

Blinded predictions of host-guest standard free energies of binding in the SAMPL5 challenge

Stefano Bosisio¹ · Antonia S. J. S. Mey¹ · Julien Michel¹

Received: 19 June 2016 / Accepted: 27 July 2016 / Published online: 8 August 2016
© Springer International Publishing Switzerland 2016

Abstract In the context of the SAMPL5 blinded challenge standard free energies of binding were predicted for a dataset of 22 small guest molecules and three different host molecules octa-acids (OAH and OAMe) and a cucurbituril (CBC). Three sets of predictions were submitted, each based on different variations of classical molecular dynamics alchemical free energy calculation protocols based on the double annihilation method. The first model (*model A*) yields a free energy of binding based on computed free energy changes in solvated and host-guest complex phases; the second (*model B*) adds long range dispersion corrections to the previous result; the third (*model C*) uses an additional standard state correction term to account for the use of distance restraints during the molecular dynamics simulations. *Model C* performs the best in terms of mean unsigned error for all guests (MUE $3.2 < 3.4 < 3.6$ kcal mol⁻¹—95 % confidence interval) for the whole data set and in particular for the octa-acid systems (MUE $1.7 < 1.9 < 2.1$ kcal mol⁻¹). The overall correlation with experimental data for all models is encouraging (R^2 $0.65 < 0.70 < 0.75$). The correlation between experimental and computational free energy of binding ranks as one of the highest with respect to other entries in the challenge. Nonetheless the large MUE for the best performing model highlights systematic errors, and submissions from other groups fared better with respect to this metric.

Keywords SAMPL5 · Binding free energies · Host-guest systems

Introduction

An accurate and reliable computational prediction of binding affinities of small molecules binding to larger molecules, such as proteins, remains a major objective of computer simulations for molecular design [1]. In order to assess state-of-the-art tools for computational predictions of thermodynamic properties of binding the Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) challenge was formulated almost 10 years ago [2–4]. The goal of SAMPL5, as in previous years, was to compare different computational approaches in blinded challenges for different properties. This report is concerned with host-guest standard binding free energies, as predicted by our group.

Host-guest systems can be regarded as a toy model for protein ligand systems and represent a good play ground for testing the accuracy of thermodynamic property predictions. Various computational methods are available to compute free energies of binding of the host-guest systems [5–8]. Previous SAMPL challenges have featured potential energy functions ranging from quantum chemical [9–11] to molecular mechanical approaches [12, 13]. Molecular dynamics (MD) or Monte Carlo (MC) simulations are frequently carried out to estimate the ensemble averages that yield standard free energies of binding. These methodologies still face three major problems: the sampling problem [14], the translation of host-guest systems into force field terms [15] and the presence of finite size effects [16]. Different approximations lead to various ways of estimating free energies of binding from molecular

✉ Julien Michel
mail@julienmichel.net

¹ EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, UK

simulations trajectories, e.g. free energy perturbations (FEP) [17], finite difference thermodynamic integration (FDTI) [18], or end-states only variants such as MM-PBSA [19].

In this study a trajectory based alchemical free energy approach was used to predict standard free energies of binding for 22 host-guest complexes. The dataset consists of 16 guests that bind to three different molecules: two octa-acid hosts (OAH and OAMe), and a cucurbituril clip (CBC) as shown in Fig. 1. The octa-acid systems are basket shaped. OAH [20] has four flexible propionate side chains bearing two rotatable single bonds each, while, OAMe contains four methyl groups, which alter the shape and depth of the hydrophobic cavity. CBC [21, 22] is a more flexible host, which has shown a high binding affinity for ferrocene, adamantane and bicyclooctane guests [23]. The aim of this paper is to illustrate the accuracy and agreement with experiments that can be reached by means of standard free energy of binding calculations using a molecular mechanics approach with the general Amber forcefield (GAFF) [24]. In recent studies by Mirshra et al. and Aldeghi et al. [25, 26] the GAFF force field has not systematically been the most accurate forcefield, but it remains an attractive choice due to the ease of parameter generation, especially given the limited time available in SAMPL between datasets release and deadline for predictions submissions. The performance of three different variants of a double annihilation methodology for binding free energy predictions is critically assessed, as well as

overall standing with respect to other SAMPL5 submissions.

Theory and methods

Several approaches have been proposed to compute standard free energies of binding from molecular simulations.

Computing free energies of binding: models A, B, and C

One way of estimating a free energy of binding ΔG_{bind} from MD simulations is by means of a double annihilation technique proposed originally by Jorgensen et al. [27] and discussed extensively by Gilson et al. [28]. The free energy of binding ΔG_{bind} is given by:

$$\Delta G_{\text{bind}} = -k_B T \ln \frac{Z_{\text{HG,solv}} Z_{\text{solv}}}{Z_{\text{G,solv}} Z_{\text{H,solv}}}, \quad (1)$$

where k_B is the Boltzmann constant, T the temperature, $Z_{\text{HG,solv}}$, $Z_{\text{G,solv}}$, $Z_{\text{H,solv}}$ and Z_{solv} are the configuration integrals for host-guest complex, the guest, the host and the solvent molecules respectively. Figure 2 depicts how the double annihilation approach may be used to evaluate ΔG_{bind} by means of thermodynamic cycles. First the guest's partial charges are turned off both in water and in the host-guest-complex phase (discharging step), giving the discharging free energy change $\Delta G_{\text{elec}}^{\text{solv}}$ and $\Delta G_{\text{elec}}^{\text{host}}$

