Application of the Alchemical Transfer and Potential of Mean Force Methods to the SAMPL8 Host-Guest Blinded Challenge

Solmaz Azimi · Joe Z. Wu · Sheenam Khuttan · Tom Kurtzman · Nanjie Deng · Emilio Gallicchio*

Abstract We report the results of our participation in the SAMPL8 GDCC Blind Challenge for host-guest binding affinity predictions. Absolute binding affinity prediction is of central importance to the biophysics of molecular association and pharmaceutical discovery. The blinded SAMPL series have provided an important forum for assessing the reliability of binding free energy methods in an objective way. In this challenge, we employed two binding free energy methods, the newly developed alchemical transfer method (ATM) and the well-established potential of mean force (PMF) physical pathway method, using the same setup and force field model. The calculated binding free energies from the two methods are in excellent quantitative agreement. Importantly, the results from the two methods were also found to agree well with the experimental binding affinities released subsequently, with R values of 0.89 (ATM) and 0.83 (PMF). These results were ranked among the best of the SAMPL8 GDCC challenge and second only to those obtained with the more accurate AMOEBA force field. Interestingly, the two host molecules included in the challenge (TEMOA and TEETOA) displayed distinct binding mechanisms, with TEMOA undergoing a dehydration transition whereas guest binding to TEETOA resulted in the opening of the binding cavity that remains essentially dry during the process. The coupled reorganization and hydration equilibria observed in these systems is a useful prototype for the study of these phenomena often observed in the formation of protein-ligand complexes. Given that the two free energy methods employed here are based on entirely different thermodynamic pathways, the close agreement between the two and their general agreement with the experimental binding free energies are a testament to the high quality and precision achieved by theory and methods. The study provides further validation of the novel ATM binding free energy estimation protocol and paves the way to further extensions of the method to more complex systems.

S. Azimi
Department of Chemistry, Brooklyn College of the City University of New York
PhD Program in Biochemistry, Graduate Center of the City University of New York

J. Z. Wu
Department of Chemistry, Brooklyn College of the City University of New York
PhD Program in Chemistry, Graduate Center of the City University of New York

S. Khuttan
Department of Chemistry, Brooklyn College of the City University of New York
PhD Program in Biochemistry, Graduate Center of the City University of New York

T. Kurtzman
Department of Chemistry, Lehman College of the City University of New York
PhD Program in Chemistry, Graduate Center of the City University of New York
PhD Program in Biochemistry, Graduate Center of the City University of New York

N. Deng
Department of Chemistry and Physical Sciences, Pace University, New York, New York

E. Gallicchio
Department of Chemistry, Brooklyn College of the City University of New York
PhD Program in Chemistry, Graduate Center of the City University of New York
PhD Program in Biochemistry, Graduate Center of the City University of New York
E-mail: egallicchio@brooklyn.cuny.edu

1 Introduction

The Statistical Assessment of Modeling of Proteins and Ligands (SAMPL) series of community challenges[1, 2, 3] have been organized to validate computational methods of molecular solvation and binding in an unbiased way. SAMPL participants are required to quantitatively predict experimental measurements that are publicly disclosed only after the predictions are submitted. The format of the challenges allows the robust assessment of computational methods and have significantly contributed to their advancement.[4] As computational models of small molecule binding to protein receptors increasingly emerge as important elements of structure-based drug discovery,[5, 6] it is critical that the reliability of these models is independently assessed and validated. We have contributed to several editions of the SAMPL challenges to validate the ability of our computational models to accurately predict host-guest and protein-ligand binding affinities[7, 8, 9, 10, 11].

In this work, we apply two conceptually orthogonal yet equivalent binding free energy estimation methods, the Alchemical Transfer Method (ATM)[12] and the Potential of Mean Force (PMF)[13] method, to the SAMPL8 GDCC challenge set[1]. The modeled predictions are tested against each other, as well as with the blinded experimental binding free energies measured by the Gibb Group.[14][2]

In principle, computational models should yield equivalent binding free energy predictions as long as they are based on the same chemical model and physical description of inter-atomic interactions. By ensuring consistency between two independent computational estimates, we can achieve an increased level of confidence in the theoretical accuracy of the models and in the correctness of their implementation. Furthermore, by comparing the computational predictions to the experimental measurements in a blinded, unbiased fashion, we can assess the predictive capability that can be expected of the models in actual chemical applications.

While a variety of empirical methods are commonly used to model the binding affinities of molecular complexes,[15, 16] here we are concerned with methods based on physical models of inter-atomic interactions and a rigorous statistical mechanics theory of the free energy of molecular binding.[17, 18, 19] Binding free energy methods are classified as physical or alchemical depending on the nature of the thermodynamic path employed to connect the unbound to the bound states of the molecular complex for computing the reversible work of binding.[20] Physical pathway methods define a physical path in coordinate space in which the reversible work for bringing the two molecules together is calculated. Conversely, alchemical methods connect the bound and unbound states by a series of artificial intermediate states in which the ligand is progressively decoupled from the solution environment and coupled to the receptor.

In this work, we compare the results of the PMF method,[13] a physical pathway approach, to that of the ATM alchemical method[12] on identically-prepared molecular systems. Because free energy is a thermodynamic state function, binding free energy estimates should be independent of the specific path employed, whether physical or alchemical. Obtaining statistically equivalent estimates of the binding free energies using these two very different thermodynamic paths constitutes a robust validation of both methods. The very recently developed ATM, in particular, benefits from the backing of the more established PMF method in this application.

We also investigate the mechanisms of binding for the systems under question, of which all five guests exhibit lower binding affinity to the TEETOA host than to the TEMOA host. We do this in context of our prior work [21], in which we investigate the connection between the thermodynamics of host reorganization and the thermodynamics of binding site solvation. The simulation study elucidates a reorganization transition that displaces unfavorable water molecules in the binding cavity of a beta-cyclodextrin host, a process that does not generally benefit from favorable thermodynamic contribution of water expulsion from the binding site cavity to facilitate ligand binding. A similar phenomenon is observed in this study in the TEETOA host, which also does not benefit from the generally favorable thermodynamic contribution of water displacement in the binding site cavity. A guest binding to TEETOA competes with an internal reorganization of the host that is not coupled to a hydration/dehydration transition, which is observed in the TEMOA host and is attributed to more favorable binding.

This paper is organized as follows. To begin, we review the theory behind both PMF and ATM methods, describe the host-guest systems included in the SAMPL8 GDCC challenge, and provide the system setup and simulation details of our free energy calculations. We then present the binding free energy estimates we obtained with the PMF and ATM approaches and compare them to each other and with the experimental measurements that were disclosed only after the predictions were submitted to the SAMPL8 organizers.

[1] github.com/samplchallenges/SAMPL8/tree/master/-host_guest/GDCC

[2] github.com/samplchallenges/SAMPL8/blob/master/-host_guest/Analysis/ExperimentalMeasurements/-Final-Data-Table-031621-SAMPL8.docx

We conclude this report with a preliminary analysis of the distinctive binding mechanism of the TEETOA and TEMOA hosts. Overall, the work shows that the ATM and PMF methods provide consistent binding free energy estimates that, in conjunction with the force field model employed here, are in statistical agreement with experimental observations. Moreover, the coupled reorganization and hydration equilibria observed in these systems is a useful prototype for the study of these phenomena in the formation of protein-ligand complexes.

2 Theory and Methods

2.1 The Potential of Mean Force Method

The Potential of Mean Force method, hereon PMF, employed in this work is a physical binding pathway approach fully described in reference 13. Here, we briefly summarize the statistical mechanics basis of the method. Implementation details specific to this work are described in the Computational Details section.

The PMF method estimates the standard free energy of binding as the sum of the free energy changes of the following processes:

1. The transfer of one ligand molecule from an ideal solution at the standard concentration $C^\circ = 1M$ to a region in the solvent bulk of volume equal to the volume of the receptor binding site, followed by the imposition of harmonic restraints that keep the ligand in a chosen reference binding orientation. The free energy term corresponding to this process, denoted as $\Delta G^{\text{bulk}}_{\text{restr}}$, is evaluated analytically.
2. The transfer of the ligand molecule from the solvent bulk to the receptor binding site along a suitable physical pathway (see Computational Details). The free energy change along this pathway is described by a potential of mean force parameterized by the distance between two reference atoms each of the ligand and the receptor (Figures 1). The free energy change for this process, denoted by $w(r_{\text{min}}) - w(r^*)$, is given by the value at the minimum of the potential of mean force relative to the value in the bulk (Figure 3, PMF panel).
3. ΔG_{vibr} is related to the ratio of the configurational partition functions of the ligand within the binding site of the receptor reflecting the bound basin of the PMF, versus when it is harmonically restrained at the bulk location r^*.
4. The release of the harmonic restraints while the ligand is bound to the receptor. The free energy change for this process, denoted by $-\Delta G^{\text{bound}}_{\text{restr}}$, is evaluated by Bennett's Acceptance Ratio method (BAR).

Hence, the PMF estimate of the free energy of binding is given by

$$\Delta G^\circ_b = \Delta G^{\text{bulk}}_{\text{restr}} + [w(r_{\text{min}}) - w(r^*)] + \Delta G_{\text{vibr}} - \Delta G^{\text{bound}}_{\text{restr}} \quad (1)$$

Additional computational details and parameters used in this work to implement the PMF calculations are described in the Computational Details section.

