J Comput Aided Mol Des (2012) 26:569–576
DOI 10.1007/s10822-012-9542-5

Thermodynamic integration to predict host-guest binding affinities

Morgan Lawrenz · Jeff Wereszczynski · Juan Manuel Ortiz-Sánchez · Sara E. Nichols · J. Andrew McCammon

Received: 13 October 2011 / Accepted: 9 January 2012 / Published online: 16 February 2012

Abstract An alchemical free energy method with explicit solvent molecular dynamics simulations was applied as part of the blind prediction contest SAMPL3 to calculate binding free energies for seven guests to an acyclic cucurbit-[n]uril host. The predictions included determination of protonation states for both host and guests, docking pose generation, and binding free energy calculations using thermodynamic integration. We found a root mean square error (RMSE) of 3.6 kcal mol^{-1} from the reference experimental results, with an R^2 correlation of 0.51. The agreement with experiment for the largest contributor to this error, guest 6, is improved by 1.7 kcal mol^{-1} when a periodicity-induced free energy correction is applied. The corrections for the other ligands were significantly smaller, and altogether the RMSE was reduced by 0.4 kcal mol^{-1}. We link properties of the host-guest systems during simulation to errors in the computed free energies. Overall, we show that charged host-guest systems studied here, initialized in unconfirmed docking poses, present a challenge to accurate alchemical simulation methods.

Keywords Thermodynamic integration · Molecular dynamics · Docking · Host-guest · Blind prediction

M. Lawrenz (✉) · J. Wereszczynski · J. M. Ortiz-Sánchez · S. E. Nichols · J. A. McCammon
Department of Chemistry and Biochemistry, Center for Theoretical Biological Physics, University of California, San Diego, La Jolla, CA, USA
e-mail: mlawrenz@ucsd.edu

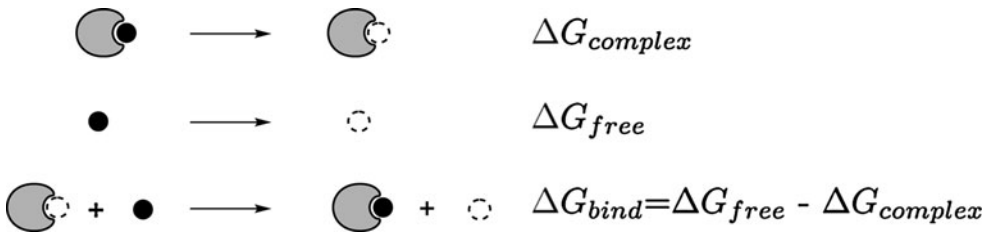
S. E. Nichols · J. A. McCammon
Department of Pharmacology, Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA, USA

Introduction

Calculation of accurate and precise ligand binding affinities has potential for significant impact on drug discovery, providing high fidelity guidance for molecular synthesis and experimental affinity assays. The varied available computational approaches for calculating affinities involve a balance of computational cost with both the detail of the model employed for the binding partners and the physical approximations applied to the calculation. In this study, a computationally intensive method was used with atomistic molecular dynamics (MD) simulations to compute binding free energies. While this approach is based on rigorous statistical mechanics principles [1–8], its practical application for estimation of free energies is still challenging for systems with many degrees of freedom.

To examine the predictive power of binding free energy calculation approaches, a simplified system of binding partners is useful for method comparison and identification of specific challenges. Blind predictions are key for eliminating bias for the investigators and emulating practical applications of these approaches. This study reports blind free energy predictions for host-guest systems, which exemplify protein-ligand association in the types of interactions formed and the applicable statistical mechanics [9], but with fewer degrees of freedom to sample.

The calculations reported here are alchemical *absolute* binding free energy calculations, which employ MD simulations of unphysical intermediates in order to compute the free energy change for the transfer of the ligand from the unbound to the bound state [10]. The binding free energy is computed using the well-established thermodynamic cycle (Scheme 1), in which the guest is transformed into a non-interacting molecule within both the host-bound and unbound solvated environments [11]. In this study,



Scheme 1 Thermodynamic cycle employed for alchemical free energy calculations

independent MD simulations are employed for each intermediate alchemical state, which has been shown to improve reliability in free energy results [12–14].

Simulation methods

Overview

In this work, we computed binding free energies for seven guests, provided by SAMPL3, to the acyclic cucurbit-[n]uril host *5a* from Ma et al. [15]. Binding free energies were computed with an alchemical technique, using multiple, independent explicit solvent MD simulations and an all-atom force field. We performed absolute binding free energy calculations, during which ligand non-bondeded interactions are slowly turned off over several unphysical, alchemical intermediate states. This process requires two separate calculations, in which the ligand is decoupled [4, 16] from both the host-bound and unbound environments, to respectively compute $\Delta G_{complex}$ and ΔG_{free} (Scheme 1). The difference of these results gives ΔG_{bind}, which is corrected for the standard state ΔG°_{bind}, as described below.

