J Comput Aided Mol Des (2012) 26:543–550
DOI 10.1007/s10822-011-9525-y

Predicting binding affinities of host-guest systems in the SAMPL3 blind challenge: the performance of relative free energy calculations

Gerhard König · Bernard R. Brooks

Received: 15 October 2011 / Accepted: 8 December 2011 / Published online: 24 December 2011

Abstract Relative free energy calculations based on molecular dynamics simulations are combined with available experimental binding free energies to predict unknown binding affinities of acyclic Cucurbituril complexes in the blind SAMPL3 competition. The predictions yield root mean square errors between 2.6 and 3.2 kcal/mol for seven host-guest systems. Those deviations are comparable to results for solvation free energies of small organic molecules. However, the standard deviations found in our simulations range from 0.4 to 2.4 kcal/mol, which indicates the need for better sampling. Three different approaches are compared. Bennett's Acceptance Ratio Method and thermodynamic integration based on the trapezoidal rule with 12 λ-points exhibit a root mean square error of 2.6 kcal/mol, while thermodynamic integration with Simpson's rule and 11 λ-points leads to a root mean square error of 3.2 kcal/mol. In terms of absolute median errors, Bennett's Acceptance Ratio Method performs better than thermodynamic integration with the trapezoidal rule (1.7 vs. 2.9 kcal/mol). Simulations of the deprotonated forms of the guest molecules exhibit a poorer correspondence to experimental results with a root mean square error of 5.2 kcal/mol. In addition, a decrease of the buffer concentration by approximately 20 mM in the simulations raises the root mean square error to 3.8 kcal/mol.

Keywords Binding free energy calculations · Bennett's acceptance ratio · Thermodynamic integration · Protonation state · Buffer concentration

G. König (✉) · B. R. Brooks
National Institutes of Health, National Heart, Lung and Blood Institute, Laboratory of Computational Biology, 5635 Fishers Lane, T-900 Suite, Rockville, MD 20852, USA
e-mail: gerhard.koenig@nih.gov

Introduction

So-called "free energy simulations" are considered among the most accurate and general methodologies in the field of computational chemistry. They provide means to study diverse processes such as the binding affinities of ligands [1, 2], enzymatic reactions [3], the solvation of organic molecules [4], and the effect of point mutations [5].

However, most applications of free energy simulations do not provide reliable data on the accuracy of the method, since either there is no reference data for an assessment of its quality, or the free energy calculations were conducted after the experimental reference results became available. In the latter case, the reference results lead to a selection process, where simulations with a high agreement with experiment become published, while simulations with a poor agreement are not disclosed. Since only successes are reported this way, an unrealistically positive picture is presented.

Such forms of bias can be avoided by employing blind studies, where the reference results are not known *a priori*. In computational chemistry, the SAMPL blind challenges were established in 2007 for assessing methods for the prediction of free energies [6–8]. Those challenges provided data on the accuracies of solute-water interactions in current force fields. In SAMPL0 and SAMPL1, the root mean square errors (RMSE) of solvation free energy calculations based on molecular dynamics simulations with explicit solvent ranged between 1.3 [6] and 3.5 kcal/mol [7, 9]. These results provide a useful range of the errors that can be expected from this method for small systems. For example, the same approach yielded a RMSE of 2.8 kcal/mol when predicting the solvation free energies of 23 small organic compounds in the SAMPL2 competition [10].

In SAMPL3, the prediction of binding affinities of host-guest systems (ΔG_{bind}) was added to the list of challenges. These systems include Cucurbituril molecular containers that are able to selectively bind ligands with cationic groups. Through their binding-pocket-like structure, host-guest systems exhibit many of the features of protein-ligand complexes, while still being small enough to be computationally tractable. No global conformational changes or unfolding events can occur during the simulation. This relative simplicity and robustness makes them a very useful benchmark system for computational methods. For example, Moghaddam et al. [11] employed M2 free energy calculations to design guest systems with ultrahigh affinity to Cucurbit-7-uril, obtaining RMSE from experimental results between 2.7 and 4.6 kcal/mol for all compounds included in their study.