2.2 The Alchemical Transfer Method

The Alchemical Transfer Method, hereon ATM, is a recently-developed method to compute the absolute binding free energy of molecular complexes. The method is fully described in reference 12. Here, we give only a brief overview of ATM, particularly focusing on the aspects specific to this work. Further implementation details are described in the Computational Details section.

Given the standard free energy of binding ΔG°_b, defined as the difference in free energy between the bound complex and the unbound components, $\Delta G^\circ_{\text{b}} = \Delta G^\circ_{\text{site}} + \Delta G^*_b$. ATM computes the excess component of the binding free energy, ΔG^*_b, defined as the reversible work for transferring the ligand from a region of volume V_{site} in the solvent bulk to a region of the same volume in the receptor binding site.[18] The standard free energy of binding is given by the excess component plus the ideal component, $\Delta G^\circ_{\text{site}} = -k_B T \ln C^\circ V_{\text{site}}$, which corresponds to the free energy change of transferring one ligand molecule from an ideal solution at the standard concentration $C^\circ = 1M$ to a region in the solvent bulk of volume that is equal to the volume of the receptor binding site, V_{site}.[17] The concentration-dependent ideal term is computed analytically and the excess component is computed by ATM using numerical molecular simulations described in Computational Details and below.

As discussed in 12, in ATM, the transfer of the ligand from the solvent bulk to the receptor binding site could not be designed in terms of a single continuous transformation. Rather, it is carried out in two alchemical steps that connect the bound and unbound end states to a common alchemical intermediate (Figure 2), in which the ligand molecule interacts equally with both the receptor and the solvent bulk at half strength. The potential energy function of the alchemical intermediate is defined as

$$U_{1/2}(x_S, x_L) = \frac{1}{2} U(x_S, x_L) + \frac{1}{2} U(x_S, x_L + h)\,, \quad (2)$$

where x_S denotes the coordinates of the atoms of the receptor and of the solvent, which are unmodified by the coordinate transformation, x_L denotes the coordinates of the atoms of the ligand while in the receptor binding site, and h is the constant displacement vector that

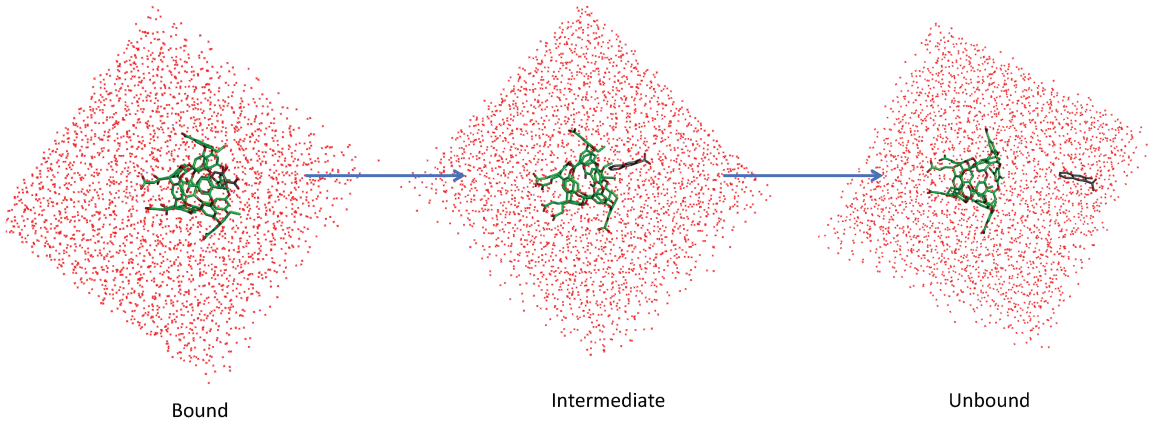

Fig. 1 Schematic of Potential of Mean Force (PMF) method. From left to right, the figure represents the physical pathway that the ligand undergoes from the bound to unbound state. Shown above is a sequence of 3 snapshots representing 3 of the 20 umbrella windows, where the ligand gets pulled at varying distances along the physical pathway away from the host (through the use of reference atoms assigned to both the ligand and host). The red dots represent the oxygen atoms of water molecules. The big bulky molecule represents the TEMOA host, while the small molecule represents the G1 guest.

brings the atoms of the ligand from the receptor site to the solvent bulk site. In this scheme, $U(x_S, x_L)$ is the potential energy of the system when the ligand is in the binding site, $U(x_S, x_L + h)$ is the potential energy after translating the ligand rigidly into the solvent bulk, and $U_{1/2}(x_S, x_L)$ is the hybrid alchemical potential given by the average of the two. In the alchemical intermediate state, receptor atoms and solvent molecules interact with the ligand at half strength but at both ligand locations. Similarly, the force that ligand atoms interact with receptor atoms and solvent molecules at the intermediate state is an average of the forces exerted by the ligand at the two distinct locations. As discussed in reference 12, the ATM alchemical intermediate has an analogous role as the vacuum intermediate state in the conventional double-decoupling method,[17] but without fully dehydrating the ligand.

The bound state of the complex is connected through leg 1 to the alchemical intermediate (Figure 2) by means of a λ-dependent alchemical potential of the form

$$U_\lambda(x) = U(x_S, x_L) + \lambda u_{\rm sc}[u(x)]; \quad \text{Leg1} \tag{3}$$

where $x = (x_S, x_L)$ are the degrees of freedom of the system,

$$u(x) = U(x_S, x_L + h) - U(x_S, x_L) \tag{4}$$

is the binding energy function, $u_{\rm sc}(u)$ is the soft-core perturbation function defined below, and λ is a progress parameter that goes from 0 to 1/2. The corresponding expression of the alchemical potential energy function for leg 2, which connects the unbound state to the alchemical intermediate (Figure 2), is

$$U_\lambda(x) = U(x_S, x_L - h) + \lambda u_{\rm sc}[-u(x)]; \quad \text{Leg2.} \tag{5}$$

Finally,

$$u_{\rm sc}(u) = u; \quad u \le u_c \tag{6}$$

$$u_{\rm sc}(u) = (u_{\rm max} - u_c) f_{\rm sc}\left[\frac{u - u_c}{u_{\rm max} - u_c}\right] + u_c; \quad u > u_c \tag{7}$$

with

$$f_{\rm sc}(y) = \frac{z(y)^a - 1}{z(y)^a + 1}, \tag{8}$$

and

$$z(y) = 1 + 2y/a + 2(y/a)^2 \tag{9}$$

is a soft-core perturbation function that avoids singularities near the initial states of each leg ($\lambda = 0$). The parameters of the soft-core function, $u_{\rm max}$, u_c, and a used in this work are listed in Computational Details.

The free energy change for each leg as a function of λ is obtained by multi-state thermodynamic reweighting[22] using the perturbation energies $u_{\rm sc}[u(x)]$ collected during the molecular dynamics runs. Figure 3 (ATM panel) illustrates the ATM free energy profile in going to the

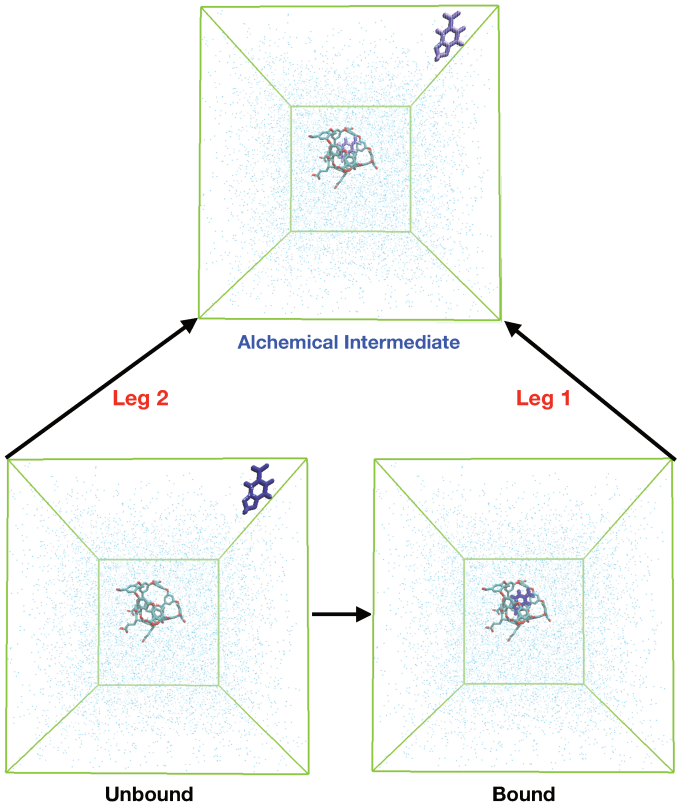

Fig. 2 The Alchemical Transfer Method (ATM) involves two simulation legs, which, in total, transfer the ligand from the solvent bulk to the binding site of the receptor. The two legs connect the bound and unbound end states through an alchemical intermediate that involves the ligand molecule interacting equally with both the receptor and the solvent bulk at half strength. Here, the receptor is the TEMOA host and the ligand is the G4 guest. The green box represents the solvent box with water molecules designated in blue. In the TEMOA structure, carbon atoms are represented in cyan and oxygen atoms in red.

unbound state at $\lambda = 0$ to the bound state at $\lambda = 1$, obtained by concatenating the free energy profile for leg 1 to that of leg 2 at the midpoint. As illustrated by the thermodynamic cycle in Figure 2, the excess component of the binding free energy is obtained by the difference of the free energies of the two legs:

$$\Delta G_b^* = \Delta G_2 - \Delta G_1 \tag{10}$$

or, equivalently, by the value of the free energy profile at $\lambda = 1$ relative to $\lambda = 0$ (Figure 3, ATM panel).