Molecular dynamics simulations

Initial coordinates for the acyclic cucurbit-[n]uril host, *5a* from Ma et al. [15] were available from X-ray crystallography experiments. The host titratable groups were treated at neutral pH (all carboxyl groups de-protonated). Ligands were prepared from their structure data files (SDF) made available through SAMPL3 in Maestro using the Schrödinger tool Ligprep [17]. All tertiary amine groups were protonated to be quaternary ammonium cations, giving all guests a +1 charge except for guest 6, at +2, and the neutral guest 4. Glide [18, 19], with the XP [20] scoring function, was used to dock ligands into the host. The top scoring docking conformation was used as a starting configuration for MD simulations.

After obtaining the docked host-guest complexes, molecular models were generated with tleap [21], using the Generalized Amber Force Field (GAFF) [22] for all bond, angle, and torsion parameters. Atomic partial charges were derived from RESP [23] fitting of Gaussian03 [24] calculated electrostatic potentials at the Hartree-Fock/6-31G* level. The compatible TIP3P [25] water model was used, with the triclinic simulation box generated to provide a 12 Å water buffer to the periodic boundary. All systems were neutralized with counter-ions, which have AMBER rescaled parameters [26].

All simulations were performed using the NAMD software package version 2.7b1 [27]. A 2 fs timestep was employed, with hydrogen-containing solute bonds constrained using RATTLE [28] and water geometries constrained using SETTLE [29]. The Particle Mesh Ewald (PME) approximation [30] with a maximum grid spacing of 1 Å was employed for electrostatics. Short-range non-bondeded interactions were evaluated every 2 fs and long-range electrostatics every 4 fs (non-bondeded interaction cutoff: 12 Å; switching distance: 10 Å) [27]. After incremental heating to 300 K at 1 atm, the system was equilibrated for 2 ns in the NPT ensemble with Langevin pressure and temperature controls [31]. Independent NPT simulations were initialized with random velocities for free energy calculations.

Free energy calculations

For free energy calculations, the SAMPL guests were alchemically decoupled from the surrounding environment with the coupling parameter λ. This parameter is bounded by 0 and 1, and linearly scales all ligand non-bondeded potential energy terms. Thermodynamic integration (TI) [32], as in Eq. 1, was used to compute the free energy of decoupling the guests from a freely solvated and the host-bound environments, giving ΔG_{free} and $\Delta G_{complex}$, respectively (Scheme 1).

$$\Delta G = \int_0^1 d\lambda \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda} \quad (1)$$

In Eq. 1, U is the total potential energy of the system, with non-bondeded terms scaled by λ. Following the traditional alchemical thermodynamic cycle, $\Delta G_{complex}$ is subtracted from ΔG_{free} to obtain ΔG_{bind} (Scheme 1).

TI was performed using the NAMD 2.7b1 free energy module, computing the $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ term from Eq. 1 as the average from multiple, independent simulations at each λ value. Such an approach has been shown in to improve the reliability of TI free energy predictions [33]. The employed simulation times, λ intermediates, and the number of independent simulations are summarized in Table 1. Electrostatics and van der Waals components of the ligands were decoupled in separate steps, with both inter- and intra-molecular non-bondeded potential terms scaled by λ. The soft-core potential by Zacharias et al. [34] was used to eliminate instabilities (shift parameter $\delta = 5$). For calculation of ΔG_{free}, all λ intermediate simulations were initialized sequentially, i.e. each simulation was initialized from the end configuration of the preceding λ simulation. To expedite wall-clock time for the more computationally expensive $\Delta G_{complex}$ calculations, each λ intermediate was initialized independently from an equilibrated starting structure and run in parallel. For the case of guest 1, which has two stereoisomers, free energies for both R and S enantiomers were computed and the mean of the two was submitted as the binding free energy of a racemic mixture. The binding free energy for guest 6 was computed by Boltzmann weighting the independent results for its three distinct stereoisomers, which are all equivalent at $\lambda = 1$ (Table 1).

