



Blinded predictions of standard binding free energies: lessons learned from the SAMPL6 challenge

Michail Papadourakis¹ · Stefano Bosisio¹ · Julien Michel¹

Received: 2 June 2018 / Accepted: 24 August 2018 / Published online: 29 August 2018
© The Author(s) 2018

Abstract

In the context of the SAMPL6 challenges, series of blinded predictions of standard binding free energies were made with the SOMD software for a dataset of 27 host–guest systems featuring two octa-acids hosts (*OA* and *TEMOA*) and a cucurbituril ring (*CB8*) host. Three different models were used, *ModelA* computes the free energy of binding based on a double annihilation technique; *ModelB* additionally takes into account long-range dispersion and standard state corrections; *ModelC* additionally introduces an empirical correction term derived from a regression analysis of SAMPL5 predictions previously made with SOMD. The performance of each model was evaluated with two different setups; *buffer* explicitly matches the ionic strength from the binding assays, whereas *no-buffer* merely neutralizes the host–guest net charge with counter-ions. *ModelC/no-buffer* shows the lowest mean-unsigned error for the overall dataset (MUE $1.29 < 1.39 < 1.50$ kcal mol⁻¹, 95% CI), while explicit modelling of the buffer improves significantly results for the *CB8* host only. Correlation with experimental data ranges from excellent for the host *TEMOA* (R^2 $0.91 < 0.94 < 0.96$), to poor for *CB8* (R^2 $0.04 < 0.12 < 0.23$). Further investigations indicate a pronounced dependence of the binding free energies on the modelled ionic strength, and variable reproducibility of the binding free energies between different simulation packages.

Keywords SAMPL6 · SAMPLing · Binding free energy · Alchemical free energy

Introduction

Computer-aided drug design (CADD) is a powerful methodology for early stage drug-discovery [1]. In particular there is much interest in the use of molecular simulations methods to support drug-discovery efforts [2], via for instance investigation of protein folding mechanisms [3, 4], or ligand modulation of millisecond time-scale conformational changes in proteins [5]. Another application of molecular simulations in CADD is potency predictions to decrease time and costs of hit-to-lead and lead optimization stages needed before molecules may be progressed towards clinical studies [6]. This requires accurate description of ligand–protein energetics,

which is nowadays increasingly sought via use of free energy calculations methods.

Among various existing free energy calculation methodologies, alchemical free energy calculations (AFE) have attracted much interest in recent years [7–9], due to their robust grounding in statistical physics. AFE calculations capture non-additivity of structure–activity relationship in congeneric series that are overlooked by empirical scoring methods [10], and have given useful potency estimates for a range of protein–ligand systems [11–13]. AFE methods may also be used to predict physical properties, such as lipophilicity coefficients [14–16]. In spite of encouraging successes, there are still important technical hurdles to overcome. Usual concerns involve finite-sampling effects that introduce statistical errors [17–20], whereas inaccuracies in potential energy functions contribute to systematic errors [21]. Additionally, algorithmic decisions for the handling of long range electrostatic interactions and finite-size artefacts affect simulation results in ways that are still poorly understood, with effects particularly apparent in the modelling of charged species [22–24]. Thus, it is important to improve the

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10822-018-0154-6>) contains supplementary material, which is available to authorized users.

✉ Julien Michel
mail@julienmichel.net

¹ EaStCHEM School of Chemistry, Joseph Black Building, King's Buildings, Edinburgh EH9 3FJ, UK

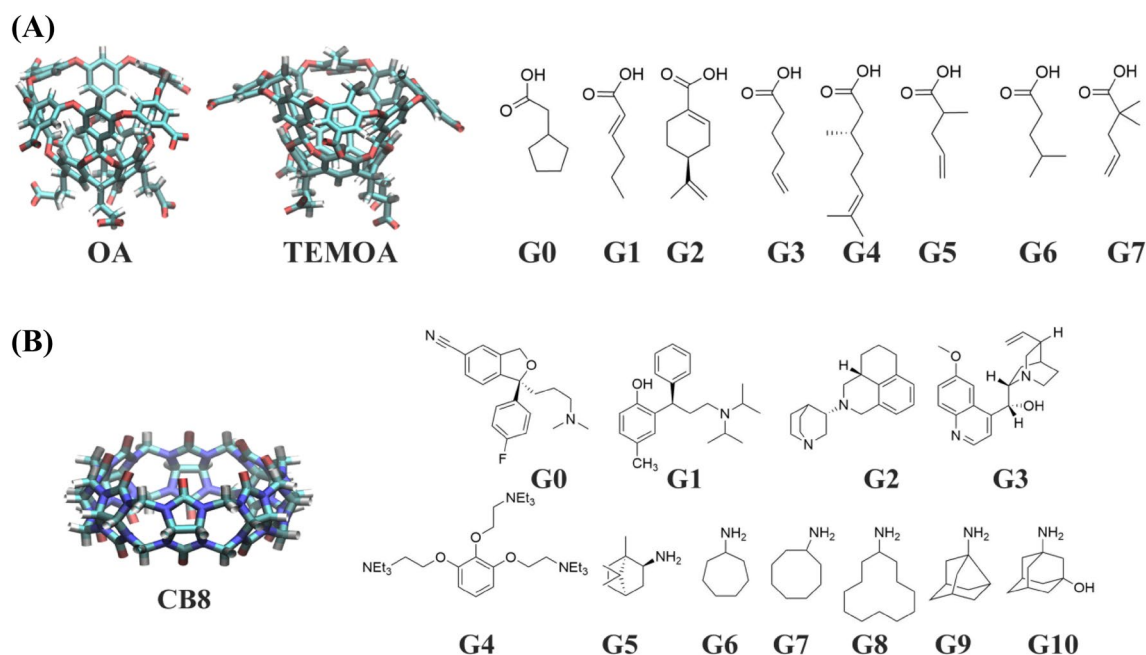


Fig. 1 Depiction of the SAMPL6 host–guest dataset. **a** OA and TEMOA host–guest systems. **b** CB8 host–guest systems

robustness of AFE protocols to enable their reliable application to structure-based drug design problems.

Blinded prediction competitions offer a valuable resource to reduce bias in validation studies and to test practical utility of a methodology in a setting that more closely resembles CADD in practice [25]. The D3R Grand Challenges have proven a popular blinded competition, with a focus on validating computational methods for modelling of protein–ligand interactions [25, 26]. The Statistical Assessment of Modelling of Proteins and Ligands (SAMPL) is also a well-established blinded competition for free energy science in drug discovery [27]. The SAMPL challenge was founded in 2007 and usually requests participants to predict physical chemical properties, such as binding affinities for host–guest systems, or hydration free energies of small drug-like molecules [28, 29]. Host–guest systems are attractive since they provide more tractable milestones towards validation of protocols for modelling protein–ligand binding energetics [30].

The 6th SAMPL (SAMPL6) competition was launched in September 2017. Our group focused on the host–guest leg of this contest, which requested predictions of standard free energies of binding for 27 guests across three different hosts [31]. The host molecules consisted in two octa-acids, OA and TEMOA molecules [32–35], and a cucurbituril ring clip CB8 [36–39], as shown in Fig. 1. The octa-acid systems (Fig. 1a) are basket shaped; OA contains four flexible propionate side chains bearing two rotatable single bonds each, while TEMOA contains four methyl groups, which alter the shape of the hydrophobic cavity. CB8 (Fig. 1b) is a heteroaromatic multicyclic molecule, chemically related to the cucurbiturils,

made of methylene bridges containing eight glycoluril units. CB8 is considered a more flexible host than OA and TEMOA, though the latter two also contain flexible groups at the top and bottom of their cavities [38, 39]. Additionally, SAMPL6 introduced a SAMPLing challenge focused on evaluating convergence and reproducibility (across codes) of free energy predictions. To this end, input files for parameterized host–guests OA–G3, OA–G6 and CB8–G3 were provided and participants requested to evaluate convergence of their binding free energy estimates.