Because the end states of ATM are similar to that of the PMF method summarized above, the two methods compute the same free energy of binding. However, each employs a different thermodynamic path (Figure 3). The PMF method progressively displaces the ligand from the binding site to the bulk along a physical path, whereas ATM employs an unphysical alchemical path, in which the ligand is displaced directly from the binding site to the solvent bulk.

2.3 SAMPL8 Systems

The chemical structures of the two hosts and five guests molecules are shown in Fig. 4. Both the hosts TEETOA and TEMOA are octaacids that carry a net charge of -8 at the pH value of 11.5 used in the experiment. The five guests, with the exception of the protonated G2 (namely G2P), are carboxylate derivatives that are also negatively charged at the same pH. The calculations in this work employed the initial host and guest structure files provided in the SAMPL8 dataset found at `https://github.com/samplchallenges/SAMPL8/tree-/master/host_guest/GDCC`.

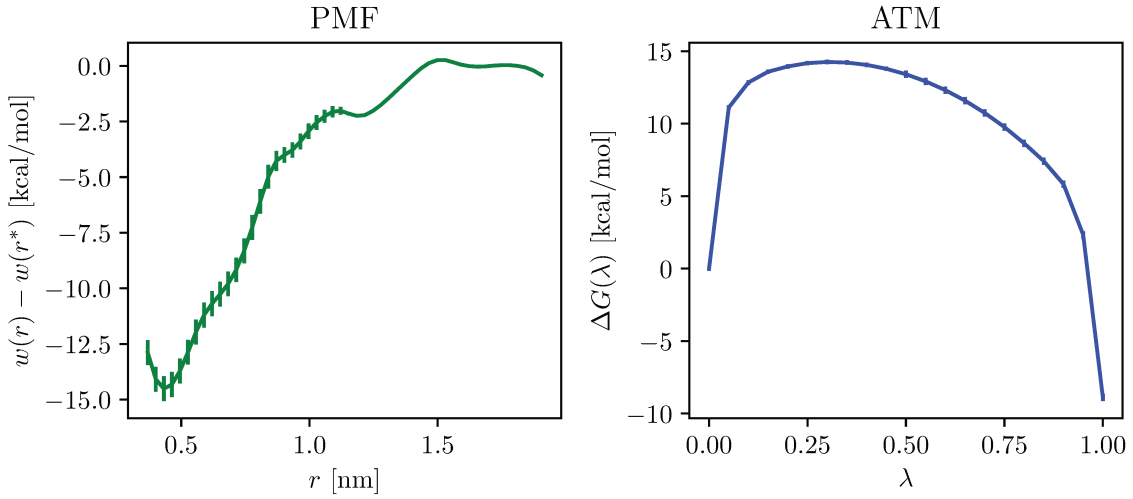

Fig. 3 Free energy profiles of the system TEETOA-G2P computed by the potential of mean force method (green) and the alchemical transfer method (blue). In the PMF plot, the x-axis represents the pulling distance in nanometers. The minimum of the curve corresponds to the free energy change of transferring the ligand from the solvent bulk to the receptor binding site along the physical pathway. For TEETOA-G2P this value is reported as $w(r_{\min}) - w(r^*)$ in Table 4. In the ATM plot, the x-axis represents the alchemical progress parameter λ. The excess binding free energy can be inferred by taking the difference between the potential energy of the unbound state ($\lambda = 0$), and that of the bound state ($\lambda = 1$).

2.4 Computational Details

The guests were manually docked to each host using Maestro (Schrödinger, Inc.) to render a set of host-guest molecular complexes that were then used to derive forcefield parameters with AmberTools. The complexes were assigned GAFF2/AM1-BCC[23] parameters and solvated in a TIP3P water box with a 12 Å solvent buffer and sodium counterions to balance the negative charge. The position and orientation of the host for each complex were restrained near the center of the box and along the box's diagonal with a flat-bottom harmonic potential of force constant 25.0 kcal/(mol Å^2) and a tolerance of 1.5 Å applied on the heavy atoms at the lower cup of the molecule (the first 40 atoms of the host as listed in the provided files). The generated Amber structure files were converted to Gromacs and Desmond formats using Intermol.[24] The systems were energy minimized and thermalized at 300 K prior to proceeding with the ATM and PMF calculations.

2.4.1 PMF Setup

The computation of the standard binding free energies using the PMF method involves the following steps:[13] (1) applying a harmonic restraint on the three Euler angles of the guest in the bound state to restrain guest orientation; (2) applying a harmonic restraint on the polar and azimuthal angles in spherical coordinates to restrain the guest center along a fixed axis when it binds/unbinds; (3) reversibly extracting the guest from the binding pocket along the chosen axis until it reaches the bulk region; (4) release the restraints on the guest center and guest orientation, which allows the guest to occupy the standard volume and rotate freely in the bulk solvent. The standard binding free energy is then obtained by summing up the reversible work associated with each of the above steps using Eq. (1).

The position and orientation of the guest relative to the host was controlled using coordinate systems which consisted of 3 reference atoms of the host (P1, P2, and P3) and 3 reference atoms of the guest (L1, L2, and L3).[25] For all the hosts, P1 was chosen to be the center of the bottom ring of each host and L1 the center of each guest molecule which lies approximately 4 Angstroms away from P1. The PMF was calculated along the P1-L1 distance using umbrella sampling with biasing potentials having a force constant of 1000 kJ/(mol nm^2). The three Euler angles and two polar and azimuthal angles were restrained using harmonic potentials with a force constant of 1,000 kJ/(mol rad^2) centered on the angles of the thermalized structures such that the guest is pulled straight out of the pocket of the host while minimizing collisions with the sidechains of the rim of the host. It is important to note that an unobstructed path is necessary for the guest's pull axis for the PMF method.

Equilibration (1.2 ns) and production (20 ns) umbrella sampling was then initiated over 20 umbrella windows to cover a distance of 4.0 to 18.0 Angstroms, i.e.

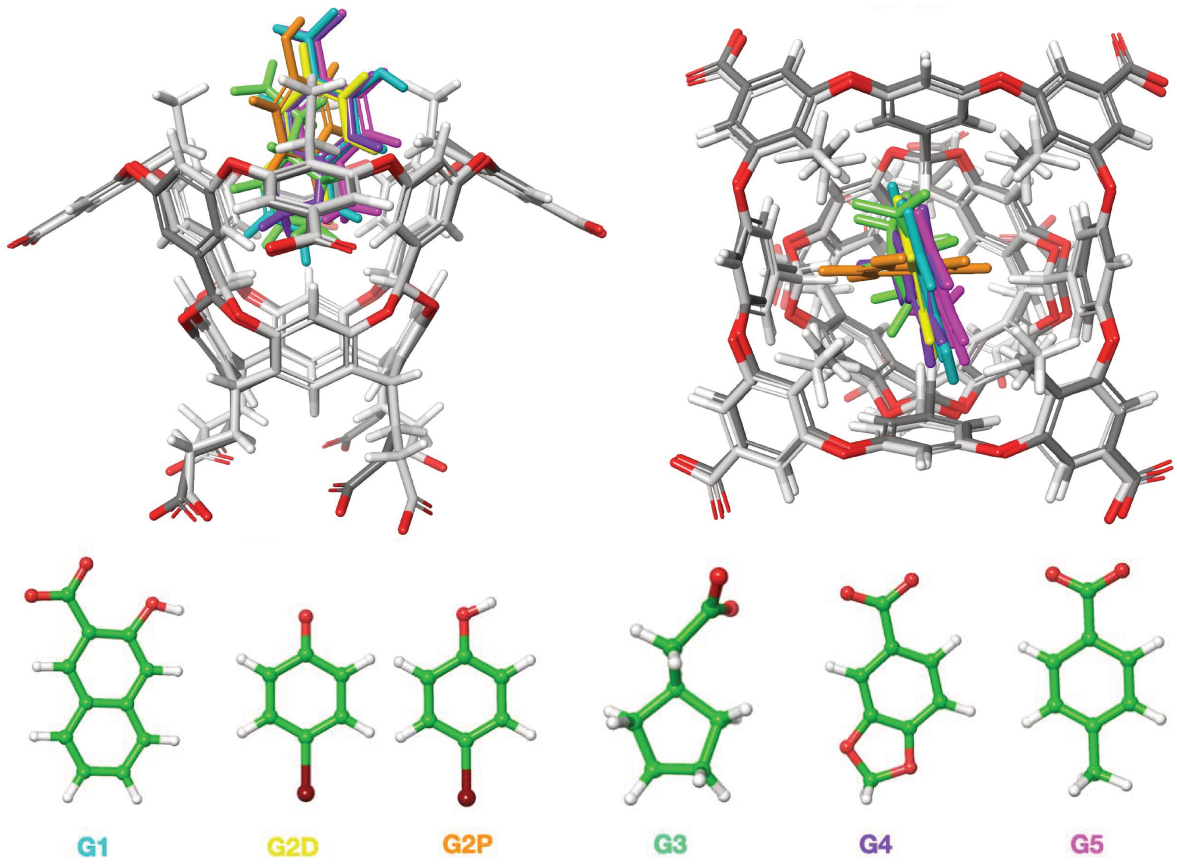

Fig. 4 Superimposed SAMPL8 host–guest systems considered in this study. The two hosts, tetramethyl octa acid (TEMOA) and tetraethyl octa acid (TEETOA), are shown in licorice representation, with light gray corresponding to TEETOA and dark gray to TEMOA. Both light and dark gray represent carbon atoms and red, oxygen atoms. The structures of the six guests in this study (bottom) are shown in ball-and-stick (CPK) representation, with the color of the label corresponding to their approximate location in the binding cavity of the hosts above. 2D designates deprotonated G2 and G2P, protonated G2. For the guests, green corresponds to carbon atoms, red oxygen atoms, brown to bromine atoms, and white hydrogen atoms.