In a recent publication, Ma et al. [12] described the synthesis of the acyclic Cucurbituril congener employed in the SAMPL3 challenge. They also tested its function as a host to a structurally diverse set of ammonium ions. This study included experimentally determined binding constants between $\sim 10^5$ and 10^9 M^{-1} for 26 guest molecules. The present study employs these experimental binding affinities as a starting point of binding affinity predictions. The absolute binding free energy predictions (ΔG_{bind}^{target}) are a combination of the experimentally derived absolute binding free energy of a reference molecule (ΔG_{bind}^{ref}) as given in Ref. [12] and computed relative binding free energies to the target molecules of the SAMPL3 competition ($\Delta\Delta G_{bind}$).

$$\Delta G_{bind}^{target} = \Delta G_{bind}^{ref} + \Delta\Delta G_{bind}^{ref \rightarrow target} \quad (1)$$

The reference molecules [12] and the corresponding targets of the SAMPL3 competition are shown on the left and the right side of Fig. 1. By employing relative (alchemical) free energy calculations rather than absolute calculations, the mutations of the system are relatively mild, i.e. most interactions within the complex are only slightly affected by the transformation. Through cancelation of errors, we expect this approach to reduce the errors due to the imperfections of the force field [13].

To calculate such relative free energy differences, several free energy methods are available, whereof Bennett's acceptance ratio method (BAR) [14] and thermodynamic integration (TI) [15] are among the most widely used (for a short description of the two methods, see the "Methods"). While BAR is generally considered more efficient than TI, a recent publication by Bruckner and Boresch suggests that TI can be as efficient as BAR [16] if a good numerical quadrature scheme is employed. If the available simulation lengths are short, TI sometimes even outperforms BAR in terms of efficiency. Therefore, we decided to employ both

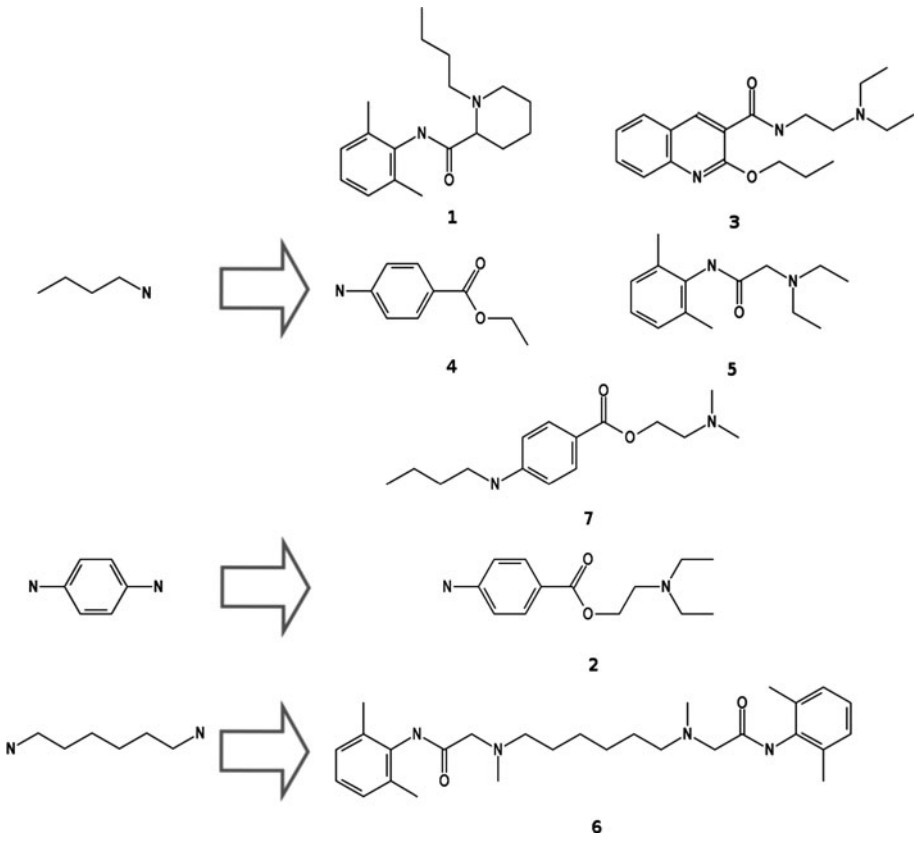

Fig. 1 Chemical structures of the guest molecules. Reference molecules of known binding affinity are shown on the left side of the *arrows*, and the corresponding target molecules of the free energy calculations for the SAMPL3 challenge are shown on the *right side*

BAR and TI for the analysis of the trajectories, using both the trapezoidal (TI-TR) and Simpson's rule (TI-SI) for the numerical quadrature step in TI. This allows a comparison of the relative competitiveness of those three methods.