The $\frac{\partial U}{\partial \lambda}$ values of Eq. 1 were printed for each λ every 0.1 ps and their forward cumulative average was monitored to evaluate convergence. Most $\frac{\partial U}{\partial \lambda}$ values reached local convergence around 500 ps (data not shown); thus, collection of data for $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ began after a 500 ps of equilibration period was discarded (Table 1). TI derivative data from all independent simulations was used to contribute to the ensemble average. Numerical integration of Eq. 1 was performed using an interpolated cubic spline, weighted by $\frac{1}{\sigma_\lambda}$, where σ_λ is the bootstrapped error computed on $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$. This error was computed after de-correlating the data at intervals of the statistical inefficiency g as described and coded by Chodera et al. [35] and 1,000 subsamples were used in the bootstrap method [36]. The σ_λ values were propagated in quadrature for an overall error on the free energy estimates s_{bind}, which are reported along with the free energy results in 2. λ values were chosen to span the range [0, 1], with more intermediates near the initial and final states. Additional points (particularly in the vdW region $\lambda > 0.9$) were added after inspection of the curves in order to reduce the integration error.

A Cartesian harmonic restraining potential $U(r_L) = \frac{1}{2}k_h(r_L - r_0)$ was applied to restrict ligand sampling r_L to a finite volume V_{pocket} within the active site throughout the TI calculations of $\Delta G_{complex}$. A k_h of 2 kcal mol^{-1} Å^{-2} was used for all ligands, as well as for three restrained host atoms that demonstrated minimal fluctuation from a free 2 ns NPT MD run. Then, the standard state binding free energy ΔG°_{bind} was computed through an analytical correction [4, 37, 38] to $\Delta G_{complex}$ for transferral of the ligand from the restricted volume V_{pocket} to the bulk V° as

$$\Delta G^\circ_{complex} = \int_0^1 d\lambda \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda + RT \ln\left(\frac{V_{pocket}}{V^\circ}\right) \tag{2}$$

To reflect host-guest binding at a standard ligand concentration of 1 M, values of $V^\circ = 1661$ Å^3 and $T = 300\ K$ were used. V_{pocket} was explicitly determined from multiple MD trajectories using the VMD VolMap plugin [39]. This procedure gave $RT \ln(\frac{V^{pocket}}{V^\circ})$ corrections that are listed in 2 and were typically about 3–4% of the total binding free energy change.

To correct for the periodicity-induced shift in the computed free energies, due to the change in overall system charge when decoupling a charged ligand, we applied the correction as described by Hunenberger et al. [40]:

$$\Delta G_{pbc} = \frac{q^2}{4\pi\varepsilon_0}\frac{1}{2L} \left\{ \varepsilon_s^{-1}\xi_{EW} - (\varepsilon_i^{-1} - \varepsilon_s^{-1}) \times \left(\frac{4\pi}{3}\left(\frac{R}{L}\right)^2 - \frac{16\pi^2}{45}\left(\frac{R}{L}\right)^5 \right) \right\} \tag{3}$$

Here ε_0 is vacuum permittivity, $\varepsilon_s = 97$ is the solvent relative permittivity for TIP3P water [25], $\varepsilon_i = 1$ is the relative solute internal permittivity, $\xi_{EW} = -2.837$ [40,

Table 1 Simulation setup for TI calculations

Calculation	Electro/vdW λ intermediates	Total λ	Equil/production t per λ (ns)	Independent runs per λ	Total t (ns)
$\Delta G_{complex}$	8[a]/14[b]	21	0.5/3.5	5	420
ΔG_{free}	9[c]/13[d]	21	0.5/3.5	3	252

[a] λ = [0, 0.1, 0.2, 0.4, 0.5, 0.6, 0.8, 1]

[b] λ = [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95, 0.97, 1]

[c] λ = [0, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]

[d] λ = [0, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95, 0.97, 0.99, 1]

41] is the self interaction potential for cubic periodic boundary lattice conditions, L is the length of the cube (we approximate the triclinic box as a cube, and use an L that gives an equivalent box volume), R is the host-guest complex radius of gyration, computed using the GROMACS [42] tool g_gyrate, and q is the charge of the system at the initial or final state of the free energy calculations. Hence, if the system undergoes a transition from 0 to -2 after decoupling a $+2$ charged ligand, Eq. 3 is evaluated with both $q = 0$ and $q = -2$ to compute the correction $\Delta\Delta G_{pbc}$ to be added to the TI free energy result.