from within the binding region to the bulk along the P1-L1 axis. WHAM analysis was used to generate the PMF and the corresponding uncertainties by bootstrapping. The free energy of releasing the angular restraints in the bulk and in the bound state were computed using BAR as implemented in GROMACS.[26]

2.4.2 ATM Setup

The magnitude of the translation vector h was set to 38 Å and pointing towards the corner of the solvent box with the aim to place the solvated ligand as far as possible from the host and its periodic images (Fig. 2). Beginning at the bound state at $\lambda = 0$, the systems were progressively annealed to the symmetric alchemical intermediate at $\lambda = 1/2$ during a 250 ps run using the ATM alchemical potential energy function for Leg 1 [Eq. (2)]. This step yields a suitable initial configuration of the system without severe unfavorable repulsive interactions at either end states of the alchemical path in order to facilitate Hamiltonian Replica Exchange that is subsequently conducted for each leg as described below.

In order to prevent large attractive interactions between opposite charges at small distances in nearly uncoupled states, polar hydrogen atoms with zero Lennard-Jones parameters were modified to $\sigma_{\mathrm{LJ}} = 0.1$ Å and $\epsilon_{\mathrm{LJ}} = 10^{-4}$ kcal/mol. [27] We established that the change in potential energy of the system in the unbound, bound, and symmetric intermediate states due to this modification of the Lennard-Jones parameters is below single floating point precision. Alchemical MD calculations were conducted with the OpenMM 7.3[28] MD engine and the SDM integrator plugin (`github.com/-Gallicchio-Lab/openmm_sdm_plugin.git`) using the

OpenCL platform. In order to maintain the temperature at 300 K, a Langevin thermostat with a time constant of 2 ps was implemented. For each ATM leg, Hamiltonian Replica Exchange in λ space was conducted every 5 ps with the ASyncRE software [29] that is customized for OpenMM and SDM (`github.com/Gallicchio-Lab/async_re-openmm.git`). Each leg employed 11 λ states uniformily distributed between $\lambda = 1/2$ and $\lambda = 0$ or $\lambda = 1$. All ATM calculations employed the soft-core perturbation energy with parameters $u_{\max} =$ 300 kcal/mol, $u_c = 100$ kcal/mol, and $a = 1/16$. A flat-bottom harmonic potential between the centers of mass of the host and the guest with a force constant of 25 kcal/mol Å^2 was applied for a distance greater than 4.5 Å to define the binding site region (V_{site}). The concentration-dependent term,

$$\Delta G^{\circ}_{\text{site}} = -k_B T \ln C^{\circ} V_{\text{site}} \tag{11}$$

equal in this case to 0.87 kcal/mol, which corresponds to 300 K temperature and the volume V_{site} of a sphere with a radius of 4.5 Å, was added to yield the standard binding free energy estimate.[18] Perturbation energy samples and trajectory frames were collected every 5 ps. Every replica was simulated for a minimum of 10 ns. For ATM, UWHAM was used to compute binding free energies and the corresponding uncertainties with the first 5 ns of the trajectory discarded.

2.4.3 Free Energy of Binding for Ligands in Multiple Protonation States

When multiple chemical species contribute to binding, we use the free energy combination formula[18]

$$\Delta G^{\circ}_b = -kT \ln \sum_i P_0(i) e^{-\beta \Delta G^{\circ}_b(i)}, \tag{12}$$

where $\Delta G^{\circ}_b(i)$ is the standard binding free energy for species i and $P_0(i)$ is the population of that species in the unbound state. In the case of an acid/base equilibrium with acidity constant

$$K_a = \frac{[A^-][H^+]}{[HA]} = \frac{[A^-]}{[HA]} 10^{-pH} = \alpha 10^{-pH}, \tag{13}$$

where $[\ldots]$ denotes concentration in molar units,

$$\alpha = 10^{pH - pKa}, \tag{14}$$

is the concentration ratio of the deprotonated and protonated forms, the population fraction of the deprotonated species is

$$P_0(A^-) = \frac{[A^-]}{[HA] + [A^-]} = \frac{\alpha}{1+\alpha} \tag{15}$$

and the population fraction of the protonated species is

$$P_0(HA) = \frac{[HA]}{[HA] + [A^-]} = 1 - P_0(A^-) = \frac{1}{1+\alpha}. \tag{16}$$

The populations of each protonation state of the ligands and the corresponding standard binding free energies $\Delta G^{\circ}_b(A^-)$ and $\Delta G^{\circ}_b(HA)$ are combined using Eq. (12) to obtain an estimate of the observed free energy of binding.

This strategy was employed for the guest G2, 4-bromophenol, which exists in two protonation states. A pH of 11.5, as indicated in the SAMPL8 GitHub site, and a pKa of 9.17 (`pubchem.ncbi.nlm.nih.gov/-compound/4-bromophenol`) was used to calculate the concentrations of the protonation states and combine them with the calculated binding free energies to yield a binding free energy estimate for G2 (see Table 5).

2.4.4 Structural Analysis

Water molecules were considered bound to the host binding cavity when the Euclidean distance between the water's oxygen atom and the center of mass of the host is at most 4 Å. A plot of the mean water count in both the TEETOA and TEMOA complexes with respect to the progress parameter λ can be found in Figure 7.

The conformations of TEETOA were classified as open and closed based on the dihedral angles between the ethyl sidechains and the phenyl ring of the benzoate group at the rim of the host (Figure 4). Each sidechain was classified as "facing out" or "facing in" the host cavity depending whether the dihedral angle was negative or positive, respectively. A conformation of TEETOA was classified as "open" if three or more sidechains were facing out and it was classified as closed if three or more sidechains were facing in. A plot of the population of closed conformations in the TEETOA complexes with respect to the progress parameter λ can be found in Figure 7.

Structural analysis was performed with VMD[30] after discarding the first half of each replica trajectory.

3 Results

The results are presented as follows. Table 1 summarizes the absolute binding free energy predictions from ATM and PMF submitted to the SAMPL8 organizers, compared to the experimental values which were disclosed to us only after submission. The results of the constituent calculations for each method that led to the binding free energy predictions are listed in Tables 3 and 4 for the ATM and PMF methods, respectively. These tables report the values of the free energy changes for each leg of the ATM calculations and the components of the PMF estimates, including those of the vibrational free energy and the restraint free energy that contribute to the overall PMF process. The free

Table 1 The PMF and ATM standard binding free energy blinded predictions submitted to SAMPL8 compared to the experimental values.

Complex	Experiment[a]	ATM[a]	PMF[a]
TEMOA-G1	-6.96 ± 0.2	-6.71 ± 0.3	-6.43 ± 0.4
TEMOA-G2	-8.41 ± 0.1	-9.90 ± 0.8	-9.37 ± 0.8
TEMOA-G3	-5.78 ± 0.1	-8.26 ± 0.3	-8.71 ± 0.4
TEMOA-G4	-7.72 ± 0.1	-8.63 ± 0.3	-8.79 ± 0.6
TEMOA-G5	-6.67 ± 0.1	-7.70 ± 0.3	-8.15 ± 0.8
TEETOA-G1	-4.49 ± 0.2	-1.07 ± 0.3	-1.38 ± 0.8
TEETOA-G2	-5.16 ± 0.1	-4.76 ± 0.3	-6.22 ± 1.8
TEETOA-G3	NB	-1.65 ± 0.3	-1.42 ± 0.8
TEETOA-G4	-4.47 ± 0.2	-2.51 ± 0.3	-2.25 ± 0.8
TEETOA-G5	-3.32 ± 0.1	-2.82 ± 0.3	-3.36 ± 1.9

[a] In kcal/mol.

energy analysis for the protonated and deprotonated species implicated in the complexes of the G2 guest is illustrated in Table 5. Finally, Figure 7 presents structural data to demonstrate that TEETOA undergoes a conformational reorganization upon guest binding with the binding cavity remaining essentially dry throughout the process, which is contrary to the conventional water displacement process observed for guest binding to TEMOA.

