Toward Improved Force-Field Accuracy through Sensitivity Analysis of Host-Guest Binding Thermodynamics

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JULY 2015

Impact Factor: 3.3 · DOI: 10.1021/acs.jpcb.5b04262 · Source: PubMed

4 AUTHORS, INCLUDING:

Andrew T Fenley

University of California, San Diego

35 PUBLICATIONS **343** CITATIONS

SEE PROFILE

Niel M Henriksen

University of California, San Diego

12 PUBLICATIONS **99** CITATIONS

SEE PROFILE

Michael K Gilson

University of California, San Diego

149 PUBLICATIONS **11,957** CITATIONS

SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

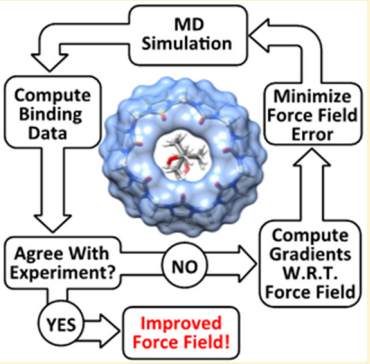
Toward Improved Force-Field Accuracy through Sensitivity Analysis of Host-Guest Binding Thermodynamics

Jian Yin,[†] Andrew T. Fenley,[†] Niel M. Henriksen, and Michael K. Gilson*

Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093-0736, United States

Supporting Information

ABSTRACT: Improving the capability of atomistic computer models to predict the thermodynamics of noncovalent binding is critical for successful structure-based drug design, and the accuracy of such calculations remains limited by nonoptimal force field parameters. Ideally, one would incorporate protein-ligand affinity data into force field parametrization, but this would be inefficient and costly. We now demonstrate that sensitivity analysis can be used to efficiently tune Lennard-Jones parameters of aqueous host-guest systems for increasingly accurate calculations of binding enthalpy. These results highlight the promise of a comprehensive use of calorimetric host-guest binding data, along with existing validation data sets, to improve force field parameters for the simulation of noncovalent binding, with the ultimate goal of making protein-ligand modeling more accurate and hence speeding drug discovery.



1. INTRODUCTION

The ability to reliably predict protein-ligand binding thermodynamics by means of molecular simulations would have enormous practical impact, such as the acceleration of drug discovery and enzyme engineering. Improving reliability is likely to require advances in two areas. One is efficient sampling, so as to obtain well-converged simulation results which correctly reflect the contributions of all thermodynamically relevant sectors of configuration space. Progress on this front includes advances in both algorithms[1−14] and computer hardware.[15−17] For example, microsecond-scale molecular dynamics (MD) simulations of biomolecular systems are now routinely achievable with commodity hardware.[18] However, a molecular simulation is only as accurate as the force field it uses, and, despite pioneering contributions[19−26] and important advances,[27−33] further improvement in force field accuracy are needed for reliable modeling of protein-ligand binding to become a reality.

Every force field includes a large set of adjustable parameters. These are typically set based on quantum chemistry data, such as gas-phase electrostatic potentials and the energetics of gas-phase clusters,[25,34−38] combined with selected experimental data. Enormous progress in force field development has already been made by using comparatively accessible experimental quantities, such as densities and heats of vaporization of neat liquids,[39−41] hydration free energies of small molecules,[42] and, more recently, conformational preferences of peptides and proteins.[43−45] However, these data sets are quite limited in size, and are scarcely expanding. For example, although a recently compiled set of ∼500 small molecule hydration free energies[46] is a powerful aid to testing and adjusting force fields, there is little prospect for increasing the number of such data. In addition, the commonly used data probe only a modest collection of interaction types, and this limitation risks compromising the generality of the resulting force fields. For example, a force field adjusted to replicate the properties of neat acetone and neat benzene may not accurately account for interactions between acetone and benzene; and hydration free energies only probe interactions of small organic molecules with one other molecule, water. Thus, it is perhaps not surprising that force fields do not routinely perform well when used to compute the properties of various chemical mixtures.[47,48] Nonetheless, such data are relevant to protein-ligand binding, because of the large variety of interactions that occur at the binding interface.

In addition, commonly used experimental observables may not strongly test the performance of force fields when they are applied to binding calculations. For example, although the TIP3P[49] and TIP4P-Ew[50] water models yield generally similar small molecule hydration free energies and enthalpies,[46,51] they yield strikingly different results for host-guest binding enthalpies[53] with mean signed errors (MSE), relative to experiment, of −3.0 kcal/mol for TIP3P and −6.8 kcal/mol for TIP4P-Ew. The magnitude and systematic character of these deviations could result from errors in the force field's representation of specific interactions present in the host-guest systems, perhaps amplified by the greater size of these host-guest systems relative to the small molecules in the hydration study. It is also worth noting, however, that neither small molecule hydration data nor the properties of neat liquids

Received: May 4, 2015
Revised: July 15, 2015

DOI: 10.1021/acs.jpcb.5b04262
J. Phys. Chem. B XXXX, XXX, XXX−XXX

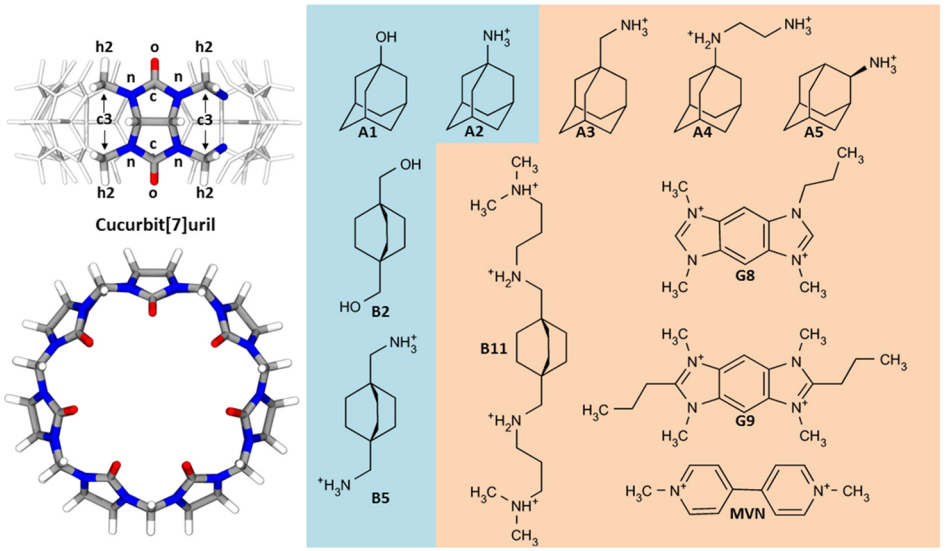


Figure 1. Cucurbit[7]uril (CB7) host and 11 guest molecules studied here. The guest molecules in the training set are shaded in blue and those in the test set are shaded in red.

probe how accurately water models treat confined water, which is present in the binding sites of host molecules and proteins, and is thought to significantly influence binding thermodynamics.[54−59]

Ideally, perhaps, one would use actual protein-ligand data to test and adjust force fields. Unfortunately, the calculation of rigorously converged absolute protein-ligand binding affinities by simulation is still too time-consuming to be incorporated into force-field optimization procedures. Moreover, protein simulations pose the challenge of establishing the protonation states of ionizable groups, such as histidine, aspartic acid, and glutamic acid, in complex, partially hydrated actives sites that can generate substantial pK_a shifts. The protonation states of such groups influence ligand affinities, but are not easily determined.