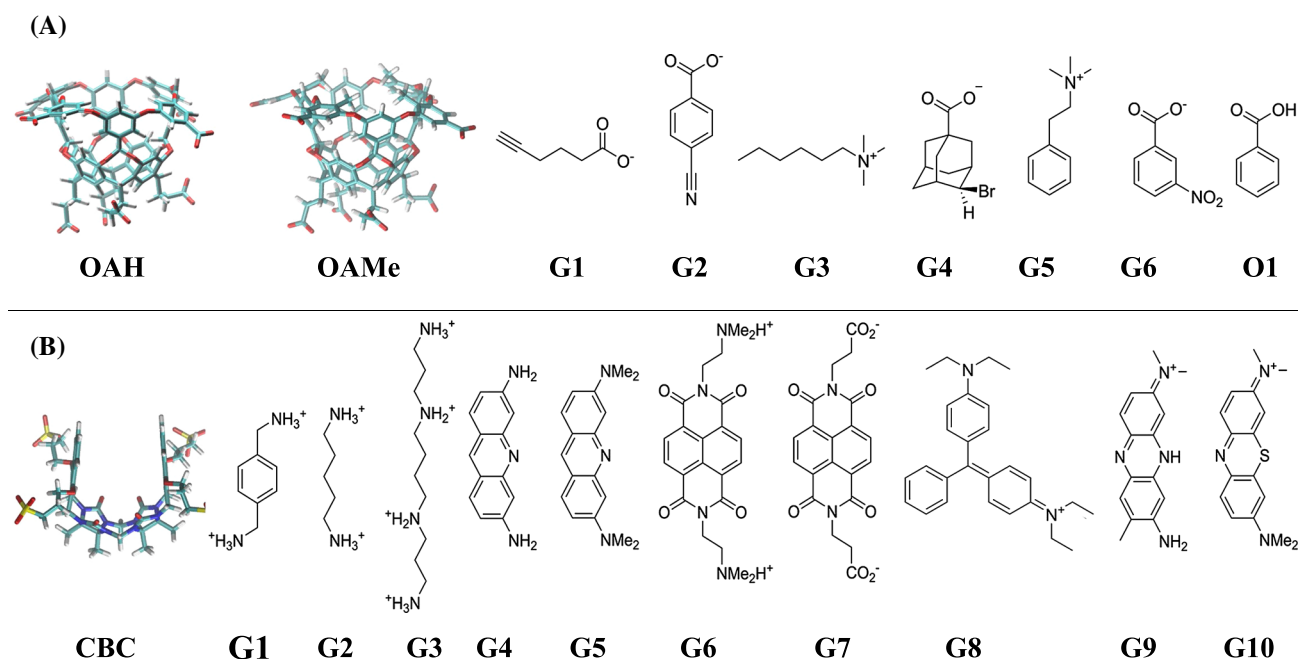
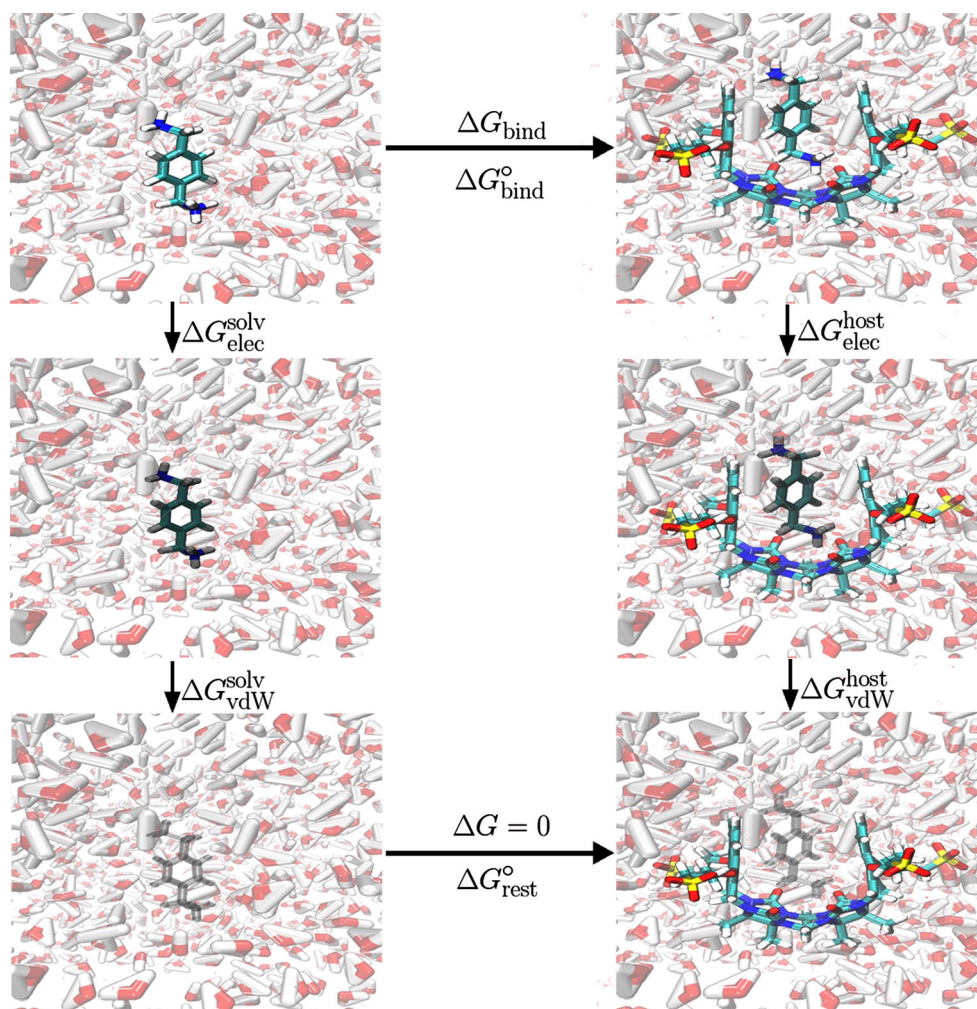


Fig. 1 a Octa-acid systems OAH and OAMe and their respective six guests, plus SAMPL4 guest O1. b Cucurbituril clip CBC and its ten guests

Fig. 2 Thermodynamic cycle for free energy of binding calculation. First, a fully interacting ligand is simulated in a water phase (*top left*), then charges and Lennard Jones terms are switched off sequentially, resulting in a fully non-interacting ligand in water (*bottom left*). The same transformation is performed in the complex, going from a fully interacting ligand (*top right*) to a dummy ligand (*bottom right*). The middle step (*middle two panels*) is concerned with the evaluation of a free energy change associated with the introduction of a host-guest restraining potential in the complex simulations. This term is neglected in *model A* and *model B*, and numerically evaluated with respect to standard state conditions in *model C*. Consequently *model A* and *model B* yield a free energy of binding ΔG_{bind} , whereas *model C* yields a standard free energy of binding $\Delta G_{\text{bind}}^{\circ}$



respectively. Secondly, the guest is fully decoupled from the solvent or host, switching off the van der Waals (vanishing step), $\Delta G_{\text{vdW}}^{\text{solv}}$ and $\Delta G_{\text{vdW}}^{\text{host}}$. The discharging and vanishing steps are usually performed with a series of intermediate simulations that depend on a coupling parameter $\lambda \in [0, 1]$. In the double annihilation method the term ΔG_{rest} shown in Fig. 2 is zero (see details below). Closure of the thermodynamic cycle in Fig. 2 shows that in the double annihilation technique the free energy of binding ΔG_{bind} is given as [29]:

$$\Delta G_{\text{bind}} = (\Delta G_{\text{elec}}^{\text{solv}} + \Delta G_{\text{vdW}}^{\text{solv}}) - (\Delta G_{\text{elec}}^{\text{host}} + \Delta G_{\text{vdW}}^{\text{host}}). \quad (2)$$