The remainder of this paper is organized as follows. First, we outline the methods employed in more detail. We then present the results for BAR, TI-TR and TI-SI, and assess their accuracy. We conclude with a short discussion on the influence of parameters such as the protonation state of the guest molecule and the buffer concentration on the binding free energy results.

Methods

We calculated the relative binding free energies of the seven guest molecules in the SAMPL3 challenge (labeled **1–7** in Fig. 1, numbers shown in bold). The relative free energy calculations were started from reference molecules of known binding affinity as published by Ma et al. [12] (numbers in italics). In particular, structures *26*, *19* and *11* from Ref. [12] were employed (in order of appearance from top to bottom on the left side of Fig. 1). Based on the reported binding constants, their absolute binding free energies are −7.0, −8.0 and −11.3 kcal/mol. Reference molecule *26* was employed for molecules **1, 3, 4, 5** and **7**. We assumed that those molecules contain only a single protonated amine group. For the other two reference molecules *19* and *11* and their corresponding target molecules **2** and **6** we assumed that they contain two protonated amine groups. In our simulations, the four carboxyl groups in the host molecule were deprotonated. Binding affinities of the two stereoisomers of guest molecule **1** were calculated separately (named **1S** and **1R**). The presented data for **1** are the average of the **1S** and **1R** results. All free energy calculations were conducted with CHARMM [17, 18], using the PERT module of CHARMM and the CHARMM General Force Field for organic molecules (CGenFF) [19], program version 0.9.1 beta, as provided on http://www.paramchem.org.

Outline of the free energy methods

TI [15], involves numerical quadrature to determine the free energy difference between two states 0 and 1. Between 0 and 1, several intermediate states can be generated by mixing the respective potential energy functions U. The mixing ratio between U_0 for state 0 and U_1 for state 1 is given by the factor λ (e.g.: $U(\lambda) = \lambda\, U_1 + (1 - \lambda)\, U_0$). In TI, λ is considered as a continuous variable that can be used for differentiation or integration. Using integration, the free energy difference between 0 and 1 is

$$\Delta G = \int_0^1 \frac{\partial G(\lambda)}{\partial \lambda} d\lambda \tag{2}$$

which leads to the equation for TI

$$\Delta G^{TI} = \int_0^1 d\lambda \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda \tag{3}$$

In practice, this integral is evaluated by conducting several simulations at discrete values of λ to evaluate $\langle \frac{\partial U(\lambda)}{\partial \lambda} \rangle_\lambda$ and then employing numerical quadrature to approximate the integral. This can be done by using the trapezoidal rule or numerical quadrature schemes of higher order such as Simpson's rule [20].

BAR [14], in contrast, requires two simulations. At each end point of the free energy calculation the potential energy differences are evaluated, using

$$\Delta G^{BAR} = -\beta^{-1} \ln \left(\frac{\langle f(U_0 - U_1 + C) \rangle_1}{\langle f(U_1 - U_0 - C) \rangle_0} \right) + C \tag{4}$$

where f denotes the Fermi function. The subscripts 0 and 1 indicate that the ensemble averages are calculated over all coordinate frames generated for the initial and final state. Bennett showed that C can found through an iterative procedure. The BAR equation can also be derived using maximum likelihood techniques [21].

Computation of relative binding affinities

Each relative binding free energy ($\Delta\Delta G_{bind}^{ref \rightarrow target}$) was calculated with the standard thermodynamic cycle, which includes two kinds of calculations: (1) Transforming the reference molecule to the target molecule in solution ($\Delta G_{\mathrm{H_2O}}^{ref \rightarrow target}$) and (2) conducting the same transformation while bound to the host ($\Delta G_{host}^{ref \rightarrow target}$). Thus,

$$\Delta\Delta G_{bind}^{ref \rightarrow target} = \Delta G_{host}^{ref \rightarrow target} - \Delta G_{\mathrm{H_2O}}^{ref \rightarrow target} \tag{5}$$

In all $\Delta G_{host}^{ref \rightarrow target}$ and $\Delta G_{\mathrm{H_2O}}^{ref \rightarrow target}$ simulations, 1492 TIP3P water molecules [22, 23] were present. Na^+ or Cl^- ions were added to neutralize the total charge of the system and an additional NaCl pair was included to improve the sampling and obtain an ionic strength similar to the experimental conditions (i.e, a buffer concentration between 40–50 mM). The simulation box was a truncated octahedron. The side length L of the cube from which the octahedron was generated was originally $L = 40.0$ Å. We used constant pressure during free energy simulations. Integration of the equations of motion was carried out with the velocity-Verlet algorithm as implemented in the TPCNTRL module of CHARMM [24]; the time step was