Host-guest systems hold great promise as a simple but informative alternative for testing and improving force fields for use in binding calculations. These miniature models of molecular recognition have affinities which span the same range as protein-ligand systems, and their small sizes, chemical simplicity, and often predictable protonation states make it far easier to carry out rigorously converged simulations which can be directly compared with experiment. Accordingly, host-guest systems have already been used to test binding calculations,[60−63] most recently through the SAMPL series of blind prediction challenges,[64−66] and also to optimize interaction potentials.[67] Both binding free energies and enthalpies are often available for host-guest systems,[62,68,69] and it has been shown that both of these quantities can be computed to high numerical precision.[53,65,66,70] Given that binding free energies and binding enthalpies are largely independent quantities,[61,71] it should be useful to tune force field parameters against both of these observables, thus taking maximal advantage of the available thermodynamic data. It is also worth noting that enthalpy changes, in the form of heats of vaporization of neat liquids, have long played a central role in force field parametrization,[72−74] so it is likely binding enthalpies will be similarly useful. For these reasons, we aim to add host-guest binding free energies and enthalpies to the types of experimental and quantum chemical data already used to optimize force field parameters.

To make efficient use of an experimental observable to adjust force field parameters, one needs a way of predicting how a prospective parameter change will affect the computed value of the observable. A natural approach to this problem is sensitivity analysis,[75] i.e., the evaluation of the partial derivatives of a simulation average with respect to simulation parameters. These derivatives provide the gradient of the computed quantity in parameter space, and hence can be used to choose parameter changes which will improve the agreement of the calculation experiment. Accordingly, partial derivatives of experimental observables or other target quantities have been incorporated into various parameter-optimization schemes.[20,50,76−80] However, we are not aware of prior efforts to use the gradients of host-guest binding thermodynamics with respect to force field parameters as a basis for force field optimization. Indeed, only recently has it been demonstrated that host-guest binding enthalpies can be computed with simulations to the level of numerical precision required for such an application.

The present study aims to prove the feasibility of using sensitivity analysis of host-guest binding enthalpies to help guide force field optimization. In particular, we use sensitivity analysis to adjust selected force field parameters to bring computed binding enthalpies for a series of cucurbit[7]uril-(CB7)-guest systems[62,81−84] into line with experiment. We focus on tuning Lennard-Jones (LJ) parameters, in part because prior energy decompositions of the computed binding enthalpies for these systems indicate that the LJ component of the force field makes dominant, favorable contributions to the computed binding enthalpies,[53] so small changes in LJ parameters should be effective force-field modifications to improve agreement with experiment. It should be noted that our main goal here is not to arrive at a specific set of recommended parameters for general use, but rather to implement and demonstrate a set of calculations which set the stage for incorporation of binding data into the broader process of force field optimization. We envision ultimately using binding data in concert with the more standard

DOI: 10.1021/acs.jpcb.5b04262
J. Phys. Chem. B XXXX, XXX, XXX−XXX

experimental and quantum chemical data sets discussed above, in order to generate force fields which work well not only for pure liquids and small molecules, but also for the noncovalent association of complex molecules in solution.

The article is organized as follows: The Methods section first describes the calculation of binding free energies and enthalpies, provides expressions for the partial derivatives of these quantities with respect to force field parameters, and explains how these were used to adjust parameters. It then provides details of the simulations and explains how statistical uncertainties were evaluated; additional methodological details are provided in the Supporting Information (SI). The Results section compares binding enthalpies calculated with the unmodified generalized Amber force field (GAFF)[32] and TIP3P water against experiment, then presents the sensitivity analysis and demonstrates its use to carry out two cycles of parameter optimization for two atoms associated with the CB7 host, based on a training set of four aliphatic guests. The consequences for the computed binding enthalpies and free energies of the training set and of a test set of seven additional guests are then presented. Finally, the Discussion section considers these particular results and looks more broadly at the implications of adding host-guest binding data to the set of data used to parametrize simulation force fields.

2. METHODS

We computed binding enthalpies and free energies for a training set of host-guest systems comprising the host CB7 and four aliphatic guests bearing either hydroxyl or ammonium groups (Figure 1, shaded in blue); applied sensitivity analysis to obtain the thermodynamic derivatives of the binding enthalpies with respect to the parameters of the Lennard-Jones force field terms; used these derivatives to guide parameter adjustments that would improve the agreement of the calculations with experiment; and reran the binding calculations to test the performance of the parameter adjustments. Two cycles of this process were carried out on the same training set, and the resulting set of parameters then was tested on a separate test set of seven guests (Figure 1, shaded in orange). Four of the test set guests are aliphatics with ammonium groups, and hence resemble the training set, while three of the test guest guests are cationic heteroaromatic compounds with little resemblance to the training set. These were included to probe the transferability of the parameter adjustments. The following subsections describe these procedures, and further details are available in the SI.

2.1. Calculation of Binding Free Energy and Enthalpy.

As previously described,[53] and detailed in the SI, binding enthalpies were computed as differences between the mean potential energy of a simulation of the solvated host-guest complex and the potential energies of two separate simulations of the solvated isolated host and the solvated isolated guest. As needed, the mean potential energy of a simulation of pure solvent was used to exactly balance the constituents of the bond versus free states; see eq 1.6 in the SI.

Host-guest binding free energies were computed via the attach-pull-release (APR) approach.[85] In brief, the APR approach involves computing a series of potentials of mean force along a path which begins with the guest unrestrained in the host's binding site, and ends with the guest held far enough from the host that their interactions are negligible. The pathway joining the initial and final states comprises an initial "attach" step, in which a system of restraints is imposed on the host-guest complex; a "pull" step, in which the restraint defining the host-guest distance is elongated to move the guest from the host; and a "release" step, in which all restraints on the host and guest that affect their internal coordinates are turned off. For the present calculations, the exit of the guest from the host is facilitated by restraints that hold open the otherwise tightly restrictive portal of the CB7 host. These restraints are turned on and off during the attachment and release steps, respectively. A final analytic correction adjusts the effective concentration of the guest to the standard concentration of 1 M. All PMFs were computed via thermodynamic integration,[86,87] using umbrella sampling[88] in a series of windows along the pathway. The SI provides further methodological details.

2.2. Sensitivity Analysis of Binding Thermodynamics.

Analyzing the sensitivity of binding thermodynamics to force field parameters requires expressions for the derivatives of host-guest binding enthalpies and free energies with respect to the chosen force field parameter. We focused on the LJ energy term, given by

$$U_{ij}^{\mathrm{LJ}} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \text{ with mixing rules } \epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \text{ and } \sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \tag{1}$$

where r_{ij} is the interatomic distance and ϵ_i and σ_i, respectively, relate to the well depth at the energy minimum and the distance at which the potential crosses zero. Note that the parameters are listed in GAFF as ϵ_i and $(1/2)R_{\mathrm{min}}$, where, $\sigma_i = 2^{-1/6}R_{\mathrm{min}}$ and R_{min} is the distance at the Lennard-Jones minimum. As noted above, the binding enthalpy of a host-guest system may be computed as the difference in the mean potential energy, $\langle U \rangle$, between the host-guest complex, hg; the free host, h; and the free guest, g; each in aqueous solution (The mean energy of a pure solvent simulation may also be required to balance the number of solvent molecules in the complex versus free simulations).[53] The sensitivity of the binding enthalpy to a force field parameter c may be obtained by recognizing that the partial derivative of $\langle U \rangle$ is given by (see SI)

$$\frac{\partial \langle U \rangle}{\partial c} = \left\langle \frac{\partial U}{\partial c} \right\rangle - \frac{1}{RT}\left(\left\langle U\frac{\partial U}{\partial c}\right\rangle - \langle U \rangle \left\langle \frac{\partial U}{\partial c}\right\rangle\right) \tag{2}$$

where R is the gas constant and T is absolute temperature, and c here will represent either ϵ_i or σ_i in eq 1. The sensitivity of the binding enthalpy is computed as the difference of $\partial\langle U \rangle/\partial c$ between the bound and unbound states:

$$\frac{\partial \Delta H}{\partial c} = \frac{\partial \langle U_{hg,N_{hg}} \rangle}{\partial c} - \frac{\partial \langle U_{h,N_h} \rangle}{\partial c} - \frac{\partial \langle U_{g,N_g} \rangle}{\partial c} - \frac{\partial \langle U_{\Delta N} \rangle}{\partial c} \tag{3}$$