Free energies of binding computed according to Eq. (2), will be referred to as *model A*. In the actual MD simulation an empirical distance-restraint term is added to the potential energy function. This is done to prevent the non-interacting guest from drifting out of the host cavity. A flat-bottom restraining potential is used between one atom of the guest, chosen to be the one closest to the centre of

mass, and four equivalent carbon atoms of the host. The restraint potential for atom j of the guest is based on work presented in Ref. [5] and takes the following form:

$$U^{\text{restr}}(d_{j1}, \dots, d_{jN_{\text{host}}}) = \sum_{i=1}^{N_{\text{host}}} \begin{cases} 0 & \text{if } |d_{ji} - R_{ji}| \leq D_{ji} \\ \kappa_{ij} (|d_{ji} - R_{ji}| - D_{ji})^2 & \text{if } |d_{ji} - R_{ji}| > D_{ji} \end{cases}, \quad (3)$$

where $U^{\text{restr}}(d_{j1}, \dots, d_{jN_{\text{host}}})$ is the potential energy of the restraint as a function of the distance between a guest atom j and a set of host atoms i , $d_{ji} = \|\mathbf{r}_i - \mathbf{r}_j\|$ where $\|\cdot\|$ denotes a 2-norm, D_{ji} is the restraint deviation tolerance, R_{ji} the reference distance between host and guest atom, κ_{ij} the restraint force constant, and N_{host} the number of host atoms that contribute to the restraint.

Model A neglects, among other things, the contribution of long range dispersions, since a cutoff for the Lennard Jones interaction was set to 12 Å to speed up simulations (see protocols). Following work from Shirts et al. [30], it is

possible to introduce a long range dispersion correction term to the free energy of binding as a post processing step of the simulation trajectories. This leads to a corrected free energy of binding $\Delta G_{\text{bind,LJRC}}$ given by:

$$\Delta G_{\text{bind,LJLRC}} = (\Delta G_{\text{elec}}^{\text{solv}} + \Delta G_{\text{vdW}}^{\text{solv}}) - (\Delta G_{\text{elec}}^{\text{host}} + \Delta G_{\text{vdW}}^{\text{host}}) + (\Delta G_{\text{LJLRC}}^{\text{host}} - \Delta G_{\text{LJLRC}}^{\text{solv}}). \quad (4)$$

Equation (4) gives the free energy of binding for *model B*, where the Lennard Jones dispersion correction $\Delta G_{\text{LJLRC}}^{\text{X}}$ can be computed making use of the Zwanzig relation [31] in the following way:

$$\Delta G_{\text{LJLRC}}^{\text{X}} = k_B T \ln \langle \exp[-\beta(U_{\text{LJ,long}}(\mathbf{r}) - U_{\text{LJ,sim}}(\mathbf{r}))_{\text{X}}] + U_{\text{LJ,ana}} \rangle, \quad (5)$$

where X = host or solv, $U_{\text{LJ,long}}$ is the Lennard Jones energy computed with an increased long range cutoff and $U_{\text{LJ,ana}}(\mathbf{r})$ is an analytical correction from the increased range long cutoff to infinity. The long range correction $U_{\text{LJ,long}}$ is estimated in a post processing step of the ‘vanishing’ trajectories generated at $\lambda = 0$ and $\lambda = 1$, by extending the domain of the typical Lennard Jones cutoff radius in the simulation from 12 Å to cover almost the entire box instead. To define this long cutoff, the minimum box length in all directions in the input coordinates is calculated, and the new cutoff radius is set to $r_{c,\text{long}} = 0.95 \min(L_x, L_y, L_z)/2$ to allow for some fluctuations in box size. This allows an averaging over the whole trajectory of the additional contribution of the long range potential $U_{\text{LJ,long}}$, with respect to the simulated Lennard Jones term $U_{\text{LJ,sim}}$. This correction, however, does not account for an infinitely large box size giving rise to an analytical correction over an infinite domain, which is given by the additive constant given below:

$$U_{\text{LJ,ana}} = 8\pi\rho \sum_i^{N_{\text{solv}}} \sum_j^{N_{\text{solv}}} \left[\frac{\epsilon_{ij}\sigma_{ij}^{12}}{9r_c^9} - \frac{\epsilon_{ij}\sigma_{ij}^6}{3r_c^3} \right], \quad (6)$$

where ρ is the solvent density in $\text{mol } \text{Å}^{-3}$, N_{solv} is the total number of atoms in the guest, N_{solv} the number of solvent molecules, ϵ_{ij} is the Lennard Jones well depth, expressed in kcal mol^{-1} , and σ_{ij} is the Lennard Jones distance, in Å, calculated with the Lorentz–Berthelot combining rule [32]. Lennard Jones parameters for the solvent are those of the oxygen atom of the TIP3P water model [33]. It is implicitly assumed that the radial distribution function $g(r) = 1$ for distance greater than r_c . Both *model A* and *model B* lack a well defined reference state in their definition of the free energy change upon binding of the guest molecules. Therefore a third model is proposed to enable a standard state definition. For this purpose the standard state

correction is subtracted from the free energy of binding given by Eq. (4). The standard free energy of binding is given by:

$$\Delta G_{\text{bind}}^{\circ} = (\Delta G_{\text{elec}}^{\text{solv}} + \Delta G_{\text{vdW}}^{\text{solv}}) - (\Delta G_{\text{elec}}^{\text{host}} + \Delta G_{\text{vdW}}^{\text{host}}) + (\Delta G_{\text{LJLRC}}^{\text{host}} - \Delta G_{\text{LJLRC}}^{\text{solv}}) + \Delta G_{\text{restr}}^{\circ}, \quad (7)$$

where $\Delta G_{\text{restr}}^{\circ}$ accounts for the introduced flat-bottom restraint. Considering the cycle in Fig. 2, the restraint free energy change can be computed as:

$$\Delta G_{\text{restr}}^{\circ} = -k_B T \ln \left(\frac{Z_{\text{H}\circ\text{G}}^{\text{ideal,solv}}}{Z_{\text{H,solv}} Z_{\text{G,gas}}} \right), \quad (8)$$

where $Z_{\text{H}\circ\text{G}}^{\text{ideal,solv}}$ is the configuration integral for the restrained decoupled guest bound to the host, $Z_{\text{H,solv}}$ is the configuration integral for the solvated host and $Z_{\text{G,gas}}$ is the configuration integral for the guest in an ideal thermodynamic state (i.e. no non-bonded interactions). Assuming that the restraint potential is decoupled from the solvent and host degrees of freedom, Eq. (8) simplifies to:

$$\Delta G_{\text{restr}}^{\circ} = -k_B T \ln \left(\frac{Z_{\text{ooG}}^{\text{ideal,solv}}}{Z_{\text{G,gas}}} \right), \quad (9)$$

where $Z_{\text{ooG}}^{\text{ideal,solv}}$ is the configuration integral for the decoupled guest. Because the guest has no intermolecular interactions in both thermodynamic states defined in Eq. (9), and because the restraint does not hinder rotational motions, internal and rotational contributions to the configuration integrals cancel out and the only term left is the translational contribution to the configuration integral. For $Z_{\text{G,gas}}$ a standard volume of measurement V° is used, with the 1 M dilute solute convention corresponding to $V^{\circ} = 1660 \text{ Å}^3 \text{ mol}^{-1}$. Therefore Eq. (9) simplifies further to:

$$\Delta G_{\text{restr}}^{\circ} = -k_B T \ln \left(\frac{V^{\text{restr}}}{V^{\circ}} \right), \quad (10)$$

where the restraint volume V^{restr} is given by:

$$V^{\text{restr}} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_j dy_j dz_j \exp(-\beta U^{\text{restr}}(d_{j1}, \dots, d_{jN_{\text{host}}})) \quad (11)$$

V^{restr} can be calculated by numerically integrating Eq. (11). The following procedure was used. First, the coordinates of the host-guest complex in the generated trajectory at $\lambda = 1$ of the vanishing step was aligned to the first frame of the trajectory. Then, the average coordinate of each of the four host atoms used for the restraint was computed. Next, a grid spacing and an integration domain needed to be defined. The grid spacing was set to 0.1 Å and the integration domain was defined by the rectangular cuboid that

is given by the minimum/maximum coordinates of the four defined host atoms with an additional buffer around the bounding domain of ± 5 Å. Numerical integration was then performed via the multidimensional trapezoidal rule.

Host-guest simulation set-up

Host-guest input files were used as provided by the challenge organizers. For the simulations of the solvated guests, guest force field parameters and coordinates were extracted from the provided topologies and the guests were solvated in a rectangular box of TIP3P water molecules [33], with a minimum distance between the solute and the box of 12 Å, using the software *tleap*. Ions were added to neutralize the overall charge of the box. The system was energy minimized with 100 steps of the steepest decent algorithm. The following equilibration protocol was used: Solute molecules were position restrained with a force constant of $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, while the water was allowed to equilibrate in an NVT ensemble for 200 ps at 298 K, followed by an NPT equilibration for a further 200 ps and a pressure of 1 atm using the Amber module Sander [34]. Lastly, 2 ns of NPT simulation was run with Sire/OpenMM6.3 (SOMD) software (revision 2015.0.0) [35, 36], to reach a final density of about 1 g/cm^3 using a timestep of 1 fs. The final coordinate files were retrieved with CPPTRAJ [37]. The same protocol for preparation and equilibration was used for the host-guest complex.

Additionally, the reference system OAH-O1, taken from the SAMPL4 challenge [38], was set-up from scratch. Guest O1 was obtained from the modification of compound G6 using Maestro (v.10.1.012, rel 2015-1, Schrödinger) [39] and further parametrized using AM1-BCC charges [40] using Antechamber 14 [34]. Complex and water phase systems were created with *tleap*, according to the above protocol, using the same binding mode as the one provided for G6.

Alchemical free energy production simulations

For the discharging steps nine equidistant λ windows were selected for the host-guest complex and the guest in water phases. For the vanishing step 12 and 18 equidistant windows were used for octa-acid guests and CBC guests respectively. The reasoning behind these different choices was that the CBC guests were larger, therefore a denser number of λ windows was deemed necessary in order to guarantee good overlap of the potential energy distributions of neighbouring λ windows, which is essential for free energy estimation via multistate Bennett's acceptance ratio (MBAR) [41]. All simulations were run for 8 ns. A velocity-Verlet integrator was used with a time step of 4 fs using a hydrogen mass repartitioning (HMR) scheme [42].

All bonds were constrained. All simulations were performed in an NPT ensemble and temperature control was achieved with an Andersen Thermostat with a coupling constant of 10 ps^{-1} [43]. Pressure was maintained by a Monte Carlo barostat that attempted isotropic box edge scaling every 100 fs. Periodic boundary conditions were imposed with a 12 Å atom-based cutoff distance for the non-bonded interactions, using a Barker Watts reaction field with dielectric constant of 78.3 [44]. In the host-guest complex the guest molecules were restrained according to Eq. (3). The parameters were $R_{ji} = 5 \text{ \AA}$, $D_{ji} = 2 \text{ \AA}$ and $\kappa_{ji} = 10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$.

Estimation of free energy changes for models A, B, and C

Individual free energy contributions from the discharging and vanishing steps were estimated by using MBAR [41]. To estimate the accuracy and consistency of the computed binding free energy from Eq. (2), each simulation was repeated twice using different initial assignments of velocities drawn from the Maxwell–Boltzmann distribution. Final binding free energies are reported as the average of both runs and statistical uncertainties were calculated according:

$$\text{err}(\Delta G) = \frac{\sigma}{\sqrt{n}}, \quad (12)$$

where σ is the standard deviation of both runs and $n=2$ unless otherwise mentioned.

The computed binding free energies with each model are then compared to experimental values considering two different measures: the determination coefficient R^2 and mean unsigned error (MUE). To gain insight into the distribution of the two different measures a bootstrapping scheme is used in which each computed free energy point is considered to parameterize a normal distribution with its mean given by the computed free energy and σ the associated computed error. Ten thousand samples are then drawn from the artificial normal distributions for each data point and correlated with the experimental values, giving rise to a distribution of R^2 and MUE. The resulting distributions are typically not symmetric around the mean and uncertainties in the dataset metrics are reported with a 95 % confidence interval. All simulation input files and post processing scripts needed for reproducing the results, as well results files, can be found in a github repository <https://github.com/michellab/Sire-SAMPL5>.

Experimental data

Experimental data for the host-guest complexes of the octa-acids were obtained by a mixture of NMR and ITC measurements. CBC host-guest complexes standard free energy

of binding were obtained using UV, visible, and fluorescent spectroscopic measurements. All data was measured in the laboratory of Bruce Gibb (Tulane University), with detailed description of the two octacid hosts found in [20, 45]. Details on the experimental procedures and error analysis of the experimental data will be described elsewhere [46]. A summary of the experimental data used for the analysis in this work, as provided by the organizers, can be found in Table 1.