2 fs. The temperature was maintained at 300 K using two separate Nosé-Hoover thermostats [25] for solute and solvent. SHAKE [26] was used to keep the water geometry rigid. Lennard-Jones interactions were switched off between 10 and 12 Å, while electrostatic interactions were computed with the Particle Mesh Ewald method [27]. Each host-guest system was equilibrated for 200 ps before production.

All alchemical mutations were split into 11–12 λ intermediate steps, using soft core Lennard Jones and electrostatic interactions. 12 λ-points were used for the BAR and TI-TR results ($\lambda = 0.0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$). Since Simpson's rule requires an odd number of equally-spaced data points for numerical quadrature, the TI-SI results are based on 11 λ-points ($\lambda = 0.0, 0.1 \ldots 1.0$). Each λ-point was simulated for 3 ns and all simulations were carried out four times, starting from different initial random velocities, to allow the calculation of standard deviations. Thus, each result submitted to SAMPL3 was based on a total simulation time of about 288 ns, leading to a combined computational effort equivalent to 2.3 µs for all simulations taken together.

Free energy simulations with different protonation states

To determine the influence of the protonation state of the guest on the binding affinity result ($\Delta\Delta\Delta G_{bind}^{deprot}$), additional simulations were conducted after the deadline for the SAMPL3 competition. For this purpose, the protonated and deprotonated state of all reference and target molecules were simulated with implicit solvent to determine the free energy differences. Again, a thermodynamic cycle was employed, using a.) free energy simulations between the protonated and deprotonated state in water for the reference ($\Delta G_{\mathrm{H_2O}}^{ref.\mathrm{H^+}\rightarrow ref}$), as well as the target ($\Delta G_{\mathrm{H_2O}}^{target.\mathrm{H^+}\rightarrow target}$) and b.) the corresponding simulations while bound to the host molecule ($\Delta G_{host}^{ref.\mathrm{H^+}\rightarrow ref}$, $\Delta G_{host}^{target.\mathrm{H^+}\rightarrow target}$), leading to

$$\Delta\Delta G_{\mathrm{H_2O}}^{deprot} = \Delta G_{\mathrm{H_2O}}^{target.\mathrm{H^+}\rightarrow target} - \Delta G_{\mathrm{H_2O}}^{ref.\mathrm{H^+}\rightarrow ref} \quad (6)$$

$$\Delta\Delta G_{host}^{deprot} = \Delta G_{host}^{target.\mathrm{H^+}\rightarrow target} - \Delta G_{host}^{ref.\mathrm{H^+}\rightarrow ref} \quad (7)$$

so that

$$\Delta\Delta\Delta G_{bind}^{deprot} = \Delta\Delta G_{host}^{deprot} - \Delta\Delta G_{\mathrm{H_2O}}^{deprot} \quad (8)$$

The free energy differences were calculated using Langevin dynamics simulations with a friction coefficient of 5 ps^{-1} on all heavy atoms. Random forces were applied according to the target temperature of 300 K. To justify a time step of 1.5 ps, hydrogen masses were set to 10 amu. The effect of the solvent was modeled with GBMV [28]. In previous studies GBMV showed a very good agreement with explicit solvent results for several relative solvation free energies (RMSE = 0.5 kcal/mol [29]), therefore the expected error due to the implicit solvent model can be assumed to be small. Free energy differences were determined with BAR, using two steps: one for changing the charges and the other step for changing the atom types. For guest molecule **6** there was not sufficient phase space overlap to obtain converged results. Therefore, no data is shown for this guest. The simulation length at each endpoint was 15 ns, the first 1.5 ns of which were discarded as equilibration. Each simulation was carried out three times with different random seeds.

