11 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
12 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. MacKerell, *J. Comput. Chem.*, 2010, **31**, 671–690.
13 M. J. Robertson, J. Tirado-Rives and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2015, **11**, 3499–3509.
14 M. M. Reif, P. H. Hünenberger and C. Oostenbrink, *J. Chem. Theory Comput.*, 2012, **8**, 3705–3723.
15 J. W. Ponder and D. A. Case, *Adv. Protein Chem.*, 2003, **66**, 27–85.
16 S. W. Rick and S. J. Stuart, in *Reviews in Computational Chemistry*, ed. K. B. Lipkowitz and D. B. Boyd, Wiley-VCH, Inc, New York, 2002, vol. 18, pp. 89–146.
17 D. E. Williams, *J. Comput. Chem.*, 1988, **9**, 745–763.
18 P. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, **107**, 5933–5947.
19 S. Patel and C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1–16.
20 C. M. Baker, V. M. Anisimov and A. D. MacKerell, *J. Phys. Chem. B*, 2011, **115**, 580–596.
21 J. A. Lemkul, J. Huang, B. Roux and A. D. MacKerell, *Chem. Rev.*, 2016, **116**, 4983–5013.
22 Y. Shi, Z. Xia, J. J. Zhang, R. Best, C. J. Wu, J. W. Ponder and P. Y. Ren, *J. Chem. Theory Comput.*, 2013, **9**, 4046–4063.
23 P. Ren, C. Wu and J. W. Ponder, *J. Chem. Theory Comput.*, 2011, **7**, 3143–3161.
24 P. Tiwary, V. Limongelli, M. Salvalaglio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, E386–E391.
25 J. C. Gumbart, B. Roux and C. Chipot, *J. Chem. Theory Comput.*, 2013, **9**, 794–802.
26 C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245–268.
27 C. D. Christ, A. E. Mark and W. F. van Gunsteren, *J. Comput. Chem.*, 2010, **31**, 1569–1582.
28 M. Z. Daniel, *Annu. Rev. Biophys.*, 2011, **40**, 41–62.
29 H. S. Muddana, C. D. Varnado, C. W. Bielawski, A. R. Urbach, L. Isaacs, M. T. Geballe and M. K. Gilson, *J. Comput.-Aided Mol. Des.*, 2012, **26**, 475–487.
30 H. S. Muddana, A. T. Fenley, D. L. Mobley and M. K. Gilson, *J. Comput.-Aided Mol. Des.*, 2014, **28**, 305–317.
31 E. Masson, X. X. Ling, R. Joseph, L. Kyeremeh-Mensah and X. Y. Lu, *RSC Adv.*, 2012, **2**, 1213–1247.
32 S. Walker, R. Oun, F. J. McInnes and N. J. Wheate, *Isr. J. Chem.*, 2011, **51**, 616–624.
33 J. W. Lee, S. Samal, N. Selvapalam, H. J. Kim and K. Kim, *Acc. Chem. Res.*, 2003, **36**, 621–630.
34 Y. J. Jeon, S. Y. Kim, Y. H. Ko, S. Sakamoto, K. Yamaguchi and K. Kim, *Org. Biomol. Chem.*, 2005, **3**, 2122–2125.
35 J. W. Ponder, C. J. Wu, P. Y. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon, *J. Phys. Chem. B*, 2010, **114**, 2549–2564.
36 A. D. Mackerell, *J. Comput. Chem.*, 2004, **25**, 1584–1604.
37 P. Cieplak, F. Y. Dupradeau, Y. Duan and J. M. Wang, *J. Phys.: Condens. Matter*, 2009, **21**, 333102.
38 P. E. M. Lopes, B. Roux and A. D. MacKerell, *Theor. Chem. Acc.*, 2009, **124**, 11–28.
39 C. Marquez and W. M. Nau, *Angew. Chem., Int. Ed.*, 2001, **40**, 4387–4390.
40 M. R. Shirts and J. D. Chodera, *J. Chem. Phys.*, 2008, **129**, 124105.
41 L. Zheng, M. Chen and W. Yang, *J. Chem. Phys.*, 2009, **130**, 234105.
42 L. Zheng, M. Chen and W. Yang, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 20227–20232.

 This journal is © the Owner Societies 2016

43 L. Zheng and W. Yang, *J. Chem. Theory Comput.*, 2012, **8**, 810–823.
44 D. Min, L. Zheng, W. Harris, M. Chen, C. Lv and W. Yang, *J. Chem. Theory Comput.*, 2010, **6**, 2253–2266.
45 D. A. Pearlman and P. A. Kollman, *J. Chem. Phys.*, 1989, **91**, 7831–7839.
46 X. J. Kong and C. L. Brooks, *J. Chem. Phys.*, 1996, **105**, 2414–2423.
47 J. R. Abella, S. Y. Cheng, Q. Wang, W. Yang and P. Ren, *J. Chem. Theory Comput.*, 2014, **10**, 2792–2801.
48 D. Jiao, P. A. Golubkov, T. A. Darden and P. Ren, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 6290–6295.
49 D. Hamelberg and J. A. McCammon, *J. Am. Chem. Soc.*, 2004, **126**, 7683–7689.
50 D. Jiao, J. J. Zhang, R. E. Duke, G. H. Li, M. J. Schnieders and P. Y. Ren, *J. Comput. Chem.*, 2009, **30**, 1701–1711.
51 G. J. Rocklin, D. L. Mobley, K. A. Dill and P. H. Hünenberger, *J. Chem. Phys.*, 2013, **139**, 184103.
52 M. Tuckerman, B. J. Berne and G. J. Martyna, *J. Chem. Phys.*, 1992, **97**, 1990–2001.
53 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
54 N. M. Henriksen, A. T. Fenley and M. K. Gilson, *J. Chem. Theory Comput.*, 2015, **11**, 4377–4394.
55 M. A. Wyczalkowski, A. Vitalis and R. V. Pappu, *J. Phys. Chem. B*, 2010, **114**, 8166–8180.
56 B. R. Brooks, D. Janezic and M. Karplus, *J. Comput. Chem.*, 1995, **16**, 1522–1542.
57 I. Andricioaei and M. Karplus, *J. Chem. Phys.*, 2001, **115**, 6289–6292.
58 D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollman, *AMBER 14*, University of California, San Francisco, 2014.
59 C. E. Chang, W. Chen and M. K. Gilson, *J. Chem. Theory Comput.*, 2005, **1**, 1017–1028.
60 R. Baron, P. H. Hünenberger and J. A. McCammon, *J. Chem. Theory Comput.*, 2009, **5**, 3150–3160.
61 J. Wereszczynski and J. A. McCammon, *Q. Rev. Biophys.*, 2012, **45**, 1–25.
62 A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 020603.