Results and discussion

Docking poses

The docked poses of the seven guests are shown in Fig. 1. The poses can be described by two general motifs. Guests 2, 4, 6 and 7 are completely threaded through the host, with functional groups on each side of the host surface. Guests 1, 3, and 5 extend a substituent into the center of the host, with the rest of the molecule extending along the surface. In particular, guests 3 and 5 have unique binding poses, with just the ethyl-substituted ammonium group inside the host. In the crystal structure used for docking, the host carboxyl groups are oriented towards each other. However, during equilibration MD simulations, these groups immediately separated due to the opposing charges. The root mean square deviation (RMSD) of host atoms from the crystal docking pose was consistently $\approx$1.5 Å during equilibration. This change forced adjustment of the guest binding pose, with only guest 1R, 2, and 7 maintaining a consistent salt-bridge distance of <4 Å between its ammonium and a single host carboxyl group. All three of these guests have a docking pose that places the ammonium group near the surface of the host (Fig. 1).

Comparing SAMPL results with experiment

Results from TI free energy calculations for seven SAMPL ligands gave a root mean squared error (RMSE) of 3.6 kcal mol^{-1} from the reference experimental results, with an R^2 correlation of 0.51. The Kendall-Tau rank correlation coefficient was 0.52. Two lists the computed free energy components from Scheme 1, the values of the standard state correction in Eq. 2, the binding free energies and errors, and the difference from the experimentally derived value. A plot of the calculated standard state binding free energies (ΔG°_{bind}) with the experimental results (ΔG_{exp}) is shown in blue in Fig. 2, along with the ideal correlation line (black) and the 2 kcal mol^{-1} RMSE boundary (green). Red circles indicate data with an additional correction described in another section.

Guests 4 and 6 contribute most to the RMSE, with guest 6 at 6.0 kcal mol^{-1} from the experimental value, and guest 4 at 5.3 kcal mol^{-1} (Table 2). As described in Table 3, guest 6 is the largest and most flexible ligand, similar in size to its host binding partner, and has the largest charge

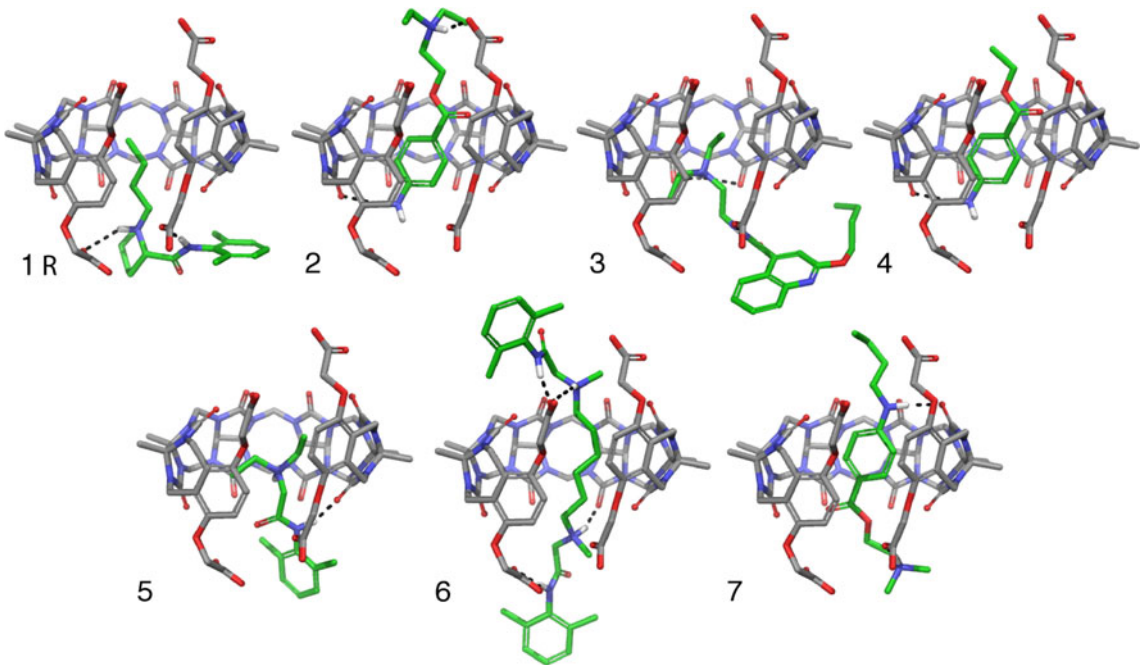


Fig. 1 Docking poses for seven host-guest systems. Guests are highlighted with *green*, and *dashed lines* indicate hydrogen bonding interactions