Simulations with different buffer concentrations

To determine the influence of the buffer concentration on the binding affinity result ($\Delta\Delta\Delta G_{bind}^{\Delta c_{buffer}}$), the additional NaCl pairs present in the binding affinity calculations (discussed in "Computation of relative binding affinities") were removed alchemically from all host-guest systems. Again, a thermodynamic cycle was employed, using free energy simulations between the systems that include the additional NaCl pair and systems where the electrostatic and Lennard-Jones interactions of that pair are turned off. This was done a.) in aqueous solution for the reference molecule ($\Delta G_{\mathrm{H_2O}}^{ref+\mathrm{NaCl}\rightarrow ref}$), as well as the target molecule ($\Delta G_{\mathrm{H_2O}}^{target+\mathrm{NaCl}\rightarrow target}$) and b.) the corresponding simulations while bound to the host molecule ($\Delta G_{host}^{ref+\mathrm{NaCl}\rightarrow ref}$, $\Delta G_{host}^{target+\mathrm{NaCl}\rightarrow target}$). This leads to

$$\Delta\Delta G_{\mathrm{H_2O}}^{\Delta c_{buffer}} = \Delta G_{\mathrm{H_2O}}^{target+\mathrm{NaCl}\rightarrow target} - \Delta G_{\mathrm{H_2O}}^{ref+\mathrm{NaCl}\rightarrow ref} \quad (9)$$

$$\Delta\Delta G_{host}^{\Delta c_{buffer}} = \Delta G_{host}^{target+\mathrm{NaCl}\rightarrow target} - \Delta G_{host}^{ref+\mathrm{NaCl}\rightarrow ref} \quad (10)$$

so that

$$\Delta\Delta\Delta G_{bind}^{\Delta c_{buffer}} = \Delta\Delta G_{host}^{\Delta c_{buffer}} - \Delta\Delta G_{\mathrm{H_2O}}^{\Delta c_{buffer}} \quad (11)$$

The corresponding free energy simulations were calculated using the same simulation setup as described in "Computation of relative binding affinities". The electrostatic and Lennard-Jones interactions of the Na^+ and Cl^- ions were turned off in 12 λ-steps ($\lambda = 0.0, 0.1, \ldots, 0.9, 0.95, 1.0$), using soft cores. Each λ-point was simulated for 2 ns. Free energy differences were calculated using TI and the trapezoidal rule.

Results and discussion

Table 1 lists the results for the absolute binding free energies (ΔG_{bind}) of the seven host-guest systems. Three free energy methods were employed: BAR (second column), TI-TR (third column) and TI-SI (fourth column).

Table 1 Computed absolute binding free energies ΔG_{bind} and their corresponding standard deviations for the seven guest molecules of the SAMPL3 challenge

Guest	ΔG_{bind}^{BAR}	ΔG_{bind}^{TI-TR}	ΔG_{bind}^{TI-SI}	Exp.
1	−9.1 ± 0.9	−8.5 ± 0.4	−7.5 ± 0.7	−5.8
2	−7.3 ± 0.5	−7.4 ± 0.6	−6.8 ± 0.6	−7.1
3	−5.3 ± 1.3	−3.8 ± 1.9	−2.0 ± 2.4	−6.8
4	−8.6 ± 0.7	−7.3 ± 1.3	−8.7 ± 1.4	−4.2
5	−9.3 ± 0.8	−9.0 ± 0.7	−6.6 ± 0.6	−6.1
6	−9.0 ± 1.0	−8.7 ± 1.2	−7.1 ± 1.0	−10.7
7	−6.4 ± 0.7	−5.0 ± 0.9	−4.5 ± 1.0	−7.9
RMSE	2.6	2.6	3.2	

Three different methods were used: Bennett's acceptance ratio method (BAR) with 12 λ-points and thermodynamic integration with the trapezoidal rule and 12 λ-points (TI-TR), and Simpson's rule with 11 λ-points (TI-SI). The experimental results (Exp.) of the binding free energies are listed in the last column. The root mean square errors (RMSE) of the computational results are included in the last row. All free energy differences are in kcal/mol

The ± sign represents the corresponding standard deviations, which were calculated from four replicas of each calculation. Generally, the standard deviations are very high, ranging between 0.5 and 1.3 kcal/mol for BAR, 0.4–1.9 kcal/mol for TI-TR and rising up to 0.6–2.4 kcal/mol for TI-SI. However, the average standard deviations were similar for all three methods (0.8 for BAR and TI-TR, 1.0 for TI-SI). Since the standard deviations reflect the quality of the sampling, this indicates that the binding free energy results are not well converged. Significantly longer trajectories would have been required to achieve what we consider adequate standard deviations of about 0.3 kcal/mol.