Results

Host-guest binding free energy predictions

To test the precision and accuracy of the protocols implementing the three *models A, B, and C* the free energy of binding of guest O1 to host OAH (used in SAMPL4) was retrospectively predicted. Figure 3 compares the results with experimental data [38]. Both *models A, B* yield a similar free energy of binding $\Delta G_{bind} = -6.1 \pm 0.5 \text{ kcal mol}^{-1}$. This is because the long-range corrections for Lennard

Table 1 Binding free energy, in kcal mol^{-1} , determination coefficient R^2 , mean unsigned error (MUE), in kcal mol^{-1} , *model A, B and C*

	$\Delta G_{bind} A$	$\Delta G_{bind} B$	$\Delta G_{bind} C$	Experimental
<i>OAH</i>				
G1	-8.5 ± 0.3	-8.6 ± 0.3	-6.9 ± 0.3	-5.04 ± 0.01
G2	-6.9 ± 0.2	-6.8 ± 0.1	-5.2 ± 0.1	-4.25 ± 0.01
G3	-9.0 ± 0.2	-8.9 ± 0.1	-7.3 ± 0.1	-5.06 ± 0.01
G4	$-14. \pm 0.4$	-14.4 ± 1.2	-12.8 ± 0.4	-9.37 ± 0.002
G5	-8.7 ± 0.4	-8.7 ± 0.4	-7.7 ± 0.4	-4.50 ± 0.001
G6	-7.8 ± 0.1	-7.8 ± 0.1	-6.1 ± 0.1	-5.33 ± 0.004
MUE	$3.3 < 3.7 < 4.0$	$3.3 < 3.6 < 4.0$	$1.7 < 2.1 < 2.5$	
R^2	$0.83 < 0.91 < 0.96$	$0.83 < 0.91 < 0.96$	$0.77 < 0.87 < 0.93$	
<i>OAMe</i>				
G1	-8.4 ± 0.1	-8.4 ± 0.1	-6.8 ± 0.1	-5.24 ± 0.05
G2	-9.2 ± 0.7	-9.3 ± 0.7	-7.6 ± 0.7	-5.04 ± 0.03
G3	-9.1 ± 0.1	-9.1 ± 0.1	-7.4 ± 0.1	-5.94 ± 0.12
G4	-5.0 ± 0.1	-5.0 ± 0.2	-3.4 ± 0.2	-2.38 ± 0.102
G5	-8.4 ± 0.7	-8.3 ± 0.8	-6.7 ± 0.8	-3.9 ± 0.02
G6	-7.3 ± 0.1	-7.3 ± 0.2	-5.6 ± 0.1	-4.52 ± 0.02
MUE	$3.1 < 3.4 < 3.7$	$3.1 < 3.4 < 3.7$	$1.5 < 1.7 < 2.0$	
R^2	$0.51 < 0.73 < 0.93$	$0.52 < 0.74 < 0.92$	$0.52 < 0.74 < 0.93$	
<i>CBC</i>				
G1	-7.5 ± 0.4	-7.3 ± 0.4	-6.1 ± 0.1	-5.8 ± 0.03
G2	-0.1 ± 0.1	-0.1 ± 0.1	1.2 ± 0.6	-2.5 ± 0.07
G3	0.1 ± 0.1	0.2 ± 0.1	1.5 ± 0.1	-4.02 ± 0.03
G4	-11.4 ± 0.2	-11.3 ± 0.2	-10.1 ± 0.2	-7.24 ± 0.03
G5	-15.4 ± 0.3	-15.2 ± 0.3	-14.0 ± 0.4	-8.53 ± 0.07
G6	-18.6 ± 0.2	-18.5 ± 0.2	-17.1 ± 0.2	-8.64 ± 0.05
G7	-13.6 ± 1.2	-13.5 ± 1.2	-12.3 ± 1.2	-5.17 ± 0.02
G8	-4.4 ± 1.1	-4.4 ± 1.1	-3.0 ± 1.1	-6.17 ± 0.04
G9	-16.2 ± 0.1	-16.2 ± 0.1	-15.0 ± 0.1	-7.4 ± 0.02
G10	-18.6 ± 0.2	-18.5 ± 0.2	-17.3 ± 0.2	-10.35 ± 0.02
MUE	$5.4 < 5.7 < 6.0$	$5.3 < 5.6 < 5.9$	$4.8 < 5.1 < 5.4$	
R^2	$0.70 < 0.76 < 0.81$	$0.70 < 0.76 < 0.82$	$0.69 < 0.76 < 0.82$	

Uncertainties in the calculated individual binding free energies are the standard error of the mean. Dataset metrics are given with a 95 % confidence interval

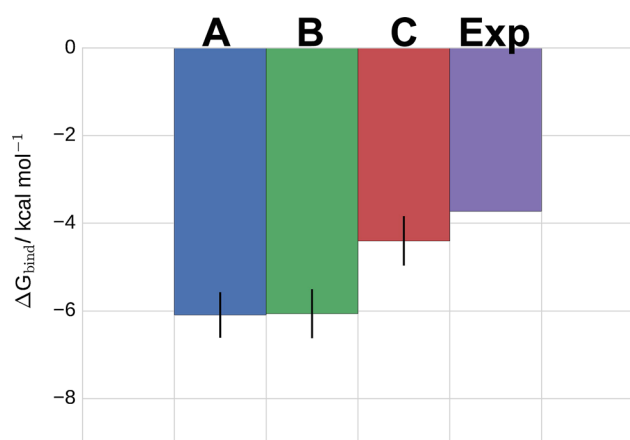
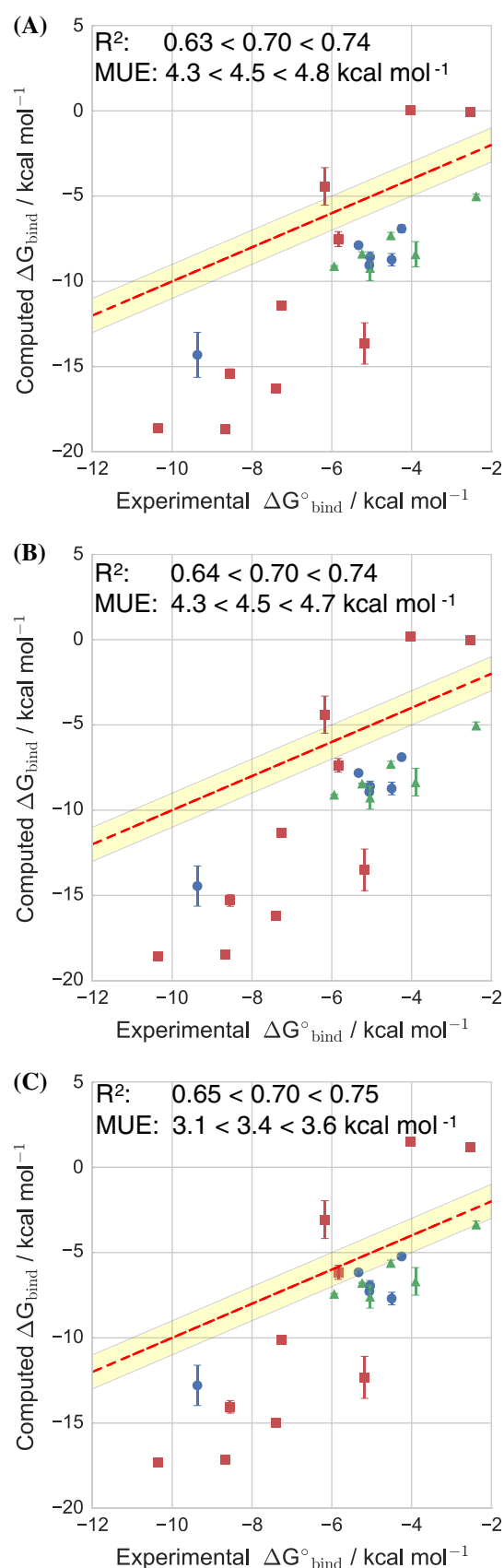