This report summarizes the performance of our free energy code Sire/OpenMM Molecular Dynamics (SOMD) against the SAMPL6 host–guest dataset, as well as the lessons learned for continuing efforts to improve the robustness of AFE methods in CADD.

Theory and methods

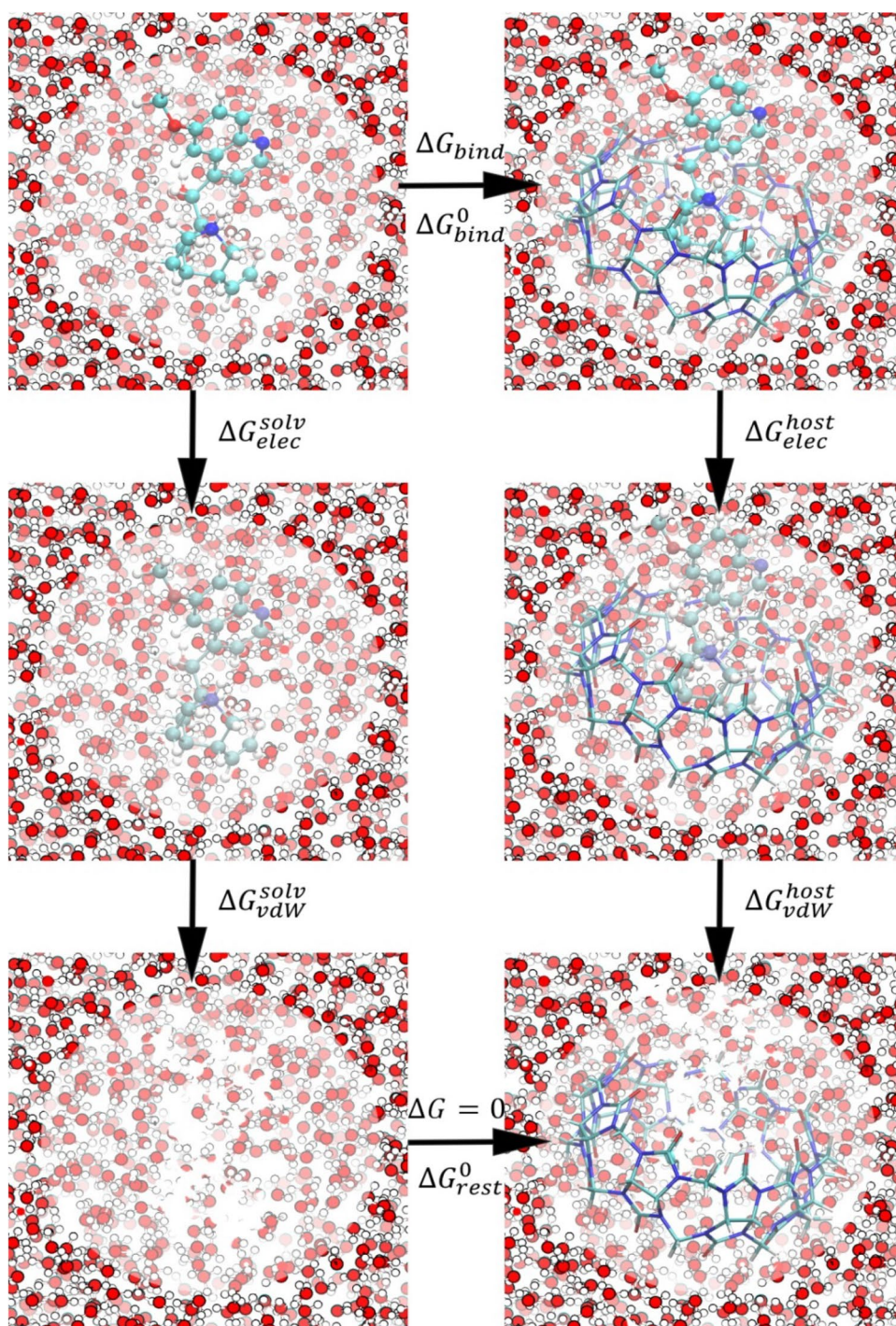
Definition of binding affinity

The reversible binding of a ligand L to a receptor P can be written as:



where ΔG_{bind}° is the standard free energy of binding of ligand L to receptor P . A statistical thermodynamics treatment leads to Eq. 2 [40]:

Fig. 2 Thermodynamic cycle for standard binding free energy calculations. Firstly, the fully interacting guest is simulated in a free phase (top left) and a bound phase (top right), then the charges and the van der Waals terms are switched off, resulting in a non-interacting guest in water (bottom left), and bound to the host (bottom right)



$$\Delta G_{bind}^0 = -k_B T \ln \frac{Z_{PL,solv} Z_{solv} V}{Z_{L,solv} Z_{P,solv} V_0} \quad (2)$$

where $Z_{PL,solv}$, Z_{solv} , $Z_{L,solv}$ and $Z_{P,solv}$ are the configuration integrals for complex system, the solvent molecules, the ligand and the protein system respectively, V is the volume of binding, namely the volume available to the ligand to

bind the protein, and V_0 the standard state volume, which is usually equal to $1661 \text{ \AA}^3/\text{molecule}$.

Computing free energies of binding through models A, B, C

Equation 2 can be applied to estimate the binding free energy for host–guest systems. Computationally, the free

energy is evaluated by using molecular dynamics simulations (MD) by means of a double annihilation technique [13, 41, 42]. Figure 2 shows how this approach is used to evaluate ΔG_{bind}^0 by means of a thermodynamic cycle. In the first step (*discharging step*) the charges of the guest's atoms are turned off both in the solvated phase and in the bound phase, providing the discharging free energy changes ΔG_{elec}^{solv} and ΔG_{elec}^{host} respectively. In the second step (*vanishing step*) a “non-interacting” guest is obtained by now switching off the van der Waals parameters of the discharged guest both in solvent and complex phase, giving the vanishing free energy changes, ΔG_{vdW}^{solv} and ΔG_{vdW}^{host} , respectively. To prevent the ligand from drifting away from the host cavity a series of a flat-bottom distance restraints are defined between one guest atom j closest to the center of mass of the guest and four host atoms i . The restraint potential is given by Eq. 3 [13]:

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

where $U_{(d_{j1}, \dots, d_{jN_{host}})}^{restr}$ is the potential energy of the restraint as a function between a guest atom j and a set of host atoms i , $| \cdot |$ denotes the absolute value, D_{ji} is the restraint deviation tolerance, R_{ji} is the reference distance between host and guest atom, κ_{ji} is the restraint force constant and N_{host} is the number of host atoms that contribute to the restraint.

From the closure of the thermodynamic cycle (Fig. 2) the binding free energy ΔG_{bind} is given by Eq. 4:

$$\Delta G_{bind}^{ModelA} = (\Delta G_{elec}^{solv} + \Delta G_{vdW}^{solv}) - (\Delta G_{elec}^{host} + \Delta G_{vdW}^{host}) \quad (4)$$

Free energies of binding computed with Eq. 4 will be referred to as *ModelA* binding energies.

ModelA does not take into account the contribution of long range dispersions interactions due to the use of non-bonded cutoffs. Thus, to improve over *ModelA*, a long-range dispersion correction term is added to the free energy of binding by post-processing of the end states trajectories [43]. Additionally, a free energy correction term is introduced to relate the volume available to the restrained but non-interacting ligand to standard state conditions. This leads to Eq. 5 for predictions of binding free energies via *ModelB*.

$$\Delta G_{bind}^{0,ModelB} = \Delta G_{bind}^{ModelA} + (\Delta G_{LJLRC}^{host} - \Delta G_{LJLRC}^{solv}) + \Delta G_{restr}^0 \quad (5)$$

ΔG_{LJLRC}^{host} is the long range correction term for the bound phase, and ΔG_{LJLRC}^{solv} is the LRC term for the solvated phase. Details for the evaluation of these terms have been provided elsewhere [13]. ΔG_{restr}^0 is the free energy cost for imposing the host–guest restraint which is given by Eq. 6:

$$\Delta G_{restr}^0 = -k_B T \ln \left(\frac{Z_{H \cdot G_{ideal}}}{Z_{H,solv} Z_{G,gas}} \right) \quad (6)$$

where $Z_{H \cdot G_{ideal}}$ is the configuration integral for the restrained decoupled guest bound to the host, $Z_{H,solv}$ is the configuration integral for the solvated host and $Z_{G,gas}$ is the configuration integral for the guest in an ideal thermodynamic state. Equation 6 is evaluated by numerical integration as described elsewhere [13].