Here N_{hg}, N_h, and N_g represent the number of solvent molecules in the hg, h, and g simulations respectively, and $\Delta N = N_{hg} - N_h - N_g$. As shown in the SI, the sensitivity of the binding free energy, ΔG, to parameter c is given by

$$\frac{\partial \Delta G}{\partial c} = \left\langle \frac{\partial U_{hg,N_{hg}}}{\partial c} \right\rangle - \left\langle \frac{\partial U_{h,N_h}}{\partial c} \right\rangle - \left\langle \frac{\partial U_{g,N_g}}{\partial c} \right\rangle - \left\langle \frac{\partial U_{\Delta N}}{\partial c} \right\rangle \tag{4}$$

DOI: 10.1021/acs.jpcb.5b04262
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 1. Experimental and Calculated Binding Enthalpies and Free Energies (kcal/mol) for Host CB7 with 11 Different Guest Molecules[a]

	binding enthalpy			binding free energy		
	exp.[b]	initial	final	exp.[b]	initial	final
A1	**−19.0 ± 0.2**	**−24.9 ± 0.2**	**−20.2 ± 0.2**	−14.1 ± 0.1	−23.7 ± 0.1	−22.6 ± 0.2
A2	**−19.3 ± 0.2**	**−22.7 ± 0.2**	**−19.0 ± 0.2**	−19.4 ± 0.1	−27.4 ± 0.1	−27.8 ± 0.2
B2	**−15.8 ± 0.1**	**−21.9 ± 0.2**	**−16.9 ± 0.2**	−13.4 ± 0.1	−21.3 ± 0.1	−19.8 ± 0.2
B5	**−15.6 ± 0.2**	**−18.3 ± 0.2**	**−14.4 ± 0.2**	−19.5 ± 0.1	−27.1 ± 0.2	−26.6 ± 0.2
A3	−21.9 ± 0.2	−24.8 ± 0.2	−19.6 ± 0.2	−20.3[c]	−28.0 ± 0.1	−26.6 ± 0.2
A4	−20.4 ± 0.2	−23.6 ± 0.2	−18.5 ± 0.3	−21.5[c]	−30.0 ± 0.2	−28.5 ± 0.2
A5	−19.5 ± 0.2	−23.2 ± 0.3	−18.8 ± 0.2	−19.1 ± 0.1	−27.3 ± 0.2	−27.1 ± 0.2
B11	−16.3 ± 0.2	−17.8 ± 0.2	−16.3 ± 0.2	−20.6 ± 0.2	−30.4 ± 0.4	−31.5 ± 0.6
G8	−8.5 ± 0.3	−6.2 ± 0.2	−0.8 ± 0.2	−6.1 ± 0.1	−13.9 ± 0.1	−13.4 ± 0.2
G9	−3.8 ± 0.1	−11.6 ± 0.2	−7.9 ± 0.2	−7.4 ± 0.1	−18.5 ± 0.3	−19.6 ± 0.3
MVN/Tris[c]	−3.4 ± 0.2	−2.1 ± 0.2	−2.6 ± 0.2	−7.1 ± 0.1	−11.1 ± 0.3	−9.7 ± 0.3

[a]Training set data are highlighted with bold font. Computational results are reported for both unmodified GAFF parameters (initial) and for the adjusted parameters from the second round of optimization (final). Uncertainties are reported as standard deviations of the mean. For methylviologen (MVN) both the measurements and calculations included Tris buffer, which influences the binding enthalpy (see ref 53 and Kimoon Kim, personal communication). Initial parameters: σ_n, ϵ_n = 3.2500, 0.1700; σ_o, ϵ_o = 2.9599, 0.2100. Final parameters: σ_n, ϵ_n = 3.1511, 0.1405; σ_o, ϵ_o = 3.2995, 0.3108. [b]See refs 62, 65, 83. [c]Experimental uncertainties were not reported for the binding free energies of A3 and A4.

The mean partial derivatives in these expressions were evaluated with custom-developed code which used analytic expressions for the derivatives of the LJ expression in eq 1 to postanalyze stored simulation trajectories.

2.3. Parameter Optimization. After the enthalpy derivatives of the training set binding enthalpies had been computed, several selected LJ parameters for atom types associated with the CB7 host were optimized simultaneously. As detailed in Results, we chose to optimize parameters with particularly large values of $\partial\Delta H/\partial c$, in order to minimize the need for large parameter adjustments. Optimization involved minimizing the squared deviation between the computed and experimental ΔH values for all CB7-guest pairs in the training set, using a nonlinear minimizer within the "Solver" tool of Microsoft Excel. A weak L2 regularization term, restraining the parameters toward their original GAFF values (see below), was included, to avoid implausibly large parameter modifications. The regularization term was weighted so that a 50% change in any parameter would generate a penalty as large as that associated with the deviation between the full set of initially computed binding enthalpies and their corresponding experimental values. To avoid overfitting, and to test for transferability of the revised host parameters, guest molecules A1, A2, B2, and B5 (Figure 1) were used as a training set for the parameter adjustments, while the remaining guests formed a test set.

2.4. Simulation Setup and Execution. All the MD simulations were performed with the AMBER14[89] suite of programs. Topology files for the systems were generated using tleap from AmberTools.[89] All amines were treated as fully protonated, because such simple aliphatic amines have p*K*a values larger than 9 in water, and these groups remain well-hydrated when bound to CB7, so their pKas are not expected to fall significantly in the bound state; and the experiments were done at neutral pH. The bonded and initial Lennard-Jones parameters were obtained from GAFF. Partial atomic charges were generated with the AM1-BCC method[90,91] as implemented in the program Antechamber.[92] The edge lengths of the cubic TIP3P water boxes ranged between 32 and 34 Å. Counterions were added to neutralize the total charge of each given system. The molecular dynamics simulations followed the same protocol as our previous study,[53] except that, instead of running the production simulations for 1 μs and saving snapshots every 2 ps, we ran the production simulations for 750 ns and saved snapshots every 1 ps. This change was implemented because we discovered, via statistical inefficiency analysis,[93] that the two approaches provide effectively the same statistical power for the standard error of the mean of the total potential energy of a simulation. Each system was initially energy minimized with 500 steps of steepest descent, followed by up to 10 000 steps of conjugate gradient with the host and/or guests restrained by a 100 kcal/mol/Å^2 harmonic restraint. The harmonic restraints were then removed, and the system was again minimized for 500 steps of steepest descent, followed by up to 10 000 steps of conjugate gradient. The system was equilibrated with a 2 ns NVT simulation followed by a 20 ns NPT simulation, and then the production simulation was begun. The temperature of the simulations was set to 300 K, using the Langevin thermostat[94] with a collision frequency of 1.0 ps^{-1}. The pressure was set to 1 bar for both the NPT stage of equilibration and the production simulation, using the Berendsen barostat[95] with the relaxation time set to 2 ps. The particle mesh Ewald (PME) method[96] was used for electrostatic interactions. The cutoff distance of Lennard-Jones interactions was set to 10 Å during the minimization and equilibration phases, and then set to 9 Å for the production simulations. (A sample calculation with a distance cutoff of 14 Å led to an unchanged computed binding enthalpy, to within the numerical uncertainty of the simulations.)