Fig. 3 Comparison of the computed binding free energy ΔG_{bind} according to *models A, B* and the computed standard binding free energy $\Delta G_{\text{bind}}^{\circ}$ according to *model C* with respect to experimental data for the SAMPL4 OAH-O1 complex

Jones interactions implemented in *model B* produce a negligible correction term of $0.03 \text{ kcal mol}^{-1}$. By contrast, the addition of a standard state correction in *model C* leads to a standard free energy of binding of $\Delta G_{\text{bind}}^{\circ} = -4.4 \pm 0.5 \text{ kcal mol}^{-1}$ which is in good agreement with the experimental data of $\Delta G_{\text{bind}}^{\circ} = -3.7 \pm 0.1 \text{ kcal mol}^{-1}$ [38].

Next, blinded predictions were performed for each SAMPL5 host-guest. Figure 4 contrasts the predictive power of the different models against the experimental data that was released after submission of the predictions. Figure 4a, shows the results for *model A*, b for *model B* and c for *model C* respectively. Results for each host-guest system are also reported in Table 1. Taking the full dataset into account, all three models yield a similar R^2 value of ca. $0.65 < 0.70 < 0.75$. *Models A, B* have a similar MUE of ca. $4.3 < 4.5 < 4.7 \text{ kcal mol}^{-1}$, whereas *model C* is statistically more accurate, with a MUE of ca. $3.2 < 3.4 < 3.6 \text{ kcal mol}^{-1}$. The accuracy of the predictions for the three different hosts was also considered individually and summarised in Table 1. As judged by the MUE measures, the models perform better across the octa-acid systems than for CBC. In particular, *model C* gives the best predictions compared to *A* and *B* for octa-acid systems, with a MUE of $1.7 < 2.1 < 2.4 \text{ kcal mol}^{-1}$ and $1.4 < 1.7 < 2.0 \text{ kcal mol}^{-1}$ for OAH and OAMe

Fig. 4 a Comparison of the binding free energy ΔG_{bind} to experiments according to *model A* and *model B* in **b** and standard binding free energy of $\Delta G_{\text{bind}}^{\circ}$ according to *model C* with respect to experimental results in **c** for all host-guest systems. The red line indicates ideal correlation between experiments and computed results and the yellow shaded region gives a binding free energy error bound of $1 \text{ kcal} \cdot \text{mol}^{-1}$. OAH systems are colored in blue, OAMe in green and CBC in red. Error bars denote $\pm \text{Eq. of (12)}$



respectively. R^2 is on average slightly higher for both octa-acid hosts (R^2 0.77 < 0.87 < 0.93, *model C* for OAH and R^2 0.52 < 0.74 < 0.93, *model C* for OAMe) than CBC (R^2 0.70 < 0.76 < 0.82) for models A, B and C, but the trend is not strong given statistical uncertainties.

Next, attention was focussed on the guests for which predictions showed the largest discrepancy with respect to experimental data. For instance, the standard free energy of binding of guest G5 in complex with both OAH and OAMe is overestimated by -3.6 and -2.8 kcal mol $^{-1}$ respectively. Guest G4 is also significantly stabilised in complex with OAH. G4 is arguably the most hydrophobic guest in the series studied and evidence from our accompanying distribution coefficient article suggest that the GAFF force field appear to favor the transfer of hydrophobic solutes into hydrophobic environments [47]. Since G5 is the largest outlier in both octa-acid hosts, the validity of its simulated binding mode was evaluated. For this purpose, G5 in the host-guest complexes was rotated by approximately 180° about its centre of mass such that the amine group pointed towards the bottom of the host cavities, and calculations were repeated using these new coordinates after solvent equilibration. Binding free energy predictions from *model C* obtained for this alternative binding mode were poor ($\Delta G_{bind}^\circ = +1.8 \pm 0.1$ and $+16.7 \pm 0.1$ kcal mol $^{-1}$ for OAH and OAMe respectively), suggesting that the original binding mode is more likely.

For the CBC host, the best MUE is about $4.8 < 5.1 < 5.4$ kcal mol $^{-1}$ for *model C*, with no significant difference over *model A* and *B*. This is surprising since the determination coefficient R^2 0.69 < 0.76 < 0.82 is quite reasonable. In particular, *model C* performs better than *A* and *B*, but large errors are present for a series of guests. Guests G2 and G3 are predicted to bind substantially worse than observed in experiments. The main difference with other guests in this dataset is that these two molecules are made up of linear flexible alkyl chains, and contain several (presumably) positively charged ammonium groups. By contrast, G4–G7, G9 and G10 are predicted to bind significantly better than experimentally observed. These compounds present a variety of net charges, but are all made up of conjugated aromatic rings. Additionally, empirical pKA estimations [48] suggest that G5, G6, G9 and G10 could adopt multiple protonation states at the pH where binding constants were measured. Hence, it is unclear whether the discrepancies are due to forcefield errors or finite-size effects.