Finally, *ModelC* was constructed by devising an empirical correction term to account for systematic errors due to finite size artefacts and inaccuracies in potential energy functions. Linear regression models were obtained by correlating past SAMPL5 binding free energies computed with SOMD to experimental data, leading to Eq. 7 to compute *ModelC* binding free energies:

$$\Delta G_{bind}^{0,ModelC} = \frac{\Delta G_{bind}^{0,ModelB} - \beta}{\alpha} \quad (7)$$

where α and β are the slope and intercept of the linear regression model. SAMPL5 featured the same hosts OA and TEMOA but a different host CB7. Thus, separate regression models were determined for use with OA, TEMOA or CB8 hosts, the parameters are given in Table S1.

Preparation of host–guest input files for free energy calculations

The SAMPL6 organizers provided mol2 files for hosts, OA, TEMOA and CB8, and ligands, depicted in Fig. 1. Each file had the same Cartesian frame of reference and docking was performed with OpenEye toolkit [44–46] to predict the most likely binding mode. Experimental measurements were done at a pH 11.7 ± 0.1 at 298 K in presence of a buffer of 10 mM Na_3PO_4 for OA and TEMOA. CB8 was measured at pH 7.4 ± 0.1 at 298 K with 25 mM Na_3PO_4 buffer. To understand the influence of the buffer on binding free energy predictions, two different sets of input files were prepared, leading to *no-buffer* and *buffer* setups.

Input files for the no-buffer setup

In the *no-buffer* simulations, the presence of the additional Na_3PO_4 buffer was neglected. OA, TEMOA and CB8 host–guest systems were parametrized starting from the mol2 host and guest's files. The force field parameters for OA and TEMOA hosts were taken from a preceding study of host–guest binding energies carried out for the SAMPL5 contest [13]. To create the host–guest complex input files, the utilities *parmed* and *tleap* were used [47, 48]. The combined host–guest complex mol2 file was loaded in *tleap*

along with host force field parameters and GAFF1.8 and AM1/BCC parameters for the ligand as generated by *antechamber* from the AMBER16 release [49, 50]. The system was solvated in a cubic box with TIP3P water molecules [51], with a minimum distance between the solute and the box of 12 Å. Counter ions were added to neutralize the total net charge. The same approach was followed for parameterizing the ligand in a solvated phase.

Next an equilibration protocol was applied to relax the box size. Initially, energy minimization of the entire system was performed with 100 steps of steepest descent gradients, using *sander*. Then, solute molecules were position restrained with a force constant of 10 kcal mol⁻¹ Å⁻² while water molecules were allowed to equilibrate in an NVT ensemble, 200 ps at 298 K, followed by a NPT equilibration for further 200 ps at 1 atm pressure. Finally, a 2 ns NPT MD simulation was run with the *SOMD* software (revision 2017.1.0) to reach a final density of about 1 g cm⁻³ [52, 53]. The final coordinate files were retrieved with *cpptraj*. The edge length of the host–guest boxes was about 50 Å, whereas the solvated guest phase had an edge length of about 35 Å.

Input files for the buffer setup

For the second set of simulations, additional counter ions were added to mimic the presence of a buffer in the experiments. However, Na₃PO₄ was modelled by NaCl as force field parameters for multivalent ions were not readily available. Thus, for *OA* and *TEMOA* systems, the 10 mM sodium phosphate buffer was modelled with 60 mM of NaCl to match the ionic strength of the solution used for the experiments. Starting from the complex phase files, created as described previously, 4 additional Na⁺ and 4 Cl⁻ ions were added to each system, using *tleap*. The equilibration protocol described previously was reapplied to adjust the placement of the counter ions. For the preparation of the solvated phase, the host molecule was extracted from an equilibrated host–guest box and the host's heavy atoms were replaced with water molecules. After equilibration the final solvated phase system had the same amount of Na⁺ and Cl⁻ ions as in the host–guest complex system, and a similar box size dimension. The same procedure was followed for *CB8*. In this case, 25 mM Na₃PO₄ were matched with 150 mM NaCl, thus 8 Na⁺ and 8 Cl⁻ ions were added to each *CB8* host–guest system.

SAMPL6 simulation protocols

For the octa-acid hosts, both complex and solvated phase discharging step were run with nine equidistant λ windows. Twelve λ windows (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 1.0) were employed for the vanishing step, both

in bound and solvated phase. For the *CB8* host the bound and solvated phase discharging steps have been run with nine equidistant λ windows. The solvated vanishing step was carried out with the same windows setup as for the octa-acid guests. The bound vanishing step was carried out with 16 λ windows (λ 0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.70, 0.85, 1.00) as preliminary runs indicated a need for greater number of windows to obtain reliable free energy changes.

All the simulations were run for duration of 8 ns with SOMD in an NPT ensemble. Temperature control was achieved with an Andersen Thermostat with a coupling constant of 10 ps⁻¹ [54]. Pressure control was maintained by a Monte Carlo barostat that attempted isotropic box edge scaling every 100 fs. A 12 Å atom-based cutoff distance for the non-bonded interactions was used, using a Barker Watts reaction field with dielectric constant of 78.3 [55]. In the bound phase the restraints parameters of Eq. 3 were: $R_{ji} = 5$ Å, $D_{ji} = 2$ Å and $\kappa_{ij} = 10$ kcal mol⁻¹ Å⁻² for all the octa-acid systems, while $R_{ji} = 7$ Å, $D_{ji} = 2$ Å and $\kappa_{ij} = 10$ kcal mol⁻¹ Å⁻² were chosen for the *CB8* simulations. The guest atom j was taken as the atom closest to the center of mass of the guest. The atom names in the input files were for *OA*: G0 = C6; G1 = C2; G2 = C9; G3 = C6; G4 = C1; G5 = C5; G6 = C6; G7 = C6. For *TEMOA*: G1 = C5; G2 = C9; G3 = C6; G4 = C7; G5 = C5; G6 = C6; G7 = C6. The four host atom names i in *OA* and *TEMOA* were C45, C51, C57, C63. The selection of R_{ij} and D_{ij} parameters were based on an average distance of 4.9 Å measured between these four host atoms and the guest atom C6 in G0 in the input files provided by the organizers. For *CB8* the guest and host atom names were: G0 = (C11, C3, C10, C18, C26), G1 = (C20, C4, C12, C22, C31), G2 = (C13, C8, C18, C26, C32), G3 = (C18, C2, C10, C16, C24), G4 = (C5, C4, C10, C16, C24), G5 = (C7, C6, C14, C22, C28), G6 = (C5, C6, C14, C22, C32), G7 = (C6, C4, C10, C16, C24), G8 = (C10, C4, C14, C20, C26), G9 = (C6, C4, C12, C20, C28), G10 = (C7, C2, C10, C16, C24). The selection of R_{ij} and D_{ij} parameters were based on an average distance of 6.6 Å measured between the four host atoms and the guest atom in G0 in the input files provided by the organizers.

SAMPLing simulation protocols

For the SAMPLing leg of the challenge topologies and coordinate file for five replicates of *OA–G3*, *OA–G6* and *CB8–G3* were provided by the organizers for both the complex phase and the solvated phase simulations. All simulations were run for duration of 20 ns per window with SOMD with other simulation parameters identical to those used for SAMPL6 unless otherwise mentioned.