All binding free energy simulations were performed using the same settings as for the enthalpy calculations, above, except that larger simulation boxes were used to accommodate the extended geometries required by the pulling calculations. The edge lengths are approximately 36, 36, and 53 Å for all host-guest systems except for CB7-guest B11 which is 36, 36, and 60 Å. Thus, each system was solvated in a rectangular box with 2210 TIP3P water molecules, except for guest B11 which required 2500 water molecules, due to its larger size. A detailed description of the attach-pull-release framework for calculating the binding free energies is provided in the SI. With the latest generation of GPUs, a performance of over 180 ns per day can be easily achieved.

DOI: 10.1021/acs.jpcb.5b04262
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 2. Derivatives of Computed Binding Enthalpies and Binding Free Energies with Respect to the GAFF Parameters of All Atom Types in CB7, for the Four Training-Set Guests[a]

	A1		A2		B2		B5	
atom type	$\partial\Delta H/\partial\sigma$	$\partial\Delta H/\partial\epsilon$	$\partial\Delta H/\partial\sigma$	$\partial\Delta H/\partial\epsilon$	$\partial\Delta H/\partial\sigma$	$\partial\Delta H/\partial\epsilon$	$\partial\Delta H/\partial\sigma$	$\partial\Delta H/\partial\epsilon$
n	−13.20	−23.52	−12.92	−23.75	−15.92	−20.54	−15.36	−21.99
o	5.36	−1.80	3.30	−4.24	5.02	−1.49	3.13	−5.13
c	−2.32	−18.80	−1.48	−16.57	−4.25	−18.40	−4.24	−20.73
c3[b]	−3.27	−9.51	−3.28	−10.44	−2.56	−3.91	−2.39	−5.67
h2	0.02	2.27	−0.98	−10.80	−0.52	−4.94	−0.93	−12.36
atom type	$\partial\Delta G/\partial\sigma$	$\partial\Delta G/\partial\epsilon$	$\partial\Delta G/\partial\sigma$	$\partial\Delta G/\partial\epsilon$	$\partial\Delta G/\partial\sigma$	$\partial\Delta G/\partial\epsilon$	$\partial\Delta G/\partial\sigma$	$\partial\Delta G/\partial\epsilon$
n	−4.08	−26.17	−2.87	−25.91	−7.97	−25.36	−6.13	−25.37
o	0.27	−8.63	−2.92	−13.07	1.18	−8.18	−1.58	−13.74
c	2.13	−17.14	2.83	−16.47	0.38	−17.53	1.03	−18.50
c3[b]	−5.09	−25.40	−4.93	−25.21	−4.17	−19.56	−3.34	−20.01
h2	−1.10	−20.51	−1.15	−22.05	−0.51	−13.05	−0.82	−17.99

[a]The units of $\partial\Delta H/\partial\sigma$ and $\partial\Delta G/\partial\sigma$ are kcal/(mol·Å); $\partial\Delta H/\partial\epsilon$ and $\partial\Delta G/\partial\epsilon$ are unitless. [b]The **c3** atom type is present in both CB7 and all the guest molecules, but the values reported here were computed based on changes only in **c3** atoms in CB7.

2.5. Error Analysis. Numerical uncertainties of the binding calculations, representing estimated standard deviations of the mean, were obtained by conducting blocking analysis;[93] see SI for details. For a fair comparison, the experimental uncertainties listed here were computed by dividing their reported experimental standard deviations by the square root of the number of replicated experiments, in order to estimate the standard deviation of the mean. The uncertainties of the error statistics were calculated via bootstrap resampling (100 000 times) of the computed and experimental binding data, by randomly sampling from a Gaussian distribution centered on each data point, with its standard deviation set to the standard deviation of the mean.

3. RESULTS

3.1. Baseline Calculations and Selection of Parameters to be Adjusted. Binding free energies and enthalpies computed with the initial GAFF parameters correlate well with experiment (R = 0.93), but systematically overestimate the affinity and exothermicity. In particular, the MSE and RMSD of the binding enthalpies are −3.0 and 4.2 kcal/mol, respectively. The binding enthalpies of all four molecules in the training set are markedly overestimated, and, in the test set, the only exceptions to this pattern are G8 and MVN/Tris, for which the calculations underestimate the exothermicity by 2.3 and 1.3 kcal/mol, respectively. Given the high precision of the calculations, these substantial deviations from experiment must be attributed to errors in the force field. In addition, the nearly uniform tendency to overestimate suggests the existence of a problem in the force field which affects most or all of the results, and whose correction might remove the systematic errors from the calculations. Note, however, that a parameter change which makes all binding enthalpies less favorable could exacerbate the underestimations seen for G8 and MVN/Tris, so one may anticipate that the parameters for these guests will ultimately require special adjustment.

Because they are common to all of the binding calculations, the water model and the CB7 host are obvious places to look for parameters whose adjustment would weaken computed binding enthalpies across many or all of the guest molecules, as required to improve most of the results. Regarding water, we know that this can have a profound effect; as noted in the Introduction, TIP4P-Ew leads to even greater overestimation of exothermicity for these systems. However, adjusting the water model risks generating a water model with unrealistic bulk properties, and we also found that $\partial\Delta H/\partial c$ converges slowly when c pertains to the water model. Therefore, although it will be important in the long term to consider adjusting the water model to fit the properties of pure water along with binding thermodynamics, we chose here to focus on parameters associated with the CB7 host. We focused in particular on its LJ parameters, for several reasons. First, bonded parameters are not expected to strongly influence binding affinities for this rather rigid host. Second, LJ forces were previously found to contribute strongly, and always favorably, to the computed binding enthalpies of these systems,[53] so modest adjustments to these parameters might yield substantial shifts in computed binding enthalpy. Third, the existing GAFF force field assigns the same LJ parameters to many, varied atom types;[46,97] thus, all 13 nitrogen atom types are assigned identical LJ parameters, and there are only three sets of LJ parameters for all small molecule oxygen atoms. Thus, it is plausible that achieving high accuracy calculations with a general force field like GAFF will require more differentiation of LJ parameters across atom types.

The partial derivatives of the binding enthalpies and free energies for the atom types present in the host (Table 2) converged well (SI Figure S3), and the values of $\partial\Delta H/\partial\epsilon$ and $\partial\Delta G/\partial\epsilon$ are nearly all negative. This is consistent with the intuition that deepening the LJ energy well should strengthen binding, and with the prior observation that the LJ term is a major contributor to the computed binding enthalpies of these systems.[53] The sensitivities for σ are less consistent in sign, however. Interestingly, only modest correlations ($R \approx 0.6$) are observed between $\partial\Delta H/\partial\epsilon$ and $\partial\Delta G/\partial\epsilon$ and between $\partial\Delta H/\partial\sigma$ and $\partial\Delta G/\partial\sigma$, indicating that the values of ΔH and ΔG can be separately manipulated by suitable adjustment of LJ parameters; in the present study, we focus on correcting the binding enthalpies.

Given the limited number of experimental data used to guide parametrization in this initial proof-of-concept study, we adjusted parameters for only two atom types, **n** and **o** in GAFF nomenclature (Figure 1), which occur only in the host molecule. The binding enthalpies are quite sensitive to the LJ parameters of these two atom types (Table 2), which means that conservative adjustments should allow significant improvement in the computed enthalpies; and their variation across the training-set guests provides an opportunity to generate different

DOI: 10.1021/acs.jpcb.5b04262
J. Phys. Chem. B XXXX, XXX, XXX−XXX

adjustments to the computed binding enthalpies from one guest to another.