3.1 Absolute Binding Free Energy Estimates by ATM and PMF

The binding free energy estimates obtained from the two complementary computational methods, ATM and PMF, are in very good agreement with an Pearson R value of 0.965 and an RMSE value of 0.60 kcal/mol. In addition, the ranking of the binding free energies of the complexes between the ATM and PMF datasets is in perfect agreement. Both methods consistently estimated the complex with the most favorable binding free energy to be TEMOA-G2, with a free energy value of -9.90 kcal/mol predicted by ATM and -9.37 kcal/mol by PMF. The least favorable binding free energy was predicted for the complex TEETOA-G1 by both methods, -1.07 kcal/mol by ATM and -1.38 kcal/mol by PMF. Both methods predicted that all of the guests bind TEMOA more favorably than TEETOA.

All of the carboxylic acid guests were modeled as ionic. We modeled both protonation states of the G2 guest (Tables 3 and 4) and combined the corresponding binding free energies using the experimental pKa of the guest (Table 5). With a discrepancy of 2.77 kcal/mol, the deprotonated G2 molecule (hereon G2D) yielded the most divergent binding free energy estimate between the ATM and PMF datasets. Nevertheless, since this protonation state is found to contribute little to binding (Table 5), the observed discrepancy did not affect significantly the correspondence between the two sets of SAMPL8 binding free energy predictions.

The molecular dynamics trajectories consistently yielded the expected binding mode of the guests to the TEMOA and TEETOA hosts. The polar/ionic end of the guests is oriented towards the water solvent while the more non-polar end of the molecule is inserted into the binding cavity of the hosts (Figure 4). In the complexes, the ethyl sidechains of the TEETOA host point outward extending further the host binding cavity and the surface of contact between the guests and the hosts. In the apo state, however, the ethyl sidechains are observed to be mostly folded into the TEETOA cavity (Fig. 6). We hypothesize that the conformational reorganization of TEETOA, the lack of favorable water expulsion, and the poorer hydration of the bound guests are responsible for the weaker binding propensity of TEETOA relative to TEMOA. We intend to investigate further these aspects of the binding mechanism in future work.

ATM and PMF both predict that G2D is one of the weakest binders for TEMOA and TEETOA (Tables 3 and 4). G2D is expected to be frustrated in the bound state because the bromine atom prefers to be in the cavity of the host, whereas the oxide group strongly prefers to remain hydrated (Figure 4). The side chains of both hosts prevent the hydration of the negative oxygen atom. This steric hindrance is especially evident in TEETOA, which possesses four ethyl groups on its outer ring. Due to its poor binding affinity, the deprotonated G2D is not predicted to contribute significantly to binding despite its higher concentration in solution at the experimental pH. Conversely, due to its smaller desolvation penalty, both the PMF and ATM methods indicate that protonated G2 (hereon G2P) is the strongest binder in the set for both TEMOA and TEETOA (Tables 3 and 4). G2P is in fact predicted to be the dominant species for binding even after factoring in the protonation penalty at the experimental pH (11.5).

The ATM free energy components ΔG_1 and ΔG_2 for each leg of the ionic guests (Table 3), being in the 40 to 50 kcal/mol range, are significantly larger in magnitude than the resulting binding free energies. These free energies correspond to the reversible work to reach the alchemical intermediate state in which the guest interacts with both the receptor and the solvent bulk intermediates. The high free energy of the alchemical intermediate relative to the bound and solvated states suggests that the ionic group can not be properly accommodated to simultaneously interact effectively with both environments. This hypothesis is confirmed by the much smaller ATM leg free energies for the neutral

Table 2 Agreement metrics including root mean square error (RMSE), Pearson's correlation coefficient of determination (R), slope of the linear regression (m), and Kendall rank order correlation coefficient (τ) between the computed binding free energies and the experimental measurements.

	RMSE	R	m	τ
ATM/PMF	0.60	0.99	1.05	1.00
Exp./ATM	1.71	0.89	1.65	0.69[a]
Exp./PMF	1.79	0.83	1.50	0.69[a]

[a] TEETOA-G3, a non-binder experimentally, was included in the τ calculation as the weakest complex.

G2P guest. While large, the ATM leg free energies of the ionic guests are expected to be significantly smaller than those that would have obtained in a double decoupling calculation[13] that would involve displacing the guests into vacuum where hydration interactions are completely removed. The statistical uncertainties of the ATM binding free energy estimates, generally around 1/3 of a kcal/mol, are reasonably small.

As seen from Table 4 the PMF calculation of the binding free energy is largely determined by the the reversible work of releasing the restraints in the binding site (2nd column) and work of ligand extraction (3rd column). The other two terms are either very small in magnitude (the vibrational free energy difference between the bound and bulk ligand, 4th column) or constant (the free energy of releasing the restraints in the bulk, 5th column). The PMF binding free energy estimates (Table 4) come with somewhat larger uncertainties than ATM. The source of uncertainties is approximately equally split between the reversible work of releasing the restraints (2nd column) and work of ligand extraction (3rd column). However, in some cases (TEETOA-G2 and TEETOA-G5) the uncertainty of the work of extraction is particularly large and probably indicative of sampling bottlenecks at intermediate stages of the extraction process for this host. According to Table 4, the statistical errors in the free energy for ligand extraction for TEETOA complexes are consistently larger than those for the TEMOA complexes. This difference is likely attributable to the guest binding -induced reorganization of the four ethyl side chains in the TEETOA molecule, which is more difficult to sample reversibly in the umbrella sampling for computing the reversible work.

3.2 Calculated Free Energy Estimates Relative to Experimental Measurements

The two computational methods employed in this work reproduced the experimental binding free energy estimates relatively well, particularly more so for the TEMOA host than for the TEETOA host (Table 1). Even though the PMF and ATM binding free energy estimates for the deprotonated form of G2 differ significantly, both methods correctly predict TEMOA-G2 as the highest affinity complex in the set with good quantitative accuracy in the overall binding free energy predictions (-8.41 kcal/mol experimentally compared to calculated -9.90 and -9.37 kcal/mol from ATM and PMF, respectively.) Concomitantly, both methods correctly predict relatively weak absolute binding free energies of -1.65 kcal/mol and -1.42 kcal/mol, respectively, for TEETOA-G3 which is an experimental non-binder. Excluding TEETOA-G3, the least favorable binding affinity measurement was obtained for TEETOA-G5, which is correctly scored as one of the weakest complex by both computational methods. Overall, despite the the narrow range of the moderate binding free energies, the computational rankings based on the binding free energies are in good agreement with the experimental rankings with a Kendall rank-order correlation coefficient of 0.69. (Table 2.)

As illustrated in Figure 5 the calculated binding free energies are highly correlated to the experimental values with Pearson R correlation coefficients of 89% and 83% for ATM and PMF, respectively (Table 2). The calculations are also in reasonable quantitative agreement with the experimental measurements with RMSE deviations of 1.71 kcal/mol for ATM and 1.79 kcal/mol for PMF. Interestingly, the computational models tend to overestimate the binding affinity of the TEMOA complexes and to underestimate those of the complexes with TEETOA. The largest deviation occurs for TEETOA-G1 which has a moderate observed binding free energy of -4.47 kcal/mol, which is underestimated by the computational predictions by around -1 kcal/mol. A large deviation, but in the opposite direction, is also observed for TEMOA-G3 (-5.78 kcal/mol experimentally compared to -8.26 and -8.71 kcal/mol computationally) (Table 1). A poor prediction for this complex was expected based on previous efforts with the GAFF/AM1-BCC force field with TIP3P solvation used here.[31]

In summary, the predictions reported here were scored as among the best of the SAMPL8 GDCC challenge and second only to those obtained with the more accurate AMOEBA force field[32] (`github.com/sampl-`

Table 3 ATM absolute binding free energy estimates for the TEMOA and TEETOA complexes.

Complex	ΔG_1 [a]	ΔG_2 [a]	$\Delta G^\circ_{\text{site}}$ [a]	ΔG°_b [a]
TEMOA-G1	53.27 ± 0.21	45.69 ± 0.21	0.87	-6.71 ± 0.30
TEMOA-G2D	42.37 ± 0.18	35.48 ± 0.21	0.87	-6.02 ± 0.28
TEMOA-G2P	22.57 ± 0.27	8.60 ± 0.78	0.87	-13.10 ± 0.83
TEMOA-G3	56.42 ± 0.18	47.29 ± 0.18	0.87	-8.26 ± 0.25
TEMOA-G4	53.13 ± 0.24	43.63 ± 0.18	0.87	-8.63 ± 0.30
TEMOA-G5	53.49 ± 0.24	44.92 ± 0.18	0.87	-7.70 ± 0.30
TEETOA-G1	51.65 ± 0.27	49.71 ± 0.21	0.87	-1.07 ± 0.34
TEETOA-G2D	42.26 ± 0.24	39.83 ± 0.27	0.87	-1.57 ± 0.36
TEETOA-G2P	22.31 ± 0.24	13.48 ± 0.15	0.87	-7.95 ± 0.28
TEETOA-G3	55.31 ± 0.24	52.79 ± 0.18	0.87	-1.65 ± 0.30
TEETOA-G4	52.28 ± 0.24	48.90 ± 0.18	0.87	-2.51 ± 0.30
TEETOA-G5	53.58 ± 0.21	49.89 ± 0.18	0.87	-2.82 ± 0.28

[a] In kcal/mol; uncertainties are reported at the 3σ confidence level.