Estimation of free energy of binding and evaluation of dataset metrics

Free energy changes were computed by use of the multistate Bennet acceptance ratio MBAR method [43]. To achieve a more robust estimation of free energies, each simulation was repeated multiple times, using different initial velocities drawn from the Maxwell–Boltzmann distribution. Unless otherwise mentioned, the reported binding free energies are the mean of three runs, and statistical uncertainties are given one standard error of the mean.

As described in Ref. [13] for each model a population distribution for the determination coefficient R^2 , the mean unsigned error MUE and the Kendall τ parameters was computed by bootstrapping each free energy predictions for each host–guest dataset for ten thousand times. The resulting distributions may not be symmetric around the mean, thus uncertainties are reported with a 95% confidence interval.

Additionally, for the SAMPLing leg of the challenge, binding free energies were evaluated using *ModelB* by skipping the first 1.5 ns of each window, and using 1–100% of the rest of the dataset. Uncertainties were taken as the standard deviation output from *pymbar* and were propagated to obtain an uncertainty for the reported standard free energy of binding. The total wall-clock time was also estimated by summing up the wall-clock time for each λ window, in each phase and simulated process. The number of iterations was retrieved as the sum of the number of time-steps for each simulated process. For each host–guest replica 459,995,400 energy evaluations were carried on with an average wall-clock time of 245 h for *CB8* systems and 190 h for *OA*. All input files for the SAMPL6 and SAMPLing protocols are publically available in the repository <https://github.com/michellab/SAMPL6inputs>.

Results

SAMPL6 challenge

Results for the full SAMPL6 dataset are shown in Fig. 3 for each model without and with a buffer setup. As judged by mean unsigned error, *ModelA/no-buffer* is the least accurate protocol, with a MUE value ca. 5.7 kcal mol⁻¹. *ModelA/buffer* offers small improvements, with the MUE decreasing to ca. 5.1 kcal mol⁻¹. Addition of long-range dispersions and standard state correction terms in *ModelB* decreases errors further (MUE ca. 3.9 and 3.4 kcal mol⁻¹ for the *no-buffer* and *buffer* setups respectively). *ModelC* improves over *ModelB* with MUE values ca. 1.4 and 1.6 kcal mol⁻¹ for the *no-buffer* and *buffer* setups respectively. Thus, the additional counter-ions in the buffer setup

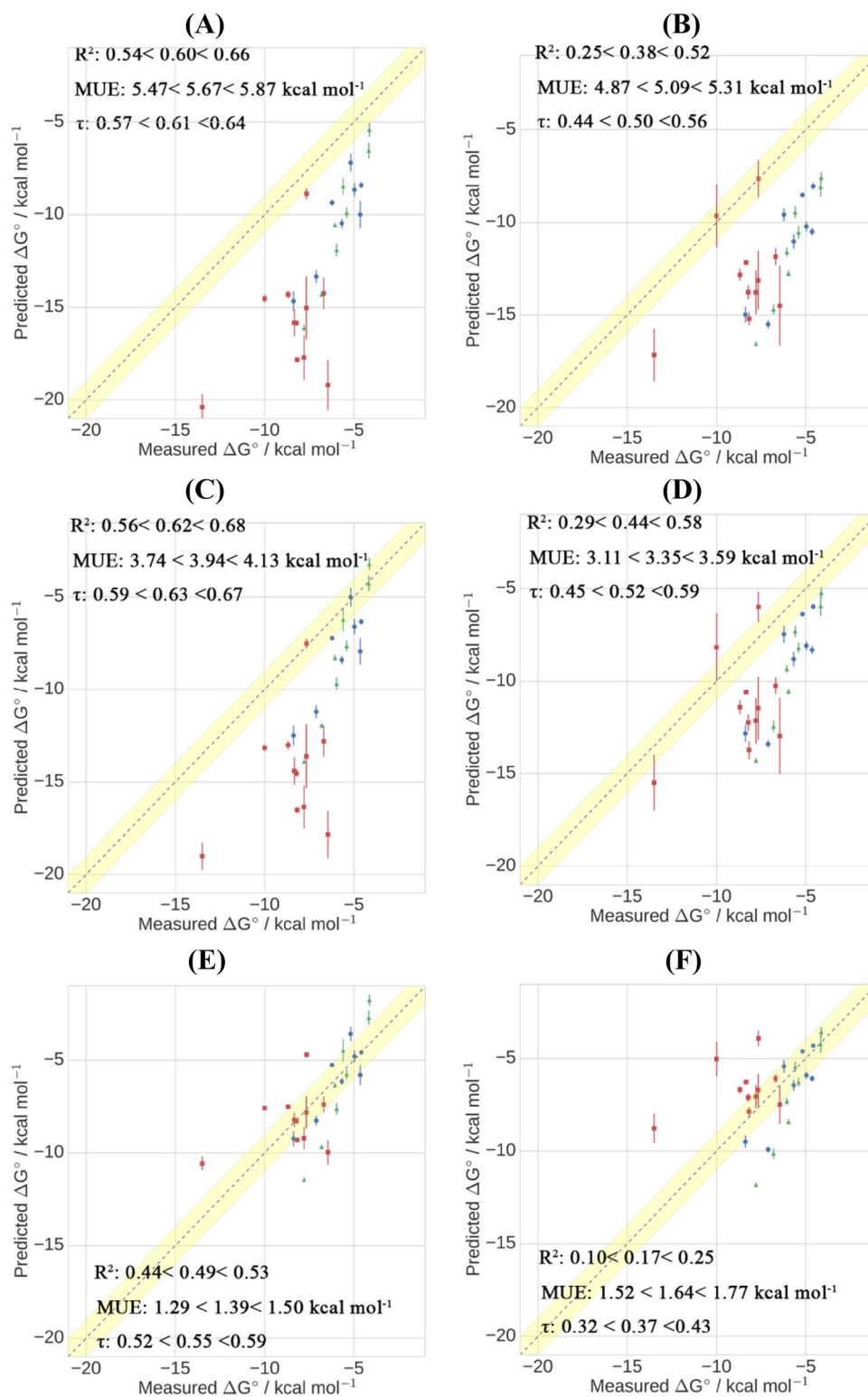
improve accuracy for *ModelA* and *ModelB* but not *ModelC*. This could be because the SAMPL5 calculations were carried out with a *no-buffer* setup [13], and the empirical correction terms used in *ModelC* do not transfer to a *buffer* setup.

Ranking of the protocols according to correlation with experimental data yields a different outcome. *ModelA/no-buffer* and *ModelB/no-buffer* perform similarly well with R^2 and τ values ca. 0.6, and a small decrease in predictive power is observed for *ModelC/no-buffer* but this is only significant for R^2 . This drop is observed because the empirical correction term works well to bring the OA host–guest binding energies in line with the experimental values, but leads to a tendency to underestimate the *CB8* binding energies. The use of a buffer also appears detrimental to predictive power, with all *buffer* protocols giving significant decreases in R^2 and τ parameters with respect to the equivalent *no-buffer* protocol.