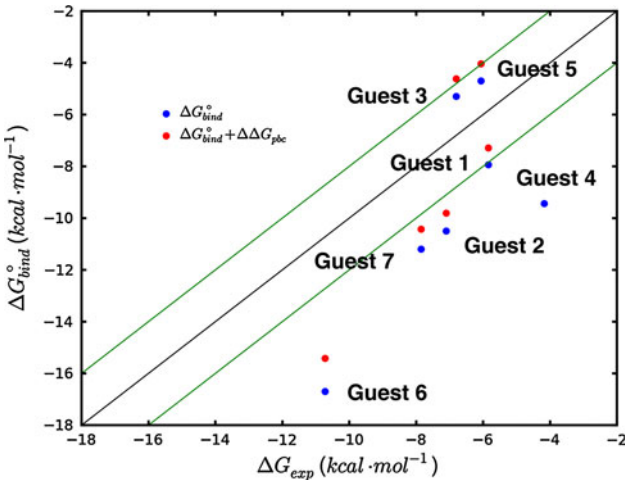


Fig. 2 Correlation of the TI predicted free energy results with experiments. ΔG°_{bind} (*blue*, 2) is plotted with the experimental value ΔG_{exp} (Table 2), as well as data with the correction in Eq. 3, $\Delta G^{\circ}_{bind} + \Delta\Delta G_{pbc}$ (*red*, Table 3). Free energies are labeled with their corresponding guest. Note that guest 4 has no corrected (*red*) result. *Lines* also depict absolute correlation (*black*) and the 2 kcal mol^{-1} RMSE boundary (*green*)

(+2); therefore this ligand may represent a greater challenge to free energy prediction, both for the accuracy of the force field description of host-guest interactions, and for sampling. Guest 4 is the smallest ligand, and also the only molecule parameterized as neutral at pH 7. The binding free energy for this particular ligand was consistently predicted to be overly favorable in submissions for the SAMPL3 blind prediction (Table 2).

Properties of host-guest complexes

Table 3 lists properties of the seven guests, along with the difference between calculated and experimental binding free energies. These differences are not correlated with molecular weight and number of rotatable bonds; the heaviest, largest guest 6 and the lightest, smallest guest 4 gave the worst predictions (Table 3). While five of the guests were predicted to be stronger binders than found by experiment, guests 3 and 5 were predicted to be weaker binders. These two ligands were docked in unique binding modes (Fig. 1) that changed significantly during equilibration, as seen with RMSD (Table 3). Guest 3 deviated 5.1 Å and guest 5 changed by 3.8 Å over 2 ns of simulation (Table 3). Guest 5 may be in a particularly unfavorable docking pose, as the molecule briefly became unbound near the end of the equilibration. These potentially unfavorable docking poses could explain the weak predicted binding free energies for these two ligands (Fig. 2 and Table 2). Guest 6 also deviated significantly from the docked pose, with an RMSD of 4.1 Å (Table 3), although this guest has the most rotatable bonds (Table 3), and the large RMSD may be mostly attributed to the flexibility of the molecule.

A comparison of the root mean square fluctuation (RMSF) of the guests throughout the alchemical simulations makes a qualitative link between the sampling and the difference between predicted and experimental free energies. The RMSF values computed over all simulations used in the TI calculation of $\Delta G_{complex}$ for each ligand is listed in

Table 3 Guest RMSF during host-bound TI calculations

Guest	Guest rotatable bonds	Guest MW (g mol^{-1})	Guest RMSD[a] (Å)	Guest RMSF[b] (Å)	$\Delta\Delta G_{calc-exp}$ (kcal mol^{-1})
1 S	8	289.4	3.1	1.4	−2.1
1 R	8	289.4	2.9	1.5	−2.1
2	10	237.3	1.9	1.6	−3.4
3	13	344.5	5.1	1.7	+1.5
4	5	165.2	2.7	0.6	−5.3
5	9	235.4	3.8	1.5	+1.4
6	19	468.7	4.1	2.7	−6.0
7	11	265.4	2.4	1.6	−3.3

[a] From docking pose

[b] During calculation of $\Delta G_{complex}$

Table 2 TI free energy results for SAMPL guests

Guest	ΔG_{free}	$\Delta G_{complex}$	$+RT \ln\left(\frac{V_{pocket}}{V^{\circ}}\right)$	$\Delta G^{\circ}_{bind} \pm s_{bind}$	Racemic	ΔG_{exp}	$\Delta\Delta G_{calc-exp}$
1 S	−8.8	−0.5	−0.3	−7.9 ± 1.1	−7.8 ± 1.7	−5.8[a]	−2.0
1 R	−2.6	5.3	−0.3	−7.6 ± 1.3	−7.8 ± 1.7	−5.8[a]	−2.0
2	−7.7	3.3	−0.4	−10.5 ± 1.8		−7.1	−3.4
3	−0.4	5.1	−0.2	−5.3 ± 1.3		−6.8	+1.5
4	8.4	18.5	−0.7	−9.4 ± 1.0		−4.2	−5.3
5	14.6	19.7	−0.4	−4.7 ± 1.2		−6.1	+1.4
6	33.5	50.2	0.1	−16.7 ± 1.9		−10.7	−6.0
7	−32.4	−20.9	−0.4	−11.2 ± 1.5		−7.9	−3.3