3.2. Parameter Adjustment for Training-Set Host-Guest Binding Enthalpies. Sensitivity analysis predicted that changing σ_n and ϵ_n by −0.1425 Å and −0.0200 kcal/mol, respectively, and σ_o and ϵ_o by 0.1782 Å and 0.0500 kcal/mol, respectively, would significantly improve the agreement of the training set calculations with experiment, with R, the correlation coefficient, increasing from 0.77 to 0.85 and the RMSD decreasing from 4.8 to 1.9 kcal/mol. We made these parameter changes and then used a new round of simulations to recalculate the binding enthalpies of the training set guests. The actual R value for the training set improved to 0.87, and the RMSD decreased to 1.4 kcal/mol, in good agreement with the predictions. Because sensitivity analysis uses a linear approximation to parameter dependencies that are not necessarily linear, we then reran the parameter optimization based on derivatives from the new MD simulations. Sensitivity analysis based on these simulations (Table S1) predicted that further changing σ_n and ϵ_n by 0.0437 Å and −0.0095 kcal/mol, respectively, and σ_o and ϵ_o by 0.1614 Å and 0.0508 kcal/mol, respectively, would improve R to 0.89 and RMSD to 1.2 kcal/mol, for the training set. In good agreement with these predictions, the results from a third set of MD calculations of binding enthalpy with these new LJ parameters (Table 1), yielded values of R and RMSD of 0.90 and 1.0 kcal/mol, respectively. Given the marked improvement and diminishing returns after these two optimization steps, no further iterations were done. The final adjusted parameters differ by an average of 20% from the initial GAFF parameters, and the net effect of the parameter adjustment is to improve agreement with experiment by making the computed binding enthalpies less favorable. It is worth noting that the conformation of free CB7 is not discernibly altered by the new parameters, as illustrated in Figure S4 (SI).

3.3. Application of Optimization Parameters to Test-Set Binding Enthalpies. We used the final modified parameters to compute the binding enthalpies of the test-set guest molecules, A3, A4, A5, B11, G8, G9, and MVN/Tris (Table 1, Final; Figure 2), and compared with the corresponding results based on the initial, unmodified GAFF LJ parameters. As anticipated, the present parameter adjustments weaken the binding enthalpies, and this leads to improved agreement with experiment for the aliphatic guests, A3, A4, A5, and B11, which are similar to the training set: the MSE and RMSD for these guests improve from −2.8 and 3.0 kcal/mol, respectively, for the initial GAFF parameters, to 1.2 and 1.6 kcal/mol, respectively, for the final optimized parameters. Thus, the parameter adjustment shows promising transferability among aliphatic guest molecules.

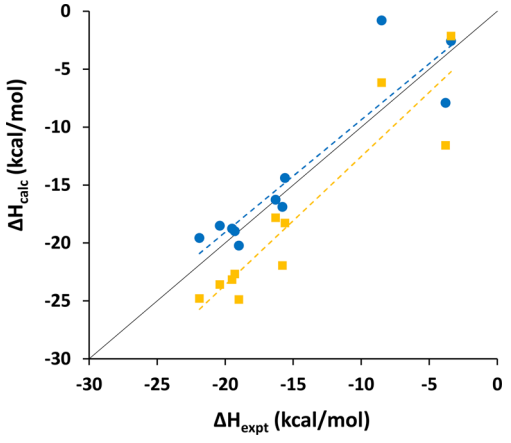


Figure 2. Calculated versus measured binding enthalpies (kcal/mol) for 11 CB7-guest systems. Orange squares: results with initial GAFF parameters. Blue circles: Results after two rounds of adjustment of LJ parameters for atom types **n** and **o** in GAFF nomenclature. Colored lines: Linear regression fits for each set of calculations. Solid black: line of identity, $\Delta H_{calc} = \Delta H_{expt}$.

The pattern of changes is more complex and less favorable for aromatic guests, which is perhaps not surprising, given that there are no aromatic guests in the training set. As in the case of the aliphatic guests, the new parameters make the computed binding enthalpies of both G8 and G9 less negative (Table 1, Final vs Initial). This change improves the agreement with experiment for G9, because its calculated binding enthalpy was too negative with the initial GAFF parameters, but it worsens the accuracy for G8, because the initial parameters GAFF made the binding enthalpy insufficiently favorable. In contrast, however, the new parameters make the binding more, rather than less, exothermic in the case of MVN/Tris (Table 1). This difference traces to the fact that the host in the absence of MVN is predicted to be occupied by a molecule of Tris, so that binding of MVN is actually a competitive process, as previously observed.[53] Further analysis of the binding enthalpy calculations indicates that changing from the initial to final LJ parameters changes the computed binding enthalpy of Tris itself to CB7 from −2.5 to 0.6 kcal/mol, while changing the binding enthalpy of MVN to CB7 in the absence of Tris from −4.7 to −1.9 kcal/mol. Thus, each individual binding event becomes less exothermic, consistent with the results for all of the other guest molecules, but the change for Tris outweighs that for MVN, leading to a more exothermic net result for the displacement of Tris by MVN. Experimental calorimetry data provided subsequent to these calculations support the view that MVN and Tris bind competitively, as the measured binding enthalpy of MVN with CB7 is −6.4 kcal/mol in the absence of Tris and −3.1 kcal/mol in its presence (Prof. Kimoon Kim, personal communication). Overall, the accuracy across all aromatic compounds in the test set does not improve on going from the initial GAFF parameters to the final adjusted ones; the MSE changes from −1.5 to 1.4 kcal/mol, while the RMSD changes from 4.7 to 5.1 kcal/mol. We conjecture that improving the accuracy for these guests would require adjusting parameters specifically associated with the two benzimidazole derivatives, G8 and G9.

The scatter plots in Figure 2 provide an overview of how the parameter adjustment affects all 11 host-guest binding enthalpies. (Additional statistical breakdowns are provided in Table S2.) The main effect of the parameter adjustment is to shift the computed enthalpies in the positive direction, bringing the scatter plot closer to the line of identity. Thus, the MSE improves from −3.1 to 0.8 kcal/mol, indicating removal of the systematic tendency to overestimate the exothermicity of these binding events, and the RMSD across all training and test set compound falls from 4.2 to 2.9 kcal/mol.

3.4. Effect of Optimized Parameters on Computed Binding Free Energies. Sensitivity analysis of the binding free energies eq 4, based on calculations with the initial GAFF parameters, suggested that the LJ parameter adjustments made here would have little effect on the computed binding free energies, with a predicted RMSD of only 0.9 kcal/mol between

DOI: 10.1021/acs.jpcb.5b04262
J. Phys. Chem. B XXXX, XXX, XXX−XXX

the initial and final results. This prediction was borne out by a fresh set of binding free energy calculations with the new parameters ("Final" in Table 1), as the RMSD of the new binding free energies, relative to the initial calculations, is 1.1 kcal/mol. In addition, the mean change in the binding free energy across all 11 cases is only 0.5 kcal/mol, much smaller than the mean change of 3.8 kcal/mol in the binding enthalpies (above). Because $\Delta G = \Delta H - T\Delta S$, these results imply that the change in parameters also changes the entropic contribution to the binding free energy by an average of −3.3 kcal/mol. Thus, these computational data provide a clear case of entropy-enthalpy compensation induced by changes in the LJ parameters.