Inspection of the results for the OA subset (Tables 1, 2) shows that *ModelB* and *ModelC* significantly improve the MUE over *ModelA* but not for R^2 or τ metrics that are ca. 0.7 and 0.5 respectively. The *buffer* protocol worsens MUE over the *no-buffer* protocol but does not influence predictive power. The same picture holds for the TEMOA subset, with improvements for MUE only observed upon switching from *ModelA* to *ModelB* and *ModelC*. Switching from *no-buffer* to *buffer* gives significant worsening of the MUE for *ModelA* and *ModelB*. The R^2 and τ metrics are high throughout (ca. 0.9 and 0.8) and insensitive to the various protocols. For the *CB8* subset dramatic improvements in MUE are also observed as correction term are introduced (*ModelA/no-buffer* MUE ca. 7.3 kcal mol⁻¹ vs. *ModelC/no-buffer* MUE ca. 1.6 kcal mol⁻¹). Unlike for the octa-acid guests switching from a *no-buffer* to *buffer* setup significantly improves the MUE for *ModelA* and *ModelB*, but not for *ModelC* where the MUE worsens. Thus, the buffer effects are host–guest dependent. For the OA and TEMOA hosts, the guests are negatively charged acids and explicit modelling of a buffer favors the binding process (average change in binding energies of -0.9 kcal mol⁻¹ for *ModelB*). For the *CB8* host, the guests are positively charged amines and explicit modelling of a buffer disfavors the binding process (average change in binding energies of $+3.1$ kcal mol⁻¹ for *ModelB*). The effect is particularly pronounced for some *CB8* guests, e.g. the binding energies of G3, G4 and G7 increase by more than 4 kcal mol⁻¹ upon switching from a *no-buffer* to *buffer* protocol. None of the models tested yield significant predictive power with R^2 and τ metrics ca. 0.1.

The largest outliers for *CB8* are guests G3, G4, G5 and G8. In particular the binding free energies of G3, G5 and G8 are lower than the experimental data by about 10 kcal mol⁻¹ with *ModelA/no-buffer* or *ModelB/no-buffer*. The statistical

Fig. 3 Comparison of the predicted and measured binding free energies for **a** *ModelA/no-buffer*, **b** *ModelA/buffer*, **c** *ModelB/no-buffer*, **d** *ModelB/buffer*, **e** *ModelC/no-buffer*, **f** *ModelC/buffer* for the 27 host–guest systems. The grey line denotes perfect correlation between predictions and measurements, while the yellow shaded region indicates a ± 1 kcal mol⁻¹ error bound. OA systems are colored in blue, TEMOA in green and CB8 in red



errors are also larger than for the octa-acids, suggesting greater challenges for converging free energy changes in *CB8* over the simulated time-scales. Switching to a *buffer* protocol decreases free energies of binding, and by up to ca. 5 kcal mol⁻¹ for G3 and G8.

Among octa-acids the models correctly capture interesting trends in the experimental data. For instance, the models correctly predict that G7 binds significantly worse to TEMOA than to OA. The bulkiness of the two methyl groups β to the carboxylic acid moiety hinders positioning

Table 1 Results for all three models (*no-buffer* protocol) for individual host–guest families

Guest	ΔG_{bind} ModelA	ΔG_{bind} ModelB	ΔG_{bind} ModelC	ΔG_{bind} experimental
OA				
G0	-10.5 ± 0.3	-8.4 ± 0.2	-6.1 ± 0.2	-5.7 ± 0.1
G1	-10.0 ± 0.7	-7.9 ± 0.7	-5.8 ± 0.5	-4.7 ± 0.1
G2	-14.7 ± 0.5	-12.5 ± 0.5	-9.2 ± 0.4	-8.4 ± 0.1
G3	-7.2 ± 0.5	-5.0 ± 0.5	-3.6 ± 0.4	-5.2 ± 0.1
G4	-13.3 ± 0.3	-11.2 ± 0.4	-8.3 ± 0.3	-7.1 ± 0.1
G5	-8.4 ± 0.2	-6.3 ± 0.2	-4.6 ± 0.1	-4.6 ± 0.1
G6	-8.7 ± 0.4	-6.6 ± 0.4	-4.8 ± 0.3	-5.0 ± 0.1
G7	-9.4 ± 0.2	-7.2 ± 0.1	-5.3 ± 0.1	-6.2 ± 0.1
R ²	$0.62 < 0.75 < 0.85$	$0.62 < 0.73 < 0.84$	$0.62 < 0.74 < 0.84$	
MUE	$4.17 < 4.41 < 4.66$	$2.15 < 2.37 < 2.60$	$0.65 < 0.82 < 1.00$	
τ	$0.43 < 0.54 < 0.64$	$0.43 < 0.54 < 0.64$	$0.43 < 0.54 < 0.64$	
TEMOA				
G0	-10.6 ± 0.1	-8.3 ± 0.2	-6.4 ± 0.2	-6.1 ± 0.1
G1	-11.9 ± 0.3	-9.7 ± 0.4	-7.6 ± 0.3	-6.0 ± 0.1
G2	-14.3 ± 0.1	-11.9 ± 0.1	-9.7 ± 0.1	-6.8 ± 0.1
G3	-8.5 ± 0.5	-6.2 ± 0.7	-4.5 ± 0.2	-5.6 ± 0.1
G4	-16.1 ± 0.2	-13.9 ± 0.1	-11.4 ± 0.1	-7.8 ± 0.1
G5	-6.5 ± 0.4	-4.3 ± 0.4	-2.7 ± 0.4	-4.2 ± 0.1
G6	-9.9 ± 0.3	-7.6 ± 0.3	-5.8 ± 0.3	-5.4 ± 0.1
G7	-5.4 ± 0.4	-3.2 ± 0.3	-1.8 ± 0.3	-4.1 ± 0.1
R ²	$0.90 < 0.93 < 0.96$	$0.89 < 0.93 < 0.97$	$0.91 < 0.94 < 0.96$	
MUE	$4.47 < 4.66 < 4.84$	$2.50 < 2.67 < 2.84$	$1.58 < 1.72 < 1.86$	
τ	$0.85 < 0.86 < 0.87$	$0.78 < 0.85 < 0.86$	$0.85 < 0.86 < 0.87$	
CB8				
G0	-14.3 ± 0.9	-12.8 ± 0.9	-7.4 ± 0.4	-6.7 ± 0.1
G1	-8.9 ± 0.3	-7.5 ± 0.3	-4.7 ± 0.1	-7.7 ± 0.1
G2	-15.0 ± 1.7	-13.6 ± 1.7	-7.8 ± 0.9	-7.7 ± 0.1
G3	-19.2 ± 1.4	-17.8 ± 1.3	-10.0 ± 0.7	-6.5 ± 0.1
G4	-17.7 ± 1.2	-16.4 ± 1.2	-9.2 ± 0.6	-7.8 ± 0.1
G5	-17.8 ± 0.1	-16.5 ± 0.2	-9.3 ± 0.1	-8.2 ± 0.1
G6	-15.8 ± 0.7	-14.4 ± 0.7	-8.2 ± 0.4	-8.3 ± 0.1
G7	-14.5 ± 0.2	-13.2 ± 0.2	-7.6 ± 0.1	-10.0 ± 0.1
G8	-20.4 ± 0.7	-19.0 ± 0.7	-10.6 ± 0.4	-13.5 ± 0.1
G9	-14.3 ± 0.2	-13.0 ± 0.3	-7.5 ± 0.1	-8.7 ± 0.1
G10	-15.9 ± 0.2	-14.5 ± 0.2	-8.3 ± 0.1	-8.2 ± 0.1
R ²	$0.04 < 0.12 < 0.23$	$0.04 < 0.12 < 0.23$	$0.04 < 0.12 < 0.23$	
MUE	$6.90 < 7.33 < 7.76$	$5.57 < 5.99 < 6.42$	$1.40 < 1.57 < 1.76$	
τ	$-0.09 < 0.04 < 0.20$	$-0.09 < 0.05 < 0.20$	$-0.09 < 0.05 < 0.20$	

of the guest in the smaller TEMOA cavity (Fig. 1a). The most significant outlier is G2 for which the models are unable to reproduce the significantly decreased binding energetics for TEMOA versus OA. A possible reason for this discrepancy is that the different ring puckering motions of the cyclohexenyl moiety in G2 may have been poorly sampled with the simulation protocols employed here.

SAMPLing challenge

Convergence plots for the calculated binding free energies of the three host–guests CB8–G3, OA–G3 and OA–G6 are presented in Fig. 4.¹ Figure 4a shows that for CB8–G3 the

¹ The SAMPLing free energies submitted on 01/19/2018 (5732q) were incorrectly evaluated due to a software bug. The results reported in this manuscript have been obtained after closure of the competition.