The experimental results for the binding free energies are presented in the rightmost column of Table 1. They form the basis for the root mean square errors (RMSE) presented in the last line for each computational method. The RMSE serve as a measure for the accuracy of each method. In terms of RMSE, the accuracies of BAR and TI-TR are equal (2.6 kcal/mol), while the errors of TI-SI are higher (3.2 kcal/mol). All three results fall into the same range of RMSE as experienced during past SAMPL competitions for the hydration free energies of organic molecules (between 1.3 [6] and 3.5 kcal/mol [7, 9]). This demonstrates that solvation free energies are a good benchmark system for free energy calculations of even larger molecular complexes.

When comparing the accuracy of the three methods, the relatively weak performance of TI-SI is interesting, since recent studies [20] demonstrated that TI, in connection with Simpson's rule (or other higher- order numerical integration schemes), is by far superior to the simple trapezoidal rule. However, better quadrature methods can enhance the efficiency of TI only if the shape of the integrand $\partial U/\partial\lambda$ is "well-behaved" and the values of $\partial U/\partial\lambda$ are converged. In Fig. 2 we show a typical $\partial U/\partial\lambda$ plot from our calculations. The four lines represent four different repetitions of the simulation. As can be seen, the aforementioned conditions are not met: Both the uncertainties of $\partial U/\partial\lambda$ (as illustrated by differences of the four repetitions) as well as the changes of $\partial U/\partial\lambda$ (a steep decrease between $\lambda = 0.0$ and $\lambda = 0.1$) are very high. In such cases, it is more efficient to introduce additional λ-points in the problematic regions of the $\partial U/\partial\lambda$ plot and to run longer simulations rather than employing high-order numerical quadrature. This is reflected by the difference of the RMSE of TI-TR and TI-SI (2.6 vs. 3.2 kcal/mol).

Figure 3 depicts the binding free energy results and standard deviations for TI-TR and BAR. The distribution of the results shows that both methods produce the same RMSE in two different ways. With the notable exception of **2**, most TI-TR predictions consistently deviate from the experimental results by about 2–3 kcal/mol (lying outside the orange lines in Fig. 3). Those deviations can probably be attributed to the problems of numerical quadrature as illustrated in Fig. 2. For BAR, the predictions can be divided into two groups: the outliers **1**, **4** and **5**, which exhibit a RMSE of 3.7 kcal/mol and the group of **2**, **3**, **6**, **7** with a relatively low RMSE of 1.4 kcal/mol. This difference to TI-TR is also highlighted by the absolute median error (1.7 kcal/mol for BAR and 2.9 kcal/mol for TI-TR). Since there are no consistent chemical patterns that distinguish the two groups in BAR, the errors are not likely to arise simply from imperfections of the force-field. The

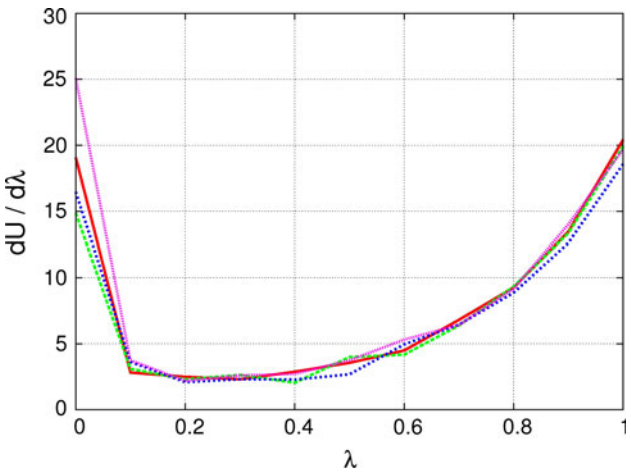

Fig. 2 Illustration of the difficulties encountered by TI: Plot of $\partial U/\partial\lambda$ as a function of λ for the free energy difference between *26* and **3** in complex with the host. The four different colors indicate the results of four different simulations. Between $\lambda = 0.0$ and $\lambda = 0.1$, the *curves* are very steep and the uncertainties of $\partial U/\partial\lambda$ are very high. This causes large errors in the numerical quadrature step of TI