4. DISCUSSION

The present study demonstrates that molecular dynamics calculations of host-guest binding thermodynamics may readily be used to optimize force field parameters. The optimization procedure is quite efficient, because it uses sensitivity analysis, in which the partial derivatives of the binding enthalpy and free energy with respect to the force field parameters are extracted from the binding calculations, at little extra computational cost. Using this approach, we obtained substantial improvement in training-set accuracy with a single round of optimization of host LJ parameters, and a second round led to additional improvement. Encouragingly, similar improvements in accuracy were observed when the new parameters were used to compute the binding thermodynamics of four test-set guest molecules chemically similar to the guests in the training set. On the other hand, application of the new parameters to chemically distinct test-set guest molecules led to mixed results. Overall, these results provide an initial look at the transferability of the parameter adjustments one may expect from the present approach. It is not unexpected that adjustments to the host molecule alone, as in the present initial study, might not suffice to generate accurate results across all guests. In some cases, such as the benzimidazole guests, G8 and G9, adjustments to guest parameters may also be required. More broadly, although the specific adjusted parameters developed here should be useful for modeling the binding of CB7 to other aliphatic guest molecules, we do not expect them to perform better than the original GAFF parameters for other experimental observables, such as the properties of neat liquids.

In the future, we envision a program of force field optimization in which parameters are adjusted against experimental data sets containing not only host-guest binding thermodynamics but also other measurements of proven value for force field optimization, such as the properties of neat liquids and small molecule hydration free energies. Quantum chemistry results, such as the interaction energies of gas-phase dimers or clusters, may also be useful. Given the large amount of available host-guest binding data,[81,82,84,98−110] we anticipate that this strategy will allow generation of force fields with well-constrained parameters that provide good agreement with experiment for a broadened set of experimental observables. Importantly, the accessibility of partial derivatives of host-guest binding enthalpies and entropies with respect to force field parameters opens the door to including these experimental observables in a variety of systematic parameter optimization schemes.[20,50,76−80] It is also conceivable that this effort will reveal weaknesses in the functional form of commonly used force fields. For example, perhaps an explicit treatment of electronic polarization will be needed to handle the transfer of guest molecules from water to the less polar interior of a host. If so, including host-guest binding data may help guide the selection and development of efficient force field models with more advanced functional forms.

The computational accessibility of binding enthalpies, binding free energies, and the gradients of these observables with respect to force field parameters, means that such a program of force field optimization can benefit from adjusting parameters against both of these thermodynamic quantities. In particular, if one parametrizes based solely on free energy, one risks getting the right free energies for one's training set, but with an incorrect enthalpy−entropy balance, and hence reducing the likelihood that the results will be transferable to other systems. It is also worth noting that including both free energies and enthalpies in the training data set is consistent with existing practice, in which force fields are adjusted based on, for example, enthalpies of vaporization as well as free energies of solvation. More generally, the weak correlation between measured binding enthalpies and free energies for many types of systems[61,71] means that these two quantities provide largely independent, and hence complementary, information to the process of parameter optimization. Indeed, we found that adjusting parameters to better replicate measured binding enthalpies can have a remarkably small effect on the computed binding free energies. This means that our adjustment of LJ parameters was associated with entropy-enthalpy compensation, as it made binding less enthalpically favored, but more entropically favored, in almost equal measure. Although these changes resulted from an artificial adjustment of force field parameters, the overall pattern of entropy-enthalpy compensation[111−115] is one that has been seen experimentally and discussed for many years. The fact that it appears spontaneously in the present simulations raises the broader possibility that analysis of simulations of experimentally interesting systems can not only help with force field optimization, but will also advance our understanding of the physical chemistry of molecular recognition.

■ ASSOCIATED CONTENT

Supporting Information

Equations deriving partial derivatives of binding thermodynamics with respect to force field parameters; protocols of calculating binding thermodynamics and numerical uncertainties; derivatives computed from the second round of binding enthalpy calculations with TIP3P water; error analysis of computed binding enthalpies in TIP3P water before and after perturbations; details of the convergence of binding enthalpy derivatives; and comparison of CB7 conformational variables with initial and final parameters. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b04262.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mgilson@ucsd.edu. Phone: 858-822-0622. Fax: 858-822-7726.

Author Contributions

†J.Y. and A.T.F. contributed equally to this work.

Notes

The authors declare no competing financial interest.

DOI: 10.1021/acs.jpcb.5b04262
J. Phys. Chem. B XXXX, XXX, XXX−XXX

■ ACKNOWLEDGMENTS

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. M.K.G. acknowledges funding from National Institute of General Medical Sciences (GM61300), and from the AFOSR through a Basic Research Initiative (BRI) grant (FA9550-12-1-0414). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS, NIH, NSF or USAF. M.K.G. has an equity interest in, and is a cofounder and scientific advisor of VeraChem LLC. We thank Professor Kimoon Kim for providing information regarding the binding thermodynamics of methylviologen in the absence of Tris, and Dr. David Slochower for useful comments.

■ REFERENCES

(1) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919−11929.

(2) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, *116*, 9058−9067.

(3) Nadler, W.; Hansmann, U. H. E. Generalized Ensemble and Tempering Simulations: A Unified View. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2007**, *75*, 026109.

(4) Nakajima, N.; Nakamura, H.; Kidera, A. Multicanonical Ensemble Generated by Molecular Dynamics Simulation for Enhanced Conformational Sampling of Peptides. *J. Phys. Chem. B* **1997**, *101*, 817−824.

(5) Elber, R.; Karplus, M. Enhanced Sampling in Molecular Dynamics: Use of the Time-Dependent Hartree Approximation for a Simulation of Carbon Monoxide Diffusion through Myoglobin. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.

(6) Wu, X.; Damjanovic, A.; Brooks, B. Efficient and Unbiased Sampling of Biomolecular Systems in the Canonical Ensemble: A Review of Self-Guided Langevin Dynamics. *Adv. Chem. Phys.* **2012**, *150*, 255−326.

(7) Okamoto, Y. Generalized-Ensemble Algorithms: Enhanced Sampling Techniques for Monte Carlo and Molecular Dynamics Simulations. *J. Mol. Graphics Modell.* **2004**, *22*, 425−439.

(8) Junghans, C.; Perez, D.; Vogel, T. Molecular Dynamics in the Multicanonical Ensemble: Equivalence of Wang-Landau Sampling, Statistical Temperature Molecular Dynamics, and Metadynamics. *J. Chem. Theory Comput.* **2014**, *10*, 1843−1847.

(9) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(10) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.

(11) Comer, J.; Gumbart, J. C.; Hénin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. *J. Phys. Chem. B* **2015**, *119*, 1129−1151.

(12) Hansen, H. S.; Hünenberger, P. H. Using the Local Elevation Method to Construct Optimized Umbrella Sampling Potentials: Calculation of the Relative Free Energies and Interconversion Barriers of Glucopyranose Ring Conformers in Water. *J. Comput. Chem.* **2010**, *31*, 1−23.

(13) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(14) Darve, E.; Pohorille, A. Calculating Free Energies Using Average Force. *J. Chem. Phys.* **2001**, *115*, 9169−9183.

(15) Ohmura, I.; Morimoto, G.; Ohno, Y.; Hasegawa, A.; Taiji, M. MDGRAPE-4: A Special-Purpose Computer System for Molecular Dynamics Simulations. *Philos. Trans. R. Soc., A* **2014**, *372*, 20130387.

(16) Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; et al. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis*; IEEE Computer Society: New York, 2014; pp 41−53.

(17) Anderson, J. A.; Lorenz, C. D.; Travesset, A. General Purpose Molecular Dynamics Simulations Fully Implemented on Graphics Processing Units. *J. Comput. Phys.* **2008**, *227*, 5342−5359.

(18) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.

(19) Hagler, A. T.; Huler, E.; Lifson, S. Energy Functions for Peptides and Proteins. I. Derivation of a Consistent Force Field Including the Hydrogen Bond from Amide Crystals. *J. Am. Chem. Soc.* **1974**, *96*, 5319−5327.

(20) Lifson, S.; Warshel, A. Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and N-Alkane Molecules. *J. Chem. Phys.* **1968**, *49*, 5116−5129.