Table 2 Results for all three models (*buffer* protocol) for individual host–guest families

Guest	ΔG_{bind} ModelA	ΔG_{bind} ModelB	ΔG_{bind} ModelC	ΔG_{bind} experimental
OA				
G0	-11.0 ± 0.4	-8.8 ± 0.4	-6.5 ± 0.3	-5.7 ± 0.1
G1	-10.0 ± 0.2	-8.3 ± 0.2	-6.1 ± 0.2	-4.7 ± 0.1
G2	-15.0 ± 0.4	-12.8 ± 0.5	-9.5 ± 0.3	-8.4 ± 0.1
G3	-8.5 ± 0.1	-6.4 ± 0.1	-4.6 ± 0.0	-5.2 ± 0.1
G4	-15.5 ± 0.2	-13.4 ± 0.2	-9.9 ± 0.2	-7.1 ± 0.1
G5	-8.1 ± 0.2	-6.0 ± 0.2	-4.3 ± 0.1	-4.6 ± 0.1
G6	-10.2 ± 0.3	-8.1 ± 0.2	-5.9 ± 0.2	-5.0 ± 0.1
G7	-9.6 ± 0.3	-7.5 ± 0.5	-5.4 ± 0.4	-6.2 ± 0.1
R ²	$0.67 < 0.74 < 0.80$	$0.63 < 0.71 < 0.78$	$0.63 < 0.71 < 0.78$	
MUE	$4.97 < 5.14 < 5.31$	$2.87 < 3.05 < 3.23$	$0.95 < 1.08 < 1.22$	
τ	$0.50 < 0.56 < 0.64$	$0.43 < 0.52 < 0.64$	$0.43 < 0.52 < 0.64$	
TEMOA				
G0	-11.6 ± 0.2	-9.3 ± 0.2	-7.3 ± 0.2	-6.1 ± 0.1
G1	-12.8 ± 0.2	-10.6 ± 0.2	-8.4 ± 0.2	-6.0 ± 0.1
G2	-14.7 ± 0.3	-12.4 ± 0.3	-10.1 ± 0.3	-6.8 ± 0.1
G3	-9.5 ± 0.3	-7.3 ± 0.3	-5.5 ± 0.3	-5.6 ± 0.1
G4	-16.6 ± 0.1	-14.3 ± 0.1	-11.8 ± 0.1	-7.8 ± 0.1
G5	-8.1 ± 0.5	-5.9 ± 0.5	-4.2 ± 0.5	-4.2 ± 0.1
G6	-10.5 ± 0.3	-8.2 ± 0.3	-6.3 ± 0.3	-5.4 ± 0.1
G7	-7.6 ± 0.3	-5.2 ± 0.3	-3.6 ± 0.3	-4.1 ± 0.1
R ²	$0.89 < 0.92 < 0.95$	$0.89 < 0.93 < 0.96$	$0.89 < 0.93 < 0.96$	
MUE	$5.48 < 5.66 < 5.84$	$3.23 < 3.41 < 3.59$	$1.51 < 1.63 < 1.77$	
τ	$0.78 < 0.84 < 0.85$	$0.78 < 0.85 < 0.86$	$0.78 < 0.84 < 0.85$	
CB8				
G0	-11.9 ± 0.5	-10.2 ± 0.5	-6.1 ± 0.2	-6.7 ± 0.1
G1	-7.6 ± 1.0	-6.0 ± 0.8	-3.9 ± 0.4	-7.7 ± 0.1
G2	-13.1 ± 1.6	-11.5 ± 1.7	-6.7 ± 0.9	-7.7 ± 0.1
G3	-14.5 ± 2.2	-13.0 ± 2.1	-7.5 ± 1.1	-6.5 ± 0.1
G4	-13.8 ± 1.2	-12.1 ± 1.2	-7.1 ± 0.6	-7.8 ± 0.1
G5	-15.2 ± 0.4	-13.7 ± 0.5	-7.9 ± 0.3	-8.2 ± 0.1
G6	-12.2 ± 0.1	-10.6 ± 0.0	-6.3 ± 0.0	-8.3 ± 0.1
G7	-9.7 ± 1.7	-8.2 ± 1.8	-5.0 ± 0.9	-10.0 ± 0.1
G8	-17.2 ± 1.4	-15.5 ± 1.5	-8.8 ± 0.8	-13.5 ± 0.1
G9	-12.3 ± 0.3	-11.4 ± 0.4	-6.7 ± 0.2	-8.7 ± 0.1
G10	-13.8 ± 0.4	-12.2 ± 0.4	-7.1 ± 0.2	-8.2 ± 0.1
R ²	$0.00 < 0.13 < 0.33$	$0.00 < 0.13 < 0.35$	$0.00 < 0.13 < 0.35$	
MUE	$4.07 < 4.59 < 5.13$	$2.96 < 3.52 < 4.08$	$1.78 < 2.06 < 2.34$	
τ	$-0.12 < 0.06 < 0.27$	$-0.09 < 0.10 < 0.30$	$-0.09 < 0.10 < 0.30$	

binding free energy estimate obtained using the full simulation dataset is -13.8 ± 0.7 kcal mol⁻¹. Although the uncertainties are high the mean free energy rapidly settles around -14 kcal mol⁻¹ and similar estimates would have been obtained with about 20% of the simulation duration. The calculated binding free energies are consistent with those obtained for this host–guest with the SAMPL6 protocol (-13.0 ± 2.1 kcal mol⁻¹, Table 2). The SAMPLing reference binding free energy computed by the organizers using the software YANK is significantly different and more precise

(-10.8 ± 0.2 kcal mol⁻¹) [56]. The reference value is also in better agreement with experimental data, though substantial differences remain (-6.5 ± 0.1 kcal mol⁻¹). It appears at least 60% of the simulation duration is needed to eliminate drifts in the running average for the reference calculation.

For OA–G3 (Fig. 4b) the binding free energies computed with SOMD and by the organizers are similarly precise and converge to -5.7 ± 0.1 kcal mol⁻¹ and -6.7 ± 0.1 kcal mol⁻¹ respectively. The SOMD SAMPLing free energies are as precise but more