Table 4 PMF absolute free energy estimates for TEMOA and TEETOA complexes.

Complex	$-\Delta G^{\text{bound}}_{\text{restr}}$ [a]	$[w(r_{\min}) - w(r^*)]$ [a]	ΔG_{vibr} [a]	$\Delta G^{\text{bulk}}_{\text{restr}}$ [a]	ΔG°_b [a]
TEMOA-G1	-4.09 ± 0.23	-12.27 ± 0.36	0.24	9.69	-6.43 ± 0.43
TEMOA-G2D	-2.05 ± 0.33	-11.01 ± 0.18	0.12	9.69	-3.25 ± 0.38
TEMOA-G2P	-5.31 ± 0.78	-17.12 ± 0.21	0.17	9.69	-12.57 ± 0.81
TEMOA-G3	-5.61 ± 0.30	-12.83 ± 0.30	0.04	9.69	-8.71 ± 0.42
TEMOA-G4	-5.00 ± 0.47	-13.72 ± 0.36	0.24	9.69	-8.79 ± 0.59
TEMOA-G5	-5.36 ± 0.81	-12.74 ± 0.15	0.26	9.69	-8.15 ± 0.82
TEETOA-G1	-3.76 ± 0.60	-7.60 ± 0.54	0.28	9.69	-1.38 ± 0.81
TEETOA-G2D	-5.50 ± 0.84	-5.25 ± 2.73	0.20	9.69	-0.86 ± 2.86
TEETOA-G2P	-4.85 ± 0.57	-14.51 ± 1.68	0.25	9.69	-9.42 ± 1.77
TEETOA-G3	-3.70 ± 0.24	-7.36 ± 0.81	-0.05	9.69	-1.42 ± 0.84
TEETOA-G4	-3.77 ± 0.12	-8.39 ± 0.75	0.22	9.69	-2.25 ± 0.76
TEETOA-G5	-4.47 ± 0.06	-8.81 ± 1.89	0.23	9.69	-3.36 ± 1.89

[a] In kcal/mol; uncertainties are reported at the 3σ confidence level.

challenges/SAMPL8/blob/master/host_guest/Analysis/Ranked_Accuracy).

3.3 Mechanisms of Binding to TEMOA and TEETOA

The results of the structural analysis of the ATM molecular dynamics trajectories shown in Figure 7 (similar behavior was observed in the PMF trajectories) indicate that TEMOA and TEETOA display very different binding mechanisms. Guest binding to TEMOA displaces water molecules from the cavity of the host (Figure 6A vs. 6B). In contrast, when unbound, the binding cavity of TEETOA is closed by the ethyl sidechains of the host (Figure 7C) and rather than displacing bound water, these sidechains are displaced upon guest binding (7C)). To illustrate this behavior quantitatively, the average number of water molecules present in the binding cavities of TEMOA and TEETOA were computed as a function of the alchemical progress parameter λ (see Computational Details) shown in Figure 7 (center and bottom panels). When not bound to the guest ($\lambda = 0$), TEMOA holds 2 to 3 water molecules, which are rapidly displaced into the bulk as the guest is coupled to the host by increasing λ. In contrast, at $\lambda = 0$ TEETOA holds almost no water molecules and the binding cavity remains essentially dry throughout the binding process (Figure 7, center panel) by the concomitant transition (Figure 7, top panel) from closed conformation of the host, where three or more sidechains are folded into the binding cavity (Figure 6C) to open conformations, in which the sidechains are rotated away from the binding cavity to accommodate the guest (Figure 6D).

4 Discussion and Conclusions

This study employed two independent binding free energy approaches, the newly developed alchemical transfer method (ATM)[27, 12] and the well established PMF physical pathway method[13] to blindly predict the absolute binding affinities of the host-guest systems as part of the SAMPL8 GDCC challenge. The SAMPL series of community challenges have consistently yielded high-quality datasets to test computational models of binding,[1, 2, 3, 9, 10, 11] and we decided to use them

Table 5 Binding free energy contributions of the protonated (G2P) and deprotonated (G2D) form of the G2 guest to the ATM and PMF binding free estimates.

	TEMOA-G2/ATM	TEMOA-G2/PMF	TEETOA-G2/ATM	TEETOA-G2/PMF
ΔG_b°(G2P)[a]	-13.10 ± 0.83	-12.57 ± 0.81	-7.95 ± 0.28	-9.42 ± 1.77
P_0(G2P)	4.66×10^{-3}	4.66×10^{-3}	4.66×10^{-3}	4.66×10^{-3}
P_0(G2P)$e^{-\beta \Delta G_b^\circ(\mathrm{G2P})}$	1.65×10^{7}	6.77×10^{7}	2.92×10^{3}	3.42×10^{4}
ΔG_b°(G2D)[a]	-6.02 ± 0.28	-3.25 ± 0.38	-1.57 ± 0.36	-0.86 ± 2.86
P_0(G2D)	0.995	0.995	0.995	0.995
P_0(G2D)$e^{-\beta \Delta G_b^\circ(\mathrm{G2D})}$	2.43×10^{4}	232	13.6	4.22
ΔG_b°[a]	-9.90 ± 0.83	-9.37 ± 0.81	-4.76 ± 0.28	-6.22 ± 1.8

[a] In kcal/mol.

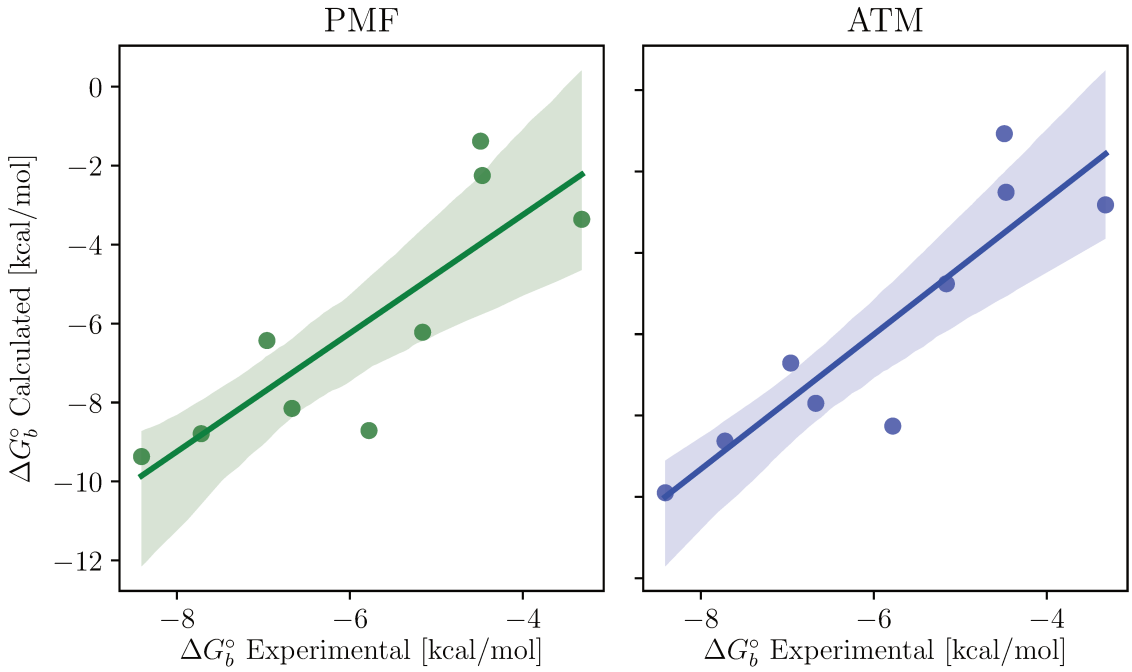

Fig. 5 Combined TEMOA and TEETOA binding free energy predictions compared to the experimental binding free energies for (left, green) the PMF predictions and (right, blue) the ATM predictions. The lines are linear regressions of the respective data.

here to stringently validate the ATM and PMF methods in an unbiased fashion.

Despite their radical differences in spirit and in practice, we find that the calculated binding affinities from the two methods are in remarkable quantitative agreement with an RMSE value of only 0.6 kcal/mol and an R value of 99%. This level of agreement, well within statistical fluctuations, gives high confidence in not only the theoretical foundations of ATM and PMF, but also in the correctness of implementation of each approach. The level of consistency of the computational methods also adds confidence that their predictions are unbiased and primarily reflective of the force field model.

We find that the standard GAFF/AM1-BCC/TIP3P model employed here tends to overestimate the binding free energies of strongly bound complexes while it tends to underestimate those of more weakly bound complexes, as also indicated by the larger than one slope of the linear regressions (Tables 1, 2). While it may be a result, in this case, of specific aspects of the TEMOA and TEETOA hosts, this trend has been generally observed with this force field combination.[31] The more accurate AMOEBA force field[32] appears to correctly predict these trends (`github.com/samplchallenges/-SAMPL8/blob/master/host_guest/Analysis/Ranked_Accuracy`).