(21) Allinger, N. L. Calculation of Molecular Structure and Energy by Force-Field Methods. *Adv. Phys. Org. Chem.* **1976**, *13*, 1−82.

(22) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.

(23) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(24) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(25) Bayly, C. C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(26) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(27) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; Distasio, R. A.; et al. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.

(28) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2005**, *1*, 153−168.

(29) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical Fluctuating Charge Force Fields: Application to Liquid Water. *J. Chem. Phys.* **1994**, *101*, 6141−6156.

(30) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. Development of an Accurate and Robust Polarizable Molecular Mechanics Force Field from Ab Initio Quantum Chemistry. *J. Phys. Chem. A* **2004**, *108*, 621−627.

(31) Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. Strike a Balance: Optimization of Backbone Torsion Parameters of AMBER Polarizable Force Field for Simulations of Proteins and Peptides. *J. Comput. Chem.* **2006**, *27*, 781−790.

(32) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(33) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; Mackerell, A. D. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013**, *9*, 5430−5449.

DOI: 10.1021/acs.jpcb.5b04262
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(34) Momany, F. A. Determination of Partial Atomic Charges from Ab Initio Molecular Electrostatic Potentials. Application to Formamide, Methanol, and Formic Acid. *J. Phys. Chem.* **1978**, *82*, 592−601.
(35) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046−4063.
(36) Paesani, F.; Iuchi, S.; Voth, G. A. Quantum Effects in Liquid Water from an Ab Initio-Based Polarizable Force Field. *J. Chem. Phys.* **2007**, *127*, 074506.
(37) Paesani, F.; Zhang, W.; Case, D. A.; Cheatham, T. E.; Voth, G. A. An Accurate and Simple Quantum Model for Liquid Water. *J. Chem. Phys.* **2006**, *125*, 184507.
(38) Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J. R.; Hwang, M. J.; Li, H.; Liang, C.; Maple, J.; Peng, Z.; Stockfisch, T. P.; et al. Derivation of Class II Force Fields. VIII. Derivation of a General Quantum Mechanical Force Field for Organic Compounds. *J. Comput. Chem.* **2001**, *22*, 1782−1800.
(39) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OLPS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(40) Chapman, D. E.; Steck, J. K.; Nerenberg, P. S. Optimizing Protein−Protein van Der Waals Interactions for the AMBER ff9x/ff12 Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 273−281.
(41) Ryckaert, J.-P.; Bellemans, A. Molecular Dynamics of Liquid Alkanes. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 95−106.
(42) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.
(43) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950−1958.
(44) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712−725.
(45) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulation. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(46) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.
(47) Kang, M.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Amides. *J. Comput. Chem.* **2006**, *27*, 1477−1485.
(48) Weerasinghe, S.; Smith, P. E. Kirkwood−Buff Derived Force Field for Mixtures of Acetone and Water. *J. Chem. Phys.* **2003**, *118*, 10663.
(49) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(50) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.
(51) Shirts, M. R.; Pande, V. S. Solvation Free Energies of Amino Acid Side Chain Analogs for Common Molecular Mechanics Water Models. *J. Chem. Phys.* **2005**, *122*, 134508.
(52) Mobley, D. L.; Dumont, É.; Chodera, J. D.; Dill, K. A. Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent. *J. Phys. Chem. B* **2007**, *111*, 2242−2254.
(53) Fenley, A. T.; Henriksen, N. M.; Muddana, H. S.; Gilson, M. K. Bridging Calorimetry and Simulation through Precise Calculations of Cucurbituril − Guest Binding Enthalpies. *J. Chem. Theory Comput.* **2014**, *10*, 4069−4078.
(54) Fornabaio, M.; Spyrakis, F.; Mozzarelli, A.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E. Simple, Intuitive Calculations of Free Energy of Binding for Protein-Ligand Complexes. 3. The Free Energy Contribution of Structural Water Molecules in HIV-1 Protease Complexes. *J. Med. Chem.* **2004**, *47*, 4507−4516.
(55) Rarey, M.; Kramer, B.; Lengauer, T. The Particle Concept: Placing Discrete Water Molecules during Protein- Ligand Docking Predictions. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 17−28.
(56) Hamelberg, D.; McCammon, J. A. Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004**, *126*, 7683−7689.
(57) Lu, Y.; Yang, C. Y.; Wang, S. Binding Free Energy Contributions of Interfacial Waters in HIV-1 Protease/inhibitor Complexes. *J. Am. Chem. Soc.* **2006**, *128*, 11830−11839.
(58) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 808−813.
(59) Nguyen, C. N.; Kurtzman Young, T.; Gilson, M. K. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137*, 044101.
(60) Rekharsky, M. V.; Mori, T.; Yang, C.; Ko, Y. H.; Selvapalam, N.; Kim, H.; Sobransingh, D.; Kaifer, A. E.; Liu, S.; Isaacs, L.; et al. A Synthetic Host-Guest System Achieves Avidin-Biotin Affinity by Overcoming Enthalpy-Entropy Compensation. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20737−20742.
(61) Chang, C.; Gilson, M. K. Free Energy, Entropy, and Induced Fit in Host−Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *J. Am. Chem. Soc.* **2004**, *126*, 13156−13164.
(62) Moghaddam, S.; Yang, C.; Rekharsky, M.; Ko, Y. H.; Kim, K.; Inoue, Y.; Gilson, M. K. New Ultrahigh Affinity Host - Guest Complexes of Cucurbit [7] Uril with Bicyclo [2. 2. 2] Octane and Adamantane Guests: Thermodynamic Analysis and Evaluation of M2 Affinity Calculations. *J. Am. Chem. Soc.* **2011**, *133*, 3570−3581.
(63) Moghaddam, S.; Inoue, Y.; Gilson, M. K. Host-Guest Complexes with Protein-Ligand-like Affinities: Computational Analysis and Design. *J. Am. Chem. Soc.* **2009**, *131*, 4012−4021.
(64) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. The SAMPL2 Blind Prediction Challenge: Introduction and Overview. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 259−279.
(65) Muddana, H. S.; Varnado, C. D.; Bielawski, C. W.; Urbach, A. R.; Isaacs, L.; Geballe, M. T.; Gilson, M. K. Blind Prediction of Host-Guest Binding Affinities: A New SAMPL3 Challenge. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 475−487.
(66) Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K. The SAMPL4 Host-Guest Blind Prediction Challenge: An Overview. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 305−317.
(67) Wickstrom, L.; He, P.; Gallicchio, E.; Levy, R. M. Large Scale Affinity Calculations of Cyclodextrin Host-Guest Complexes: Understanding the Role of Reorganization in the Molecular Recognition Process. *J. Chem. Theory Comput.* **2013**, *9*, 3136−3150.
(68) Inoue, Y.; Hakushi, T.; Liu, Y.; Tong, L.-H.; Shen, B.-J.; Jin, D.-S. Thermodynamics of Molecular Recognition by Cyclodextrins 1. Calorimetric Titration of Inclusion Complexation of Naphthalenesulfonates with a-, B-, and G-Cyclodextrins: Enthalpy-Entropy Compensation. *J. Am. Chem. Soc.* **1993**, *115*, 475−481.
(69) Rekharsky, M.; Inoue, Y. Chiral Recognition Thermodynamics of Beta-Cyclodextrin: The Thermodynamic Origin of Enantioselectivity and the Enthalpa-Entropy Compensation Effect. *J. Am. Chem. Soc.* **2000**, *122*, 4418−4435.
(70) Lai, B.; Oostenbrink, C. Binding Free Energy, Energy and Entropy Calculations Using Simple Model Systems. *Theor. Chem. Acc.* **2012**, *131*, 1272.