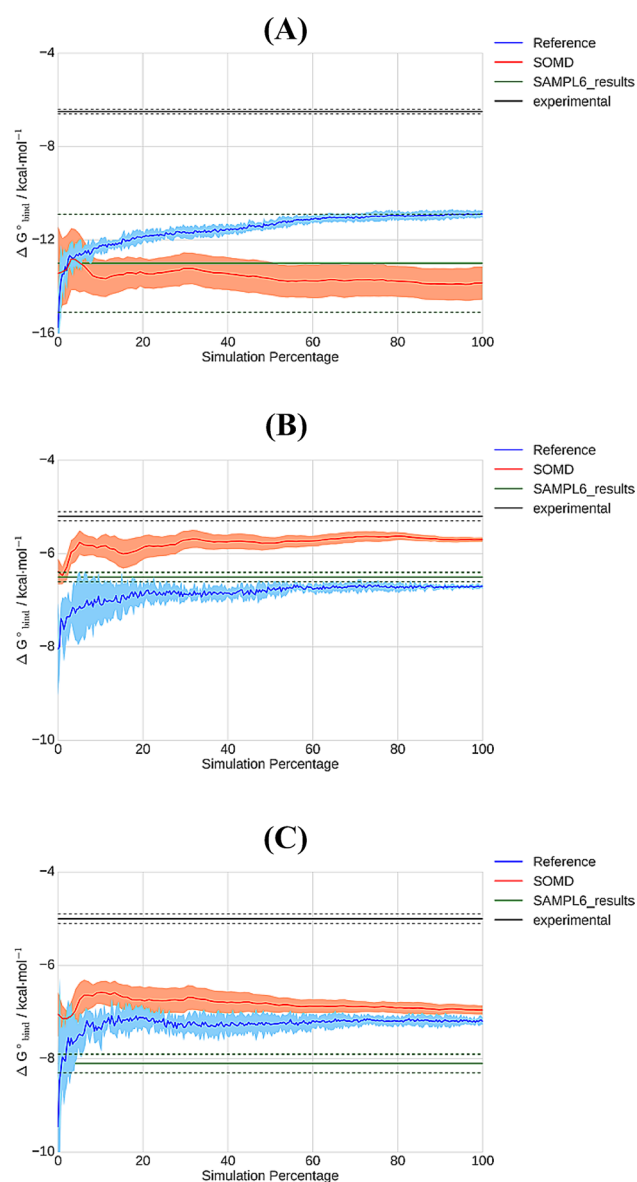


Fig. 4 Comparison of standard binding free energies computed with SOMD (red) to SAMPLing reference values (blue) for *CB8-G3* (a), *OA-G3* (b) and *OA-G6* (c). Bold lines denote the average free energy from five replicate simulations started from different coordinates. Shaded areas denote $\pm 1\sigma$. The SAMPL6 and experimental results are depicted with green and black lines respectively, and the dotted lines denote $\pm 1\sigma$

accurate than the SOMD SAMPL6 free energies (-6.4 ± 0.1 kcal mol $^{-1}$, Table 2) in comparison with experimental data (-5.2 ± 0.1 kcal mol $^{-1}$). The running average for both protocols is stable after ca. 20% of the simulation duration. For *OA-G6* the SOMD and organizer's free energies rapidly converge to very similar values (-6.9 ± 0.1 kcal mol $^{-1}$ vs. -7.1 ± 0.1 kcal mol $^{-1}$ respectively). These figures are in better agreement with

experiment (-5.0 ± 0.1 kcal mol $^{-1}$) than the SAMPL6 SOMD free energies (-8.1 ± 0.2 kcal mol $^{-1}$).

Overall comparison of free energies estimated from the SAMPL6 and SAMPLing protocols shows that averaging results over multiple starting host-guest structures improved agreement of predictions with experiment for *OA-G3* and *OA-G6* but not *CB8*. No clear reason emerges to explain differences in binding free energies computed by SOMD and YANK.

Conclusions

AFE calculations were employed to estimate standard binding free energies for 27 host-guests in the SAMPL6 competition. Protocols similar to that used in the SAMPL5 competition were adopted (*ModelA/no-buffer* and *ModelB/no-buffer*) [13], leading to results of comparable performance to SAMPL5 (SAMPL6 *ModelB/no-buffer* R^2 ca. 0.6, MUE 3.9 kcal mol $^{-1}$, $N=27$ vs. SAMPL5 *ModelC* R^2 ca. 0.7, MUE 3.4 kcal mol $^{-1}$, $N=22$). The reasons for the systematic overestimation of free energies of binding remain unclear; this could be because of a neglect of long-range correction term to electrostatics, or use of non-polarizable force-fields.

Additionally, an empirical correction term derived by a linear regression approach against SAMPL5 data was devised to correct for systematic errors in the free energy calculation protocol (*ModelC/no-buffer*). This leads to significant improvements in mean-unsigned error but a slight decrease in correlation with experimental trends (MUE ca. 1.4 kcal mol $^{-1}$, R^2 ca. 0.5). High accuracy predictions and correlations with experimental data were achieved for the *OA* and *TEMOA* hosts, but *CB8* proved more challenging, with significantly higher uncertainties in the computed binding free energies and poor correlation with experiment.

The influence of the modelled buffer on the computed binding free energies was also investigated. The main finding is that explicit modelling of the buffer enhances binding of negatively charged guests to *OA* and *TEMOA*, and weakens binding of positively charged guests to *CB8*. Overall the MUE for the dataset (*ModelA* and *ModelB*) decreases by about 0.6 kcal mol $^{-1}$ because the *CB8* binding energies are more in line with experimental data. However, this improvement is also accompanied by a drop of ca. 0.2 in R^2 . The empirical correction term derived against SAMPL5 data is incompatible with a protocol that models explicitly a buffer, presumably because no buffer was modelled in the SAMPL5 calculations [13].

With respect to other SAMPL6 submissions the results obtained with SOMD were encouraging and among the top performing models for *OA* and *TEMOA* as judged

by R^2 and MUE metrics. CB8 proved challenging for most participating groups. SOMD *ModelC/no-buffer* gave the lowest MUE values among all submissions (ca. 1.5 kcal mol⁻¹), but the predictive power was insignificant (R^2 ca. 0.1) [31].

The OA–G3 and OA–G6 binding free energies computed with the SAMPLing protocol were significantly different from those computed with SAMPL6 protocol (0.7 and 1.2 kcal mol⁻¹ respectively). A standard practice in our group is to at least estimate uncertainties in computed binding free energies from triplicate runs initiated from the same input coordinates. This gives a reasonable estimate of the extent to which free energies are reproducible *given a starting condition*, but can also give a misleading impression of convergence. Where multiple reasonable poses can be produced, efforts are better spent evaluating free energies with simulations started from different input coordinates. Comparison of SOMD's free energies with the reference values (YANK) provided by the organizers yields a mixed picture, with a substantially significant difference (CB8–G3, 3 ± 0.7 kcal mol⁻¹), a moderate difference (OA–G3, 1 ± 0.2 kcal mol⁻¹), and an insignificant difference (OA–G6 0.2 ± 0.2 kcal mol⁻¹). There are several algorithmic differences between the two codes that could explain discrepancies, a notable one being an atom-based Barker–Watts reaction-field treatment of long-range electrostatics (SOMD) versus PME (YANK). Other differences exist around the treatment of soft-cores, the coupling of non-bonded and bonded interactions with the λ schedule, and electrostatic correction terms for charged guests. More systematic reproducibility studies on larger datasets will be needed to isolate the factors that contribute to the observed variability. Such efforts are important to validate the robustness and transferability of molecular simulation algorithms.

Acknowledgements Julien Michel is supported by a University Research Fellowship from the Royal Society. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC Grant Agreement No. 336289.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Jorgensen WL (2004) The many roles of computation in drug discovery. *Science* 303(5665):1813–1818
- Michel J (2014) Current and emerging opportunities for molecular simulations in structure-based drug design. *Phys Chem Chem Phys* 16(10):4465–4477
- Larson SM, Snow CD, Shirts M, Pande VS (2009) Folding@Home and Genome@Home: using distributed computing to tackle previously intractable problems in computational biology. *ArXiv09010866 Phys. Q-Bio*
- Shaw DE, Dror RO, Salmon JK, Grossman JP, Mackenzie KM, Bank JA, Young C, Deneroff MM, Batson B, Bowers KJ et al (2009) Millisecond-scale molecular dynamics simulations on anton. In Proceedings of the conference on high performance computing networking, storage and analysis; SC'09; ACM, New York, pp 39:1–39:11
- Kohlhoff K, Shukla D, Lawrenz M, Bowman G, Konerding D, Belov D, Altman R, Pande V (2014) Cloud-Based simulations on Google Exacycle Reveal ligand modulation of GPCR activation pathways. *Nat Chem* 6:15
- Michel J, Foloppe N, Essex JW (2010) Rigorous free energy calculations in structure-based drug design. *Mol Inform* 29(8–9):570–578
- Deng Y, Roux B (2006) Calculation of standard binding free energies: aromatic molecules in the T4 lysozyme L99A mutant. *J Chem Theory Comput* 2(5):1255–1273
- Chang C-E, Gilson MK (2004) Free energy, entropy, and induced fit in host–guest recognition: calculations with the second-generation mining minima algorithm. *J Am Chem Soc* 126(40):13156–13164
- Mey ASJS, Juárez-Jiménez J, Michel J (2017) Impact of domain knowledge on blinded predictions of binding energies by alchemical free energy calculations. *J Comput Aided Mol Des* 32:199–210. <https://doi.org/10.1007/s10822-017-0083-9>
- Calabrò G, Woods CJ, Powlesland F, Mey ASJS, Mulholland AJ, Michel J (2016) Elucidation of nonadditive effects in protein–ligand binding energies: thrombin as a case study. *J Phys Chem B* 120(24):5340–5350
- Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J et al (2015) Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc* 137(7):2695–2703
- Aldeghi M, Heifetz A, Bodkin MJ, Knapp S, Biggin PC (2015) Accurate calculation of the absolute free energy of binding for drug molecules. *Chem Sci* 7(1):207–218
- Bosisio S, Mey ASJS, Michel J (2017) Blinded predictions of host-guest standard free energies of binding in the SAMPL5 challenge. *J Comput Aided Mol Des* 31(1):61–70
- Bosisio S, Mey ASJS, Michel J (2016) Blinded predictions of distribution coefficients in the SAMPL5 challenge. *J Comput Aided Mol Des* 30(11):1101–1114
- Bannan CC, Burley KH, Chiu M, Shirts MR, Gilson MK, Mobley DL (2016) Blind prediction of cyclohexane–water distribution coefficients from the SAMPL5 challenge. *J Comput Aided Mol Des* 30(11):927–944
- Rodil A, Bosisio S, Ayoub MS, Quinn L, Cordes DB, Slawin AMZ, Murphy CD, Michel J, O'Hagan D (2018) Metabolism and hydrophilicity of the polarised 'Janus Face' all-*cis* tetrafluorocyclohexyl ring, a candidate motif for drug discovery. *Chem Sci* 9(11):3023–3028
- Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS (2011) Alchemical free energy methods for drug discovery: progress and challenges. *Curr Opin Struct Biol* 21(2):150–160
- Chen I-J, Foloppe N (2011) Is conformational sampling of drug-like molecules a solved problem? *Drug Dev Res* 72(1):85–94