The stringent blinded test conducted in this work is a further validation of the ATM binding free energy method that we have recently proposed.[12] ATM, implemented on top of the versatile OpenMM molecular

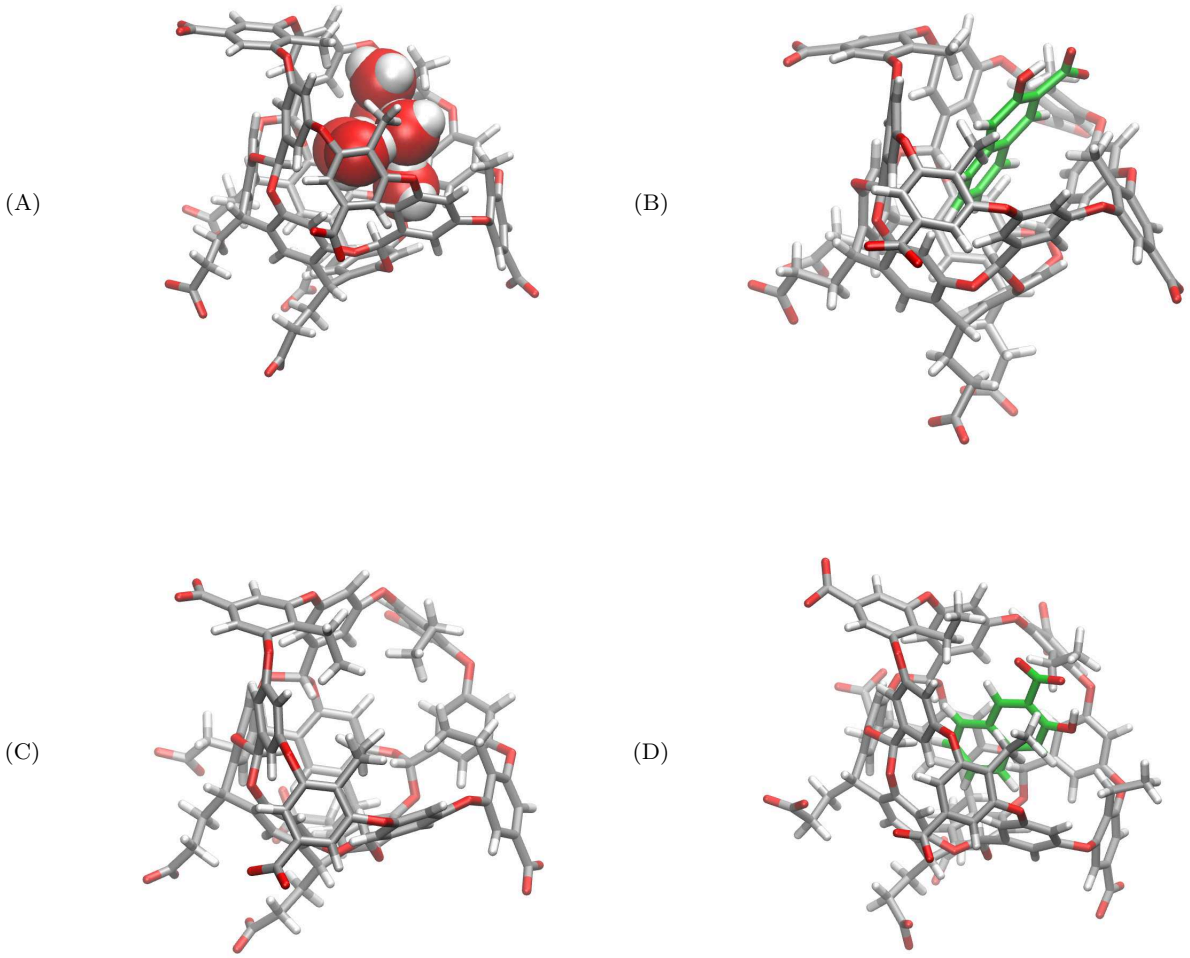

Fig. 6 Representative structures of (A) free TEMOA with water molecules present in the binding cavity, (B) TEMOA bound to the G1 guest, (C) free TEETOA with the ethyl sidechains in the closed conformation, and (D) TEETOA bound to the G1 guest with the ethyl sidechains in the open conformation. Guest binding to TEMOA displaces bound water molecules (A to B), whereas guest binding to TEETOA causes a reorganization of the host from closed to open without solvent displacement (C to D).

dynamics engine,[28] promises to provide an accurate and streamlined route to absolute[12] and relative binding free calculations.[33] While alchemical, ATM makes use of a single simulation system, which is similar to the PMF pathway method,[13] and it avoids problematic vacuum intermediates and the splitting of the alchemical path into electrostatic and non-electrostatic transformations. ATM also does not require soft-core pair potentials and modifications of energy routines, and can be easily implemented as a controlling routine on top of existing force routines of MD engines.

The distinct binding mechanisms observed for the TEMOA and the TEETOA hosts are a useful prototype for the modeling of complex protein-ligand binding equilibria coupled to dehydration and conformational reorganization processes. Binding to the rigid TEMOA host follows a conventional water displacement mechanism in which two to three water molecules bound to the host cavity are replaced by the ligand and transferred to the solvent bulk. Water displacement significantly favors guest binding to TEMOA[11] due to the high free energy of the interface between the host cavity and the bound water molecules relative to the solvent bulk. For the more flexible TEETOA host, on the other hand, water displacement does not occur because, in the absence of the guest, its cavity is occupied by the ethyl sidechains rather than by water molecules. Hence, guest binding to TEETOA involves the displacement of

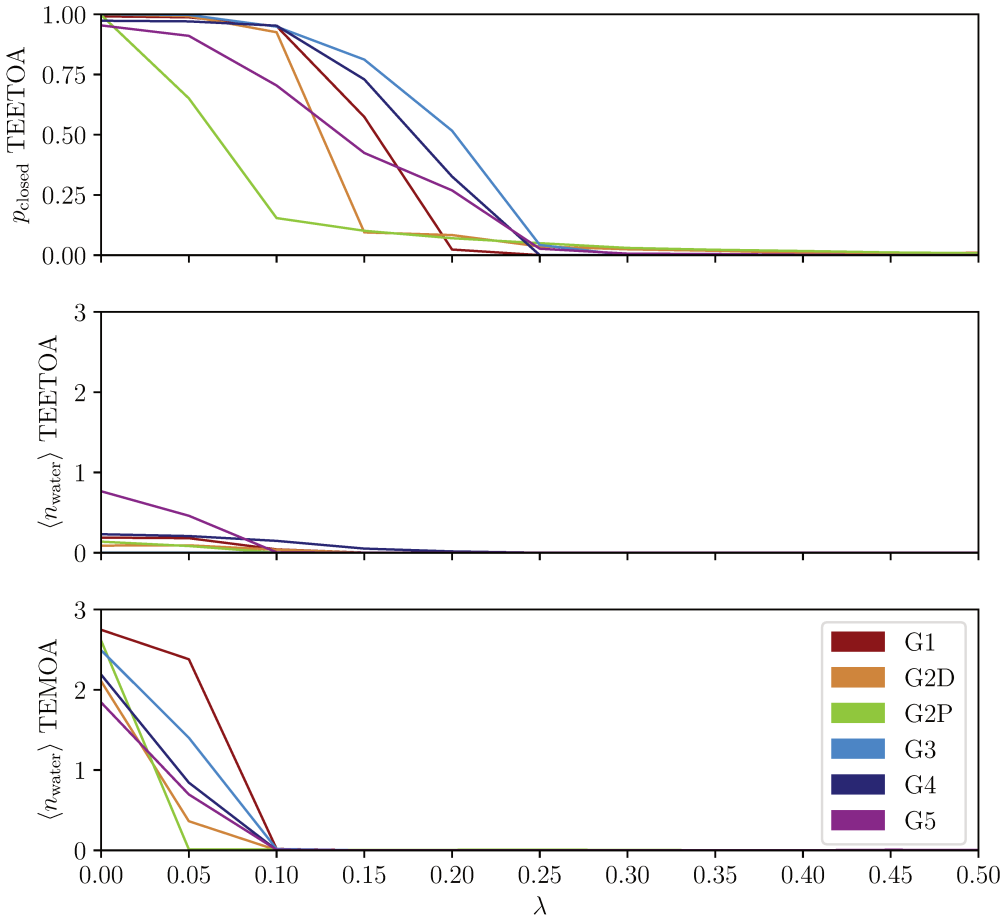

Fig. 7 (top) The population of closed conformations of TEETOA, (middle) the mean number of water molecules enclosed within TEETOA's binding cavity, and (bottom) the mean number of water molecules enclosed within TEMOA's binding cavity, respectively, as a function of the progress parameter λ. $\lambda = 0$ represent the unbound state.

the host sidechains from the cavity rather than that of water, a mechanism that does not contribute to an overall free energy gain of water expulsion that is observed in TEMOA. Evidently, TEETOA achieves a lower free energy state by undergoing an internal conformational reorganization to displace the bound water molecules. Consequently, it lacks the extra contributions towards binding provided water displacement and it has weaker affinity for the guest binding relative to TEMOA (Table 1).

Cryptic pockets induced by ligand binding, such as the one observed for TEETOA, are often observed in proteins. Cryptic pockets are unstable in the absence of the ligand because, as in TEETOA, there is a greater free energy gain in collapsing them rather than filling them with water. Being able to predict such occurrences would clearly be very valuable in the design of drug compounds. However, the balance between the various thermodynamic driving forces at play (dehydration, internal reorganization, and the formation of intramolecular contacts) can be subtle and very difficult to model accurately.[21] Studying these phenomena in host-guest model systems is a valuable opportunity to test theories and models and apply them to flexible protein receptors.