DOI: 10.1021/acs.jpcb.5b04262
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(71) Bohm, H.; Klebe, G. What Can We Learn from Molecular Recognition in Protein-Ligand Complexes for the Design of New Drugs? *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2588−2614.
(72) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(73) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(74) Caleman, C.; Van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; Van Der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61−74.
(75) Wong, C. F.; Thacher, T.; Rabitz, H. Sensitivity Analysis in Biomolecular Simulation. In *Reviews in Computational Chemistry*; Wiley-VCH: New York, 1998.
(76) Di Pierro, M.; Mugnai, M. L.; Elber, R. Optimizing Potentials for a Liquid Mixture: A New Force Field for a Tert-Butanol and Water Solution. *J. Phys. Chem. B* **2015**, *119*, 836.
(77) Di Pierro, M.; Elber, R. Automated Optimization of Potential Parameters. *J. Chem. Theory Comput.* **2013**, *9*, 3311−3320.
(78) Wang, L. P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885−1891.
(79) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B* **2013**, *117*, 9956−9972.
(80) Shell, M. S. The Relative Entropy Is Fundamental to Multiscale and Inverse Thermodynamic Problems. *J. Chem. Phys.* **2008**, *129*, 144108.
(81) Freeman, W. A.; Mock, W. L.; Shih, N.-Y. Cucurbituril. *J. Am. Chem. Soc.* **1981**, *103*, 7367−7368.
(82) Lee, J. W.; Samal, S.; Selvapalam, N.; Kim, H.-J.; Kim, K. Cucurbituril Homologues and Derivatives: New Opportunities in Supramolecular Chemistry. *Acc. Chem. Res.* **2003**, *36*, 621−630.
(83) Kim, H.-J.; Jeon, W. S.; Ko, Y. H.; Kim, K. Inclusion of Methylviologen in cucurbit[7]uril. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5007−5011.
(84) Jeon, Y. J.; Kim, S.-Y.; Ko, Y. H.; Sakamoto, S.; Yamaguchi, K.; Kim, K. Novel Molecular Drug Carrier: Encapsulation of Oxaliplatin in cucurbit[7]uril and Its Effects on Stability and Reactivity of the Drug. *Org. Biomol. Chem.* **2005**, *3*, 2122−2125.
(85) Velez-Vega, C.; Gilson, M. K. Overcoming Dissipation in the Calculation of Standard Binding Free Energies by Ligand Extraction. *J. Comput. Chem.* **2013**, *34*, 2360−2371.
(86) Zwanzig, R. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(87) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(88) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(89) Case, D. A.; Darden, T.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; et al. *AMBER 14*; University of California: San Francisco, 2014.
(90) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.
(91) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(92) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.
(93) Flyvbjerg, H.; Petersen, H. G. Error Estimates on Averages of Correlated Data. *J. Chem. Phys.* **1989**, *91*, 461−466.
(94) Zwanzig, R. Nonlinear Generalized Langevin Equations. *J. Stat. Phys.* **1973**, *9*, 215−220.
(95) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(96) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.
(97) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Dill, K. A. Predictions of Hydration Free Energies from All-Atom Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 4533−4537.
(98) Lagona, J.; Mukhopadhyay, P.; Chakrabarti, S.; Isaacs, L. The Cucurbit[n]uril Family. *Angew. Chem., Int. Ed.* **2005**, *44*, 4844−4870.
(99) Rekharsky, M. V.; Inoue, Y. Complexation Thermodynamics of Cyclodextrins. *Chem. Rev.* **1998**, *98*, 1875−1918.
(100) Zimmerman, S. C.; Vanzyl, C. M. Rigid Molecular Tweezers: Synthesis, Characterization, and Complexation Chemistry of a Diacridine. *J. Am. Chem. Soc.* **1987**, *109*, 7894−7896.
(101) Zimmerman, S. C. Rigid Molecular Tweezers as Hosts for the Complexation of Neutral Guests. In *Supramolecular Chemistry I—Directed Synthesis and Molecular Recognition*; Weber, E., Ed.; Topics in Current Chemistry *165*; Springer: Berlin, 1993; pp 71−102.
(102) Klärner, F. G.; Kahlert, B. Molecular Tweezers and Clips as Synthetic Receptors. Molecular Recognition and Dynamics in Receptor-Substrate Complexes. *Acc. Chem. Res.* **2003**, *36*, 919−932.
(103) Sinha, S.; Lopes, D. H. J.; Du, Z.; Pang, E. S.; Shanmugam, A.; Lomakin, A.; Talbiersky, P.; Tennstaedt, A.; McDaniel, K.; Bakshi, R.; et al. Lysine-Specific Molecular Tweezers Are Broad-Spectrum Inhibitors of Assembly and Toxicity of Amyloid Proteins. *J. Am. Chem. Soc.* **2011**, *133*, 16958−16969.
(104) Linton, B.; Hamilton, A. D. Host-Guest Chemistry: Combinatorial Receptors. *Curr. Opin. Chem. Biol.* **1999**, *3*, 307−312.
(105) Dotsevi, G.; Sogah, Y.; Cram, D. J. Total Chromatographic Optical Resolutions Of. Alpha.-Amino Acid and Ester Salts through Chiral Recognition by a Host Covalently Bound to Polystyrene Resin. *J. Am. Chem. Soc.* **1976**, *98*, 3038−3041.
(106) Cram, D.; Helgeson, R. Host-Guest Complexation. 8. Macrocyclic Polyethers Shaped by Two Rigid Substituted Dinaphthyl or Ditetralyl Units. *J. Org. Chem.* **1978**, *43*, 1930−1946.
(107) Rebek, J., Jr. Host-guest Chemistry of Calixarene Capsules. *Chem. Commun.* **2000**, 637−643.
(108) Rebek, J., Jr.; Nemeth, D. Molecular Recognition: Ionic and Aromatic Stacking Interactions Bind Complementary Functional Groups in a Molecular Cleft. *J. Am. Chem. Soc.* **1986**, *108*, 5637−5638.
(109) Rebek, J., Jr. Progress in Molecular Recognition. In *Environmental Influences and Recognition in Enzyme Chemistry*; Liebman, J. F., Greenberg, A., Eds.; VCH Publishers: New York, 1988; pp 291−250.
(110) Chang, S. K.; van Engen, D.; Fan, E.; Hamilton, A. D. Hydrogen Bonding and Molecular Recognition: Synthetic, Complexation, Amd Structural Studies on Barbiturate Binding to an Artificial Receptor. *J. Am. Chem. Soc.* **1991**, *113*, 7640−7645.
(111) Chodera, J. D.; Mobley, D. L. Entropy-Enthalpy Compensation: Role and Ramifications in Biomolecular Ligand Recognition and Design. *Annu. Rev. Biophys.* **2013**, *42*, 121−142.
(112) Sharp, K. Entropy-Enthalpy Compensation: Fact or Artifact? *Protein Sci.* **2001**, *10*, 661−667.
(113) Fenley, A. T.; Muddana, H. S.; Gilson, M. K. Entropy-Enthalpy Transduction Caused by Conformational Shifts Can Obscure the Forces Driving Protein-Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 20006−20011.
(114) Lumry, R.; Rajender, S. Enthalpy-Entropy Compensation Phenomena in Water Solutions of Proteins and Small Molecules: A Ubiquitous Property of Water. *Biopolymers* **1970**, *9*, 1125−1227.

DOI: 10.1021/acs.jpcb.5b04262
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(115) Dunitz, J. D. Win Some, Lose Some: Enthalpy-Entropy Compensation in Weak Intermolecular Interactions. *Chem. Biol.* **1995**, *2*, 709−712.

DOI: 10.1021/acs.jpcb.5b04262
J. Phys. Chem. B XXXX, XXX, XXX−XXX