[a] Result for a racemic mixture (see "Simulation methods"). All results are listed in kcal mol^{-1}

3. Guest 6 demonstrated the most flexibility during the alchemical simulations, with an average RMSF (2.7 Å) that is two standard deviations higher than the average RMSF over all the ligands (1.6 Å). Guest 4 was the most rigid throughout the simulations, with an average RMSF (0.6 Å) that is almost two standard deviations below the average for all guests (1.6 Å). While these findings do not fully explain the differences in free energies, they do highlight unique sampling trends that set these two ligands apart from the others. The fluctuations of guest 6 could impede convergence of the free energy data to accurate values, while guest 4 could be stuck sampling local configurations (Fig. 2; Table 3).

Spline-fit curves for host-guest TI calculations

Plots of the $\frac{\partial U}{\partial \lambda}$ data with λ, and the fitted cubic splines used to compute $\Delta G_{complex}$, are shown in 3; the curves corresponding to calculation of ΔG_{free} are very similar in functional form (not shown). Inspection of 3 reveals a significant difference in both the electrostatic and van der Waals data for guest 6 (pink), as the potential derivatives for this ligand span a much larger range of values along the alchemical pathway than do the other ligands. In the van der Waals curve for guest 6 (pink, 3b), the ligand potential derivative increases by 350 kcal mol^{-1} over the last three steps between $\lambda = 0.9$ and 1; the λ value 0.95 was added to smooth out the fitted curve in this region. In addition to integration error, force field issues behind this large jump in energy, observed for all independent simulations, could contribute to the inaccuracy of the free energy computed for this guest.

Correction to alchemical free energies computed for charged ligands

After submission of the SAMPL3 blind prediction results and discussion with contributors, we decided to evaluate the effect of the correction in Eq. 3 on our results. During alchemical simulations of charged ligands, the system charge, neutral at $\lambda = 0$, was changed after the ligand electrostatics were turned off, and a periodicity-induced shift in the free energy of binding can occur [40, 41]. As all the SAMPL guests except for guest 4 were charged, we computed the correction (see "Simulation methods") in hopes that the free energies, particularly for the +2 charged guest 6, would improve. The results with this additional correction and experimental comparison are listed in 4. As expected from Eq. 3, with $q = 2$, guest 6 was most affected by the correction, shifting the binding free energy result by 1.7 kcal mol^{-1} closer to the experimental value. Overall, the correction reduced the RMSE by 0.4 kcal mol^{-1}, but the correlation coefficient is also reduced to 0.42 (compare blue and red data in 2). The corrected predictions for the guest 3 and 5 were even less favorable than the uncorrected free energies (Fig. 3; Table 4).

Because of its neutral charge, guest 4 is unaffected by the additional periodicity correction. However, the pKa of the aniline group of guest 4 was questioned by the SAMPL contributors; the ionic environment of the host cation binding motif (four carboxylic groups) may shift the pKa of the guest aniline group. A systematic error for this molecule due to this neutral parameterization could explain the prevalence of overly favorable estimates computed in SAMPL3.

Additional potential sources of error

The highly charged nature of the host-guest binding in this study presents a challenge to the molecular models used in

Table 4 TI free energy results with periodicity correction

Guest	$\Delta G_{bind} + \Delta\Delta G_{pbc}$	ΔG_{exp}	$\lvert\Delta\Delta G_{calc2-exp}\rvert$
1 (racemic)	−7.3	−5.8	−1.5
2	−9.8	−7.1	−2.7
3	−4.6	−6.8	+2.2
4	−9.4	−4.2	−5.3
5	−4.0	−6.06	+2.0
6	−15.0	−10.7	−4.3
7	−10.3	−7.9	2-.5