As a separate issue, the reproducibility of standard free energies of binding was evaluated by comparing the results from *model C* with those reported by the Gilson lab (UCSD) for the octa-acid hosts [49]. The same input files were used, but the free energy calculations were performed

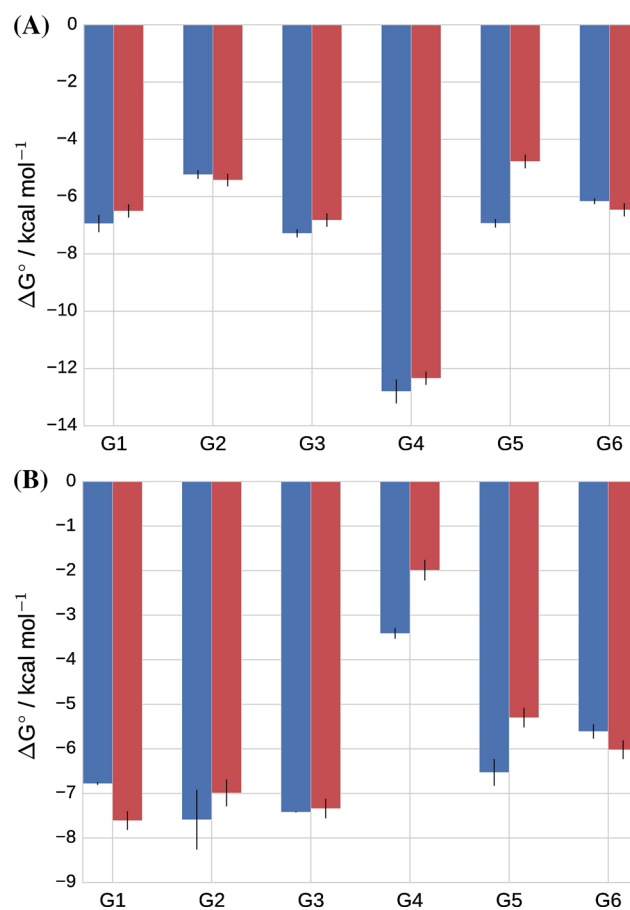


Fig. 5 Comparison between standard binding free energies computed with *model C* (blue) and with the attach-pull-release method (red) for OAH in (a) and OAMe in (b)

with the *pmemd.cuda* program from AMBER 14 [50], and a different potential of mean force based 'attach-pull-release' (APR) methodology [51]. Figure 5 shows that a good agreement is observed between both OAH (Fig. 5a) and OAMe (Fig. 5b) hosts, with a mean unsigned differences of about 0.4 kcal mol $^{-1}$ in the former case and 0.6 kcal mol $^{-1}$ for the latter. At first glance this level of variability seems reasonable given the typical statistical uncertainties of each methodology. Nonetheless closer inspection indicates that OAH-G5, OAMe-G5 and OAMe-G4 show significant discrepancies. Since the *model C* standard free energies of binding were only estimated from two repeats a concern was that the error estimates were not reliable. To test this two additional repeats were performed for these systems. The standard free energies of binding obtained from four repeats of *model C* are: ΔG_{bind}° (SOMD, OAH-G5) = -6.9 ± 0.1 , ΔG_{bind}° (SOMD, OAMe-G4) = -3.4 ± 0.2 , ΔG_{bind}° (SOMD, OAMe-G5) = -6.5 ± 0.3 kcal mol $^{-1}$ respectively. The results were statistically identical to those obtained from two repeats (Table 1) for OAMe-G4 and OAMe-G5, but not OAH-G5. Personal

discussions with the Gilson lab prompted additional APR calculations which produced revised values for ΔG_{bind}° (APR, OAMe-G4) = -4.3 ± 0.3 kcal mol⁻¹, and ΔG_{bind}° (APR, OAH-G5) = -4.5 ± 0.5 kcal mol⁻¹. However, discrepancies remain and further work is needed to establish the protocol variations that introduced this variability in the computed standard free energies of binding.

Conclusions

The present alchemical free energy calculation protocols proved reasonably reproducible (Table 1). This was unexpected given the size of the guests that was deemed large for an absolute binding free energy calculation, suggesting that longer per- λ simulation time than what was used here would be necessary. Factors that may have contributed to this outcome include the relative rigidity of some of the guests, the rapid relaxation of the hosts upon guest decoupling, the symmetry of the hosts, and the use of distance restraints to limit translational motions of the decoupled guests. Encouragingly, the results were also reasonably predictive, at least when judged by correlation with experiment (R^2 .65 < 0.70 < 0.75 for *model C*). Indeed, the SAMPL5 submissions for *models A, B, C* were among the top-performing protocols of this entire competition with respect to this metric. Nevertheless, systematic errors are present and the same models do not fare as well when ranked according to a mean unsigned error metric. *Model B* yields results that are identical to *model A* since the long range correction for missing dispersion interactions is essentially negligible. This was unexpected given previous reports were this term was found to be a significant contribution to standard free energies of binding [30]. For the systems considered here it seems that the simulation cutoffs used were sufficient to include most of the guest-host dispersion interactions. Satisfactorily, addition of a standard state correction term in *model C* systematically improves agreement with experimental data. In addition the computed standard free energies of binding for *model C* agree well with those produced independently by members of the Gilson lab (UCSD) using a different code and methodology. However, it is not currently understood why a few compounds show more significant deviations between the double decoupling and APR methodologies and this should receive further attention. It is well known that the computation of free energies of solvation of charged solutes via molecular simulations is typically affected by significant finite-size effects [52–55]. Given the broad range of net charges in the guests considered here, it is perhaps surprising that encouraging R^2 values were obtained. For the host-guest binding energies reported here errors due to finite-size electrostatics is

mitigated since partial error compensation occur between the simulations of the solvated guest and the host-guest complex. However, it seems reasonable to anticipate that the significant MUE values could be decreased with the use of suitable schemes to reduce or eliminate finite-size errors.¹ Other areas where further improvement could be sought for this dataset include the explicit consideration of multiple tautomeric forms of the guests, as well as a more systematic evaluation of alternative potential energy functions.

Acknowledgments J. M. is supported by a Royal Society University Research Fellowship. The research leading to these results has received funding from the European Research Council under the European Unions Seventh Framework Programme (FP7/ 2007-2013)/ ERC Grant Agreement No. 336289.