19. Souaille M, Roux B (2001) Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput Phys Commun* 135:40–57
20. Li H, Fajer M, Yang W (2007) Simulated scaling method for localized enhanced sampling and simultaneous “alchemical” free energy simulations: a general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations. *J Chem Phys* 126(2):024106
21. Halgren TA, Damm W (2001) Polarizable force fields. *Curr Opin Struct Biol* 11(2):236–242
22. Kastenholz MA, Hünenberger PH (2004) Influence of artificial periodicity and ionic strength in molecular dynamics simulations of charged biomolecules employing lattice-sum methods. *J Phys Chem B* 108(2):774–788
23. Reif Maria M, Oostenbrink C (2013) Net Charge changes in the calculation of relative ligand-binding free energies via classical atomistic molecular dynamics simulation. *J Comput Chem* 35(3):227–243
24. Rocklin GJ, Mobley DL, Dill KA, Hünenberger PH (2013) Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects. *J Chem Phys* 139(18):184103
25. Mey ASJS, Juárez-Jiménez J, Hennessy A, Michel J (2016) Blinded predictions of binding modes and energies of HSP90- α ligands for the 2015 D3R Grand Challenge. *Bioorg Med Chem* 24(20):4890–4899
26. Gaieb Z, Liu S, Gathiaka S, Chiu M, Yang H, Shao C, Feher VA, Walters WP, Kuhn B, Rudolph MG et al (2018) D3R Grand Challenge 2: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. *J Comput Aided Mol Des* 32(1):1–20
27. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS (2008) Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. *J Med Chem* 51(4):769–779
28. Mobley DL, Liu S, Cerutti DS, Swope WC, Rice JE (2012) Alchemical prediction of hydration free energies for SAMPL. *J Comput Aided Mol Des* 26(5):551–562
29. Peat TS, Dolezal O, Newman J, Mobley D, Deadman JJ (2014) Interrogating HIV integrase for compounds that Bind—a SAMPL challenge. *J Comput Aided Mol Des* 28(4):347–362
30. Mobley DL, Gilson MK (2017) Predicting binding free energies: frontiers and benchmarks. *Annu Rev Biophys* 46:531–558
31. Rizzi A, Murkli S, McNeill JN, Yao W, Sullivan M, Gilson MK, Chiu MW, Isaacs L, Gibb BC, Mobley DL et al (2018) Overview of the SAMPL6 host-guest binding affinity prediction challenge. *bioRxiv*. <https://doi.org/10.1101/371724>
32. Gan H, Benjamin CJ, Gibb BC (2011) Nonmonotonic assembly of a deep-cavity cavitand. *J Am Chem Soc* 133(13):4770–4773
33. Gibb CLD, Gibb BC (2014) Binding of cyclic carboxylates to octa-acid deep-cavity cavitand. *J Comput Aided Mol Des* 28(4):319–325
34. Sullivan MR, Sokkalingam P, Nguyen T, Donahue JP, Gibb BC (2017) Binding of carboxylate and trimethylammonium salts to octa-acid and TEMOA Deep-Cavity cavitands. *J Comput Aided Mol Des* 31(1):21–28
35. Gan H, Gibb BC (2013) Guest-mediated switching of the assembly state of a water-soluble deep-cavity cavitand. *Chem Commun* 49(14):1395–1397
36. Assaf KI, Nau WM (2014) Cucurbiturils: from synthesis to high-affinity binding and catalysis. *Chem Soc Rev* 44(2):394–418
37. Biedermann F, Scherman OA (2012) Cucurbit[8]uril mediated donor–acceptor ternary complexes: a model system for studying charge-transfer interactions. *J Phys Chem B* 116(9):2842–2849
38. Vázquez J, Remón P, Dsouza RN, Lazar AI, Arteaga JF, Nau WM, Pischel U (2014) A simple assay for quality binders to cucurbiturils. *Chem – Eur J* 20(32):9897–9901
39. Liu S, Ruspic C, Mukhopadhyay P, Chakrabarti S, Zavalij PY, Isaacs L (2005) The Cucurbit[n]uril family: prime components for self-sorting systems. *J Am Chem Soc* 127(45):15959–15967
40. Michel J, Essex JW (2010) Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J Comput Aided Mol Des* 24(8):639–658
41. Jorgensen WL, Buckner JK, Boudon S, Tirado-Rives J (1988) Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. *J Chem Phys* 89(6):3742–3746
42. Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys J* 72(3):1047–1069
43. Shirts MR, Mobley DL, Chodera JD, Pande VS (2007) Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J Phys Chem B* 111(45):13052–13063
44. McGann MFRED, Docking HYBRID (2012) Performance on standardized datasets. *J Comput Aided Mol Des* 26(8):897–906
45. McGann MFRED (2011) Pose prediction and virtual screening accuracy. *J Chem Inf Model* 51(3):578–596
46. Kelley BP, Brown SP, Warren GL, Muchmore SW (2015) POSIT: flexible shape-guided docking for pose prediction. *J Chem Inf Model* 55(8):1771–1780
47. ParmEd — ParmEd documentation. <http://parmed.github.io/ParmEd/html/index.html>. Accessed 29 Mar 2018
48. Case D, Cerutti DS, Cheatham T, Darden T, Duke R, Giese TJ, Gohlke H, Götz A, Greene D, Homeyer N et al (2017) Amber 2017. University of California, San Francisco
49. Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* 25(2):247–260
50. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* 25(9):1157–1174
51. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935
52. Woods C, Mey A, Calabro G, Michel J (2016) Sire molecular simulations framework. <https://www.siremol.org/>
53. Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, Ku JP, Beauchamp KA, Lane TJ, Wang L-P, Shukla D et al (2013) OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation. *J Chem Theory Comput* 9(1):461–469
54. Andersen HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. *J Chem Phys* 72(4):2384–2393
55. Tironi IG, Sperb R, Smith PE, van Gunsteren WF (1995) A generalized reaction field method for molecular dynamics simulations. *J Chem Phys* 102(13):5451–5459
56. <https://github.com/choderalab/yank>. Accessed 28 Aug 2018