All results are in kcal mol^{-1}

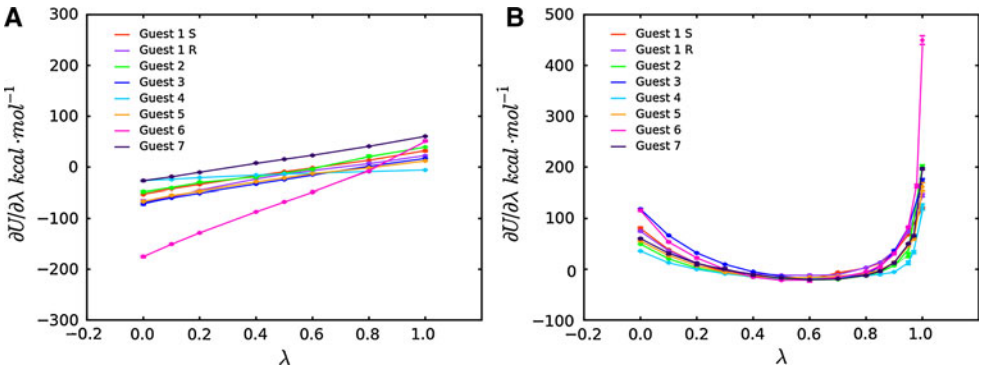


Fig. 3 Cubic spline fit of $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ with λ for ligand non-bonded potential components. The $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for electrostatic (**a**) and van der Waals (**b**) ligand potential components are plotted along with the *error bars* σ_λ (see "Simulation methods"). The integrands of these curves are summed to give $\Delta G_{complex}$

classical MD simulations. The host itself contains a cluster of 4 carboxyl groups, which interact with the cationic ammonium groups on its guests. The protonation states of these carboxyl groups were not clear, and in solution the states are likely dynamically exchanging. Thus, parameterization of the host as fully de-protonated, for a -4 charge, was an initial approximation that could affect all results. As described in a previous section, the host carboxyl group conformations change significantly from the crystal structure during equilibration (see above), which forces adjustment of the docked poses. TI calculations are dependent upon the accuracy of the docking pose used as a starting structure for the simulations. Equilibration gives limited relaxation time for the pose, and during the alchemical simulations the guest center of mass is restrained (see "Simulation methods"). If the docking pose is a metastable state with a significant barrier to reorientation, the calculations will be not be accurate. This could be a significant contributor to error in all host-guest cases, particularly guest 4, which remains largely rigid throughout the calculation (Table 3). Validation of docking pose predictions with crystallographic poses would be useful for targeting sources of error.

Conclusions

Binding free energies for the cucurbit-[n]uril host and seven ligands were predicted using alchemical MD simulations in explicit water and TI. The RMSE was 3.6 kcal mol^{-1} from the reference experimental results, with an R^2 correlation of 0.51. Two guests contributed most to these errors: the smallest, and only neutral guest 4 and the largest, only $+2$ charged guest 6. We computed a periodicity-induced correction for alchemical simulations of charged ligands, which improved the binding free energy estimate for guest 6 by 1.7 kcal mol^{-1}. The corrections for the $+1$ ligands were significantly smaller, and altogether the RMSE was reduced by 0.4 kcal mol^{-1}. Through RMSF calculations, we observed guest 4 to be significantly more rigid than other ligands; this ligand could be trapped in an incorrect binding pose that gives an overly favorable free energy estimate.

Overall, we propose that parameterization and docking poses are significant contributors to error in the free energy predictions. Fixed partial charge assignment for MD simulations of the highly charged host and guests may be inaccurate in this case, where proton exchange likely occurs in solution. We propose that such a system could be an ideal target for constant pH calculations, which allow dynamic exchange of protonation states. Validation of the docking poses can reveal additional sources of error, as the starting structure heavily influences the dynamics of the alchemical simulations and accuracy of the results.

Acknowledgments The authors thank the members of the McCammon group for useful discussions, and Gabriel Rocklin from the Shoichet/Dill group for help in determining the periodicity-induced free energy correction. This work was supported, in part, by the National Institutes of Health, the National Science Foundation, the National Biomedical Computational Resource, and the Howard Hughes Medical Institute. We thank the Center for Theoretical Biological Physics (NSF Grant PHY-0822283), and the Texas Advanced Computer Center (grant TG-MCA93S013) for distributed computing resources. J. M. Ortiz-Sánchez acknowledges the Fulbright Commission/Generalitat de Catalunya Program and the Generalitat de Catalunya for a Fulbright and a Beatriu de Pinos postdoctoral grants, respectively. J. Wereszczynski acknowledges support by Award Number F32GM093581 from the National Institute of General Medical Sciences. The project content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of General Medical Sciences or the National Institutes of Health.