References

1. Michel J (2014) *Phys Chem Chem Phys* 16(10):4465–4477
2. Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ (2010) *J Comput Aided Mol Des* 24(4):259–279
3. Skillman AG (2012) *J Comput Aided Mol Des* 26(5):473–474
4. Guthrie JP (2009) *J Phys Chem B* 113(14):4501–4507
5. Michel J, Henchman RH, Gerogiokas G, Southey MWY, Mazanetz MP, Law RJ (2014) *J Chem Theory Comput* 10(9):4055–4068
6. Jorgensen LW, Thomas LL (2008) *J Chem Theory Comput* 4(6):869–876
7. Woods CJ, Malaisree M, Hannongbua S, Mulholland AJ (2011) *J Chem Phys*. doi:10.1063/1.3519057
8. Chang C-E, Gilson MK (2004) *J Am Chem Soc* 126(40):13156–13164
9. Muddana SH, Gilson MK (2012) *J Chem Theory Comput* 8(6):2023–2033
10. Mikulskis P, Cioloboc D, Andrejić M, Khare S, Brorsson J, Genheden S, Mata RA, Söderhjelm P, Ryde U (2014) *J Comput Aided Mol Des* 28(4):375–400
11. König G, Pickard IV FC, Mei Y, Brooks BR (2014) *J Comput Aided Mol Des* 28(3):245–257
12. Beckstein O, Fourrier A, Iorga BI (2014) *J Comput Aided Mol Des* 28(3):265–276
13. Monroe JI, Shirts MR (2014) *J Comput Aided Mol Des* 28(4):401–415
14. Chen I-J, Foloppe N (2011) *Drug Develop Res* 72(1):85–94
15. Halgren TA, Damm W (2001) *Curr Opin Struct Biol* 11(2):236–242
16. Kastenzholz MA, Hnenberger PH (2004) *J Phys Chem B* 108(2):774–788
17. Jorgensen LW, Ravimohan C (1985) *J Chem Phys* 83(6):3050–3054
18. Mezei M (1987) *J Chem Phys* 86(12):7084–7088
19. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) *Acc Chem Res* 33(12):889–897

¹ A fourth model that corrects for finite-effects on electrostatic interactions was also submitted by our group, but the results are not discussed here as it was subsequently established that a coding error led to incorrect evaluation of the correction terms.

20. Gibb CLD, Gibb CB (2013) *J Comput Aided Mol Des* 28(4):319–325
21. Gilberg L, Zhang B, Zavalij PY, Sindelar V, Isaacs L (2015) *Org Biomol Chem* 13:4041–4050
22. Zhang B, Isaacs L (2014) *J Med Chem* 57(22):9554–9563
23. Sure R, Antony J, Grimme S (2014) *J Phys Chem B* 118(12):3431–3440
24. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) *J Comput Chem* 25(9):1157–1174
25. Mishra SK, Calabr G, Loeffler HH, Michel J, Koa J (2015) *J Chem Theory Comput* 11(7):3333–3345
26. Aldeghi M, Heifetz A, Bodkin MJ, Knapp S, Biggin PC (2016) *Chem Sci* 7(1):207–218
27. Jorgensen WL, Buckner JK, Boudon S, Tirado-Rives J (1988) *J Chem Phys* 89(6):3742–3746
28. Gilson MK, Given JA, Bush BL, McCammon JA (1997) *Biophys J* 72(3):1047
29. Michel J, Essex JW (2010) *J Comput Aided Mol Des* 24(8):639–658
30. Shirts MR, Mobley DL, Chodera JD, Pande VS (2007) *J Phys Chem B* 111(45):13052–13063
31. Zwanzig WR (1954) *J Chem Phys* 22(8):1420–1426
32. Frenkel D, Smit B (2001) *Understanding molecular simulation*, 2nd edn. Academic Press Inc, Orlando
33. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) *J Chem Phys* 79(2):926–935
34. Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, Cheatham III DS, Darden TA, Duke TA, Gohlke H, Goetz AW, Gusarov S, Homeyer N, Janowski P, Kaus J, Kolossvary I, Kovalenko A, Lee TS, LeGrand S, Luchko T, Luo R, Madej B, Merz KM, Paesani F, Roe DR, Roitberg A, Sagui C, Salomon-Ferrer R, Seabra G, Simmerling CL, Smith W, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Kollman PA (2014) AMBER 14, University of California, San Francisco
35. Woods C, Mey ASJ, Calabro G, Michel J (2016) Sire molecular simulations framework. <http://siremol.org>. Accessed May 31
36. Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, Ku JP, Beauchamp KA, Lane TJ, Wang L-P, Shukla D, Tye T, Houston M, Stich T, Klein C, Shirts MR, Pande VS (2013) *J Chem Theory Comput* 9(1):461–469
37. Roe RD, Cheatham TE III (2013) *J Chem Theory Comput* 9(7):3084–3095
38. Muddana HS, Fenley AT, Mobley DL, Gilson MK (2014) *J Comput Aided Mol Des* 28(4):305–317
39. Schrödinger release 2015-2: Maestro, version 10.2, Schrödinger, llc, New York, NY, 2015
40. Jakalian A, Bush BL, Jack DB, Bayly CI (2000) *J Comput Chem* 21(2):132–146 (cited By 552)
41. Shirts MR, Chodera JD (2008) *J Chem Phys*. doi:10.1063/1.2978177
42. Hopkins CW, Le Grand S, Walker RC, Roitberg AE (2015) *J Chem Theory Comput* 11(4):1864–1874
43. Andersen HC (1980) *J Chem Phys* 72:2384–2393
44. Tironi IG, Sperb R, Smith PE, van Gunsteren WF (1995) *J Chem Phys* 102(13):5451–5459
45. Gan H, Benjamin CJ, Gibb BC (2011) *J Am Chem Soc* 133(13):4770–4773
46. Yin J, Henriksen NM, Slochow DR, Chiu MW, Mobley DL, Gilson MK (2016) Overview of the SAMPL5 host-guest challenge: are we doing better? *J Comput Aided Mol Des* (under review)
47. Bosisio S, Mey ASJS, Michel J (2016) Blinded predictions of distribution coefficients in the SAMPL5 challenge. *J Comput Aided Mol Des* (under review)
48. Chemaxon. www.chemicalize.org
49. Yin J, Henriksen NM, Slochow DR, Gilson MK (2016) The SAMPL5 host-guest challenge: binding free energies and enthalpies from explicit solvent simulations. *J Comput Aided Mol Des* (under review)
50. Goetz AW, Poole D, Le Grand S, Walker RC, Salomon-Ferrer R (2013) *J Chem Theory Comput* 9:3878–3888
51. Velez-Vega C, Gilso MK (2013) *J Comput Chem* 34(27):2360–2371
52. Kastenzholz MA, Hünenberger PH (2006) *J Chem Phys* 124(12):124106
53. Kastenzholz MA, Hünenberger PH (2006) *J Chem Phys* 124(22):224501
54. Reif MM, Oostenbrink C (2014) *J Comput Chem* 35(3):227–243
55. Rocklin GJ, Boyce SE, Fischer M, Fish I, Mobley DL, Shoichet BK, Dill KA (2013) *J Mol Biol* 425(22):4569–4583