In summary, this work provides a rare blinded and stringent test of binding free energy models. It shows that the application of sound statistical mechanics theories of binding and careful modeling of chemical systems can lead to reliable predictions limited only by the quality of the force field model. The systems chosen for this round of the SAMPL series of challenges provide a valuable opportunity to study the effects of hydration and conformational reorganization equilibria coupled to binding.

5 Acknowledgements

We acknowledge support from the National Science Foundation (NSF CAREER 1750511 to E.G.) and National Institutes of Health (R01GM100946 to T.K.). Molecular simulations were conducted on the Comet and Expanse GPU clusters at the San Diego Supercomputing Center supported by NSF XSEDE award TG-MCB150001. We appreciate the National Institutes of Health for its support of the SAMPL project via R01GM124270 to David L. Mobley.

6 Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Matthew T Geballe, A Geoffrey Skillman, Anthony Nicholls, J Peter Guthrie, and Peter J Taylor. The SAMPL2 blind prediction challenge: introduction and overview. *J. Comp. Aided Mol. Des.*, 24(4):259–279, 2010.
2. David L Mobley, Shuai Liu, Nathan M Lim, Karisa L Wymer, Alexander L Perryman, Stefano Forli, Nanjie Deng, Justin Su, Kim Branson, and Arthur J Olson. Blind prediction of hiv integrase binding from the SAMPL4 challenge. *J. Comp. Aided Mol. Des.*, pages 1–19, 2014.
3. Martin Amezcua, Léa El Khoury, and David L Mobley. SAMPL7 host–guest challenge overview: assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations. *J. Comp.-Aid. Mol. Des.*, 35(1):1–35, 2021.
4. David L Mobley and Michael K Gilson. Predicting binding free energies: frontiers and benchmarks. *Ann. Rev. Bioph.*, 46:531–558, 2017.
5. William L Jorgensen. Efficient drug lead discovery and optimization. *Acc Chem Res*, 42:724–733, 2009.
6. Kira A Armacost, Sereina Riniker, and Zoe Cournia. Novel directions in free energy methods and applications, 2020.
7. E. Gallicchio and R. M. Levy. Prediction of SAMPL3 host-guest affinities with the binding energy distribution analysis method (BEDAM). *J. Comp. Aided Mol. Design.*, 25:505–516, 2012.
8. Emilio Gallicchio, Haoyuan Chen, He Chen, Michael Fitzgerald, Yang Gao, Peng He, Malathi Kalyanikar, Chuan Kao, Beidi Lu, Yijie Niu, Manasi Pethe, Jie Zhu, and Ronald M Levy. BEDAM binding free energy predictions for the SAMPL4 octa-acid host challenge. *J. Comp. Aided Mol. Des.*, 29:315–325, 2015.
9. Emilio Gallicchio, Nanjie Deng, Peng He, Alexander L. Perryman, Daniel N. Santiago, Stefano Forli, Arthur J. Olson, and Ronald M. Levy. Virtual screening of integrase inhibitors by large scale binding free energy calculations: the SAMPL4 challenge. *J. Comp. Aided Mol. Des.*, 28:475–490, 2014.
10. Nanjie Deng, William F Flynn, Junchao Xia, RSK Vijayan, Baofeng Zhang, Peng He, Ahmet Mentes, Emilio Gallicchio, and Ronald M Levy. Large scale free energy calculations for blind predictions of protein–ligand binding: the d3r grand challenge 2015. *J. Comp.-Aided Mol. Des.*, 30(9):743–751, 2016.
11. Rajat Kumar Pal, Kamran Haider, Divya Kaur, William Flynn, Junchao Xia, Ronald M. Levy, Tetiana Taran, Lauren Wickstrom, Tom Kurtzman, and Emilio Gallicchio. A combined treatment of hydration and dynamical effects for the modeling of host-guest binding thermodynamics: The SAMPL5 blinded challenge. *J. Comp. Aided Mol. Des.*, 31:29–44, 2016.
12. Joe Z Wu, Solmaz Azimi, Sheenam Khuttan, Nanjie Deng, and Emilio Gallicchio. Alchemical transfer approach to absolute binding free energy estimation. *J. Chem. Theory Comput.*, 17:3309, 2021.
13. Nanjie Deng, Di Cui, Bin W Zhang, Junchao Xia, Jeffrey Cruz, and Ronald Levy. Comparing alchemical and physical pathway methods for computing the absolute binding free energy of charged ligands. *Phys. Chem. Chem. Phys.*, 20(25):17081–17092, 2018.
14. Paolo Suating, Thong T Nguyen, Nicholas E Ernst, Yang Wang, Jacobs H Jordan, Corinne LD Gibb, Henry S Ashbaugh, and Bruce C Gibb. Proximal charge effects on guest binding to a non-polar pocket. *Chemical Science*, 11(14):3656–3663, 2020.
15. Paweł Śledź and Amedeo Caflisch. Protein structure-based drug design: from docking to molecular dynamics. *Curr. Op. Struct. Biol.*, 48:93–102, 2018.
16. Thomas Seidel, Oliver Wieder, Arthur Garon, and Thierry Langer. Applications of the pharmacophore concept in natural product inspired drug design. *Molecular Informatics*, 39(11):2000059, 2020.
17. M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.*, 72:1047–1069, 1997.

18. Emilio Gallicchio and Ronald M Levy. Recent theoretical and computational advances for modeling protein-ligand binding affinities. *Adv. Prot. Chem. Struct. Biol.*, 85:27–80, 2011.
19. Zoe Cournia, Bryce K Allen, Thijs Beuming, David A Pearlman, Brian K Radak, and Woody Sherman. Rigorous free energy simulations in virtual screening. *Journal of Chemical Information and Modeling*, 2020.
20. Emilio Gallicchio. Free energy-based computational methods for the study of protein-peptide binding equilibria. In Thomas Simonson, editor, *Computational Peptide Science: Methods and Protocols*, Methods in Molecular Biology. Springer Nature, 2021.
21. Peng He, Sheila Sarkar, Emilio Gallicchio, Tom Kurtzman, and Lauren Wickstrom. Role of displacing confined solvent in the conformational equilibrium of β-cyclodextrin. *J. Phys. Chem. B*, 123(40):8378–8386, 2019.
22. Zhiqiang Tan, Emilio Gallicchio, Mauro Lapelosa, and Ronald M. Levy. Theory of binless multi-state free energy estimation with applications to protein-ligand binding. *J. Chem. Phys.*, 136:144102, 2012.
23. Xibing He, Shuhan Liu, Tai-Sung Lee, Beihong Ji, Viet H Man, Darrin M York, and Junmei Wang. Fast, accurate, and reliable protocols for routine calculations of protein–ligand binding affinities in drug design projects using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega*, 5(9):4611–4619, 2020.
24. Michael R Shirts, Christoph Klein, Jason M Swails, Jian Yin, Michael K Gilson, David L Mobley, David A Case, and Ellen D Zhong. Lessons learned from comparing molecular dynamics engines on the sampl5 dataset. *J. Comp.-Aid. Mol. Des.*, 31(1):147–161, 2017.
25. S Boresch, F Tettinger, M Leitgeb, and M Karplus. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B*, 107:9535–9551, 2003.
26. Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R Shirts, Jeremy C Smith, Peter M Kasson, David van der Spoel, Berk Hess, and Erik Lindahl. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29:845–854, 2013.
27. S Khuttan, Solmaz Azimi, Joe Z Wu, and E Gallicchio. Alchemical transformations for concerted hydration free energy estimation with explicit solvation. *J. Chem. Phys*, 154:054103, 2021.
28. Peter Eastman, Jason Swails, John D Chodera, Robert T McGibbon, Yutong Zhao, Kyle A Beauchamp, Lee-Ping Wang, Andrew C Simmonett, Matthew P Harrigan, Chaya D Stern, et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comp. Bio.*, 13(7):e1005659, 2017.
29. Emilio Gallicchio, Junchao Xia, William F Flynn, Baofeng Zhang, Sade Samlalsingh, Ahmet Mentes, and Ronald M Levy. Asynchronous replica exchange software for grid and heterogeneous computing. *Computer Physics Communications*, 196:236–246, 2015.
30. William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
31. Andrea Rizzi, Steven Murkli, John N McNeill, Wei Yao, Matthew Sullivan, Michael K Gilson, Michael W Chiu, Lyle Isaacs, Bruce C Gibb, David L Mobley, et al. Overview of the SAMPL6 host–guest binding affinity prediction challenge. *J. Comp.-Aid. Mol. Des.*, 32(10):937–963, 2018.
32. Yuanjun Shi, Marie L Laury, Zhi Wang, and Jay W Ponder. Amoeba binding free energies for the SAMPL7 trimertrip host–guest challenge. *J. Comp.-Aid. Mol. Des.*, 35(1):79–93, 2021.
33. Solmaz Azimi, Sheenam Khuttan, Joe Z. Wu, Rajat Pal, and Emilio Gallicchio. Relative binding free energy calculations for ligands with diverse scaffolds with the alchemical transfer method. *ArXiv Preprint*, 2107.05153, 2021.