References

1. Beveridge D, DiCapua F (1989) Free energy via molecular simulation: application to chemical and biomolecular systems. Annu Rev Biophys Chem L18:431–492
2. Kollman P (1993) Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev L93:2395–2417
3. van Gunsteren WF, Beutler TC, Fraternali F, King PM, Mark AE, Smith PE (1993) Computation of free energy in practice: choice of approximations and accuracy limiting factors, vol 2. ESCOM Science Publishers, Leiden
4. Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J L72:1047–1069
5. Jorgensen W (2004) The many roles of computation in drug discovery. Science L303:1813–1818
6. Gilson MK, Zhou H-X (2007) Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct L36:21–42
7. Pohorille A, Jarzynski C, Chipot C (2010) Good practices in free-energy calculations. J Phys Chem B L114:10235–10253
8. Christ CD, Mark AE, Gunsteren WFV (2010) Basic ingredients of free energy calculations: a review. J Comput Chem L31:1569–1582
9. Moghaddam S, Inoue Y, Gilson MK (2009) Host-guest complexes with protein-ligand-like affinities: computational analysis and design. J Am Chem Soc L131:4012–4021
10. Kirkwood JG (1935) Statistical mechanics of fluid mixtures. J Chem Phys L3:300–313
11. Tembe B, McCammon J (1984) Ligand receptor interactions. J Comput Chem L8:281–283
12. Fujitani H, Tanida Y, Ito M, Jayachandran G, Snow CD, Shirts MR, Sorin EJ, Pande VS (2005) Direct calculation of the binding free energies of FKBP ligands. J Chem Phys L123:084108
13. Zagrovic B, van Gunsteren W (2007) Computational analysis of the mechanism and thermodynamics of inhibition of phosphodiesterase 5A by synthetic ligands. J Chem Theory Comput L3:301–311
14. Lawrenz M, Baron R, McCammon J (2009) Independent-trajectories thermodynamic-integration free-energy changes for biomolecular systems: determinants of H5N1 avian influenza virus

neuraminidase inhibition by peramivir. J Chem Theory Comput L5:1106–1116
15. Ma D, Zavalij PY, Isaacs L (2010) Acyclic cucurbit[n]uril congeners are high affinity hosts. J Org Chem L75:4786–4795
16. Hamelberg D, McCammon JA (2004) Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method. J Am Chem Soc L126: 7683–7689
17. LigPrep version 2.3; Schrodinger LLC: New York, NY, (2009)
18. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem L47:1739–1749
19. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem L47:1750–1759
20. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem L49:6177–6196
21. Case DA et al (2010) AMBER 11. University of California, San Francisco
22. Wang J, Wolf RM, Caldwell JW, Case PAKDA (2004) Development and testing of a general amber force field. J Comput Chem L25:1157–1174
23. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc L117:5179–5197
24. Frisch M et al (2004) Gaussian 03, Revision C02:2003. Gaussian Inc., Wallingford
25. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys L79:926–935
26. Åqvist J (1990) Ion-water interaction potentials derived from free energy perturbation simulations. J Phys Chem L94:8021–8024
27. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem L26:1781–1802
28. Andersen H (1983) Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations. J Comput Phys L52:24–34
29. Shuichi M, Peter A (1992) SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem L13:952–962, 148324
30. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N [center-dot] log(N) method for Ewald sums in large systems. J Chem Phys L98:10089–10092
31. Feller S, Zhang Y, Pastor R, Brooks B (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. J Chem Phys L103:4613–4621
32. Kirkwood J (1935) Statistical mechanics of fluid mixtures. J Chem Phys L3:300–313
33. Lawrenz M, Baron R, Wang Y, McCammon JA (2011) Effects of biomolecular flexibility on alchemical calculations of absolute binding free energies. J Chem Theory Comput L7:2224–2232
34. Zacharias M, Straatsma TP, McCammon JA (1994) Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. J Chem Phys L100:9025–9031
35. Chodera J, Swope W, Pitera J, Seok C, Dill K (2007) Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. J Chem Theory Comput L3:26–41. doi:10.1021/ct0502864
36. Carlstein E (1986) The use of subseries values for estimating the variance of a general statistic from a stationary sequence. Ann Stat L14:1171–1179
37. Boresch S, Tettinger F, Leitgeb M, Karplus M (2003) Absolute binding free energies: a quantitative approach for their calculation. J Phys Chem B L107:9535–9551
38. General IJ (2010) A note on the standard state's binding free energy. J Chem Theory Comput L6:2520–2524
39. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graphics L14:33–38, 27–28
40. Hünenberger PH, McCammon JA (1999) Ewald artifacts in computer simulations of ionic solvation and ion–ion interaction: a continuum electrostatics study. J Chem Phys L110:1856
41. Hummer G, Pratt LR, García AE (1996) Free energy of ionic hydration. J Phys Chem L100:1206–1215
42. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput L4:435–447