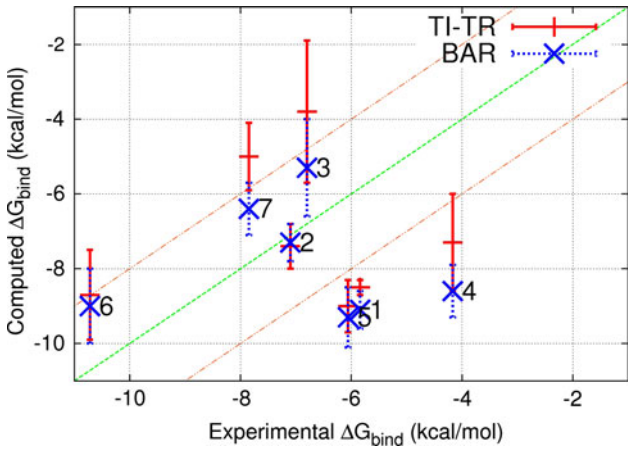

Fig. 3 Comparison of the BAR and TI-TR results for 12 λ-points. The experimental binding free energies of the guest molecules (*x*-axis) are plotted versus the computational results (*y*-axis). The TI-TR results and error bars are shown in *red*, while each BAR result and its corresponding *error bar* are marked by a *blue cross* and a *dashed line*. The line of ideal correspondence between experiments and predictions is shown in *green*, while the range within 2 kcal from the experimental results is indicated by two *orange lines*

cause of this effect is probably a mixture of error cancellation and insufficient sampling.

Dependency of the binding free energy on the pronation state

Another aspect of the accuracy of the binding free energy simulations is the selection of protonation states of both the guest molecules and the host. Since the protonation states in these simulations were picked based on similarities to molecules of known pK_a's in chemical textbooks, we were interested to see what would have happened if we had chosen the deprotonated state for our simulations. For that purpose, additional free energy calculations were conducted after the deadline for SAMPL3 to determine the change of binding affinity after deprotonation of the guest molecules ($\Delta\Delta\Delta G_{bind}^{deprot}$). The results for $\Delta\Delta\Delta G_{bind}^{deprot}$ (second column) as well as the corresponding absolute binding affinities (third column) are shown in Table 2. In addition, the standard deviations due to error propagation (after the $\pm$ sign) and the RMSE of the resulting absolute binding free energies (last row) are presented.

When taken as a whole, the binding affinity results of the deprotonated guest molecules deviate significantly from the experimental results, as indicated by the RMSE of 5.2 kcal/mol. Compared with the results of the protonated forms from Table 1, only guest molecule **5** lies closer to the experimental value of -6.1 kcal/mol than its corresponding protonated form (-7.7 kcal/mol instead of 9.3 kcal/mol). This indicates that the guest molecules are protonated in their bound form. Generally, the standard

Table 2 Effect of employing the deprotonated form of the guest molecules on the binding affinity ($\Delta\Delta\Delta G_{bind}^{deprot}$)

Guest	$\Delta\Delta\Delta G_{bind}^{deprot}$	ΔG_{bind}^{deprot}
1	0.0 ± 3.2	−9.1 ± 3.2
2	6.5 ± 1.2	−0.8 ± 1.3
3	3.8 ± 3.3	−1.5 ± 3.6
4	−2.5 ± 2.6	−11.1 ± 2.7
5	1.6 ± 2.7	−7.7 ± 2.8
7	0.6 ± 2.7	−5.8 ± 2.8
RMSE		5.2

The changes of the binding free energies due to deprotonation (and the associated standard deviations) are shown in the second column, the resulting absolute binding free energies of the deprotonated guests and the standard deviations due to error propagation are shown in the rightmost column. The corresponding RMSE for the deprotonated guests is presented in the last row. All free energy differences are in kcal/mol

deviations for $\Delta\Delta\Delta G_{bind}^{deprot}$ are very high, ranging between 1.2 and 3.3 kcal/mol. Therefore, most of the results are not statistically significant (i.e., indistinguishable from zero). This can be attributed to two effects: improper sampling and the detachment of the guest molecule from the host in some simulations. The latter problem could have been avoided by employing restraints to restrict their sampling to the binding pocket.

Dependency of the binding free energy on the buffer concentration

In their experimental paper, Ma et al. [12] presented data on the dependence of the binding affinity on the buffer concentration. They hypothesized that the cations in solution bind to the host molecule and thereby reduce the affinity toward the guests. This effect was demonstrated for one guest molecule by changing the sodium phosphate buffer concentration from 24.5 to 57.2 mM, which caused the binding affinity to change by about 0.4 kcal/mol. In our binding free energy simulations sodium chloride was employed instead of sodium phosphate, therefore, we had increased the buffer concentration relative to the experimental conditions in order to obtain about the same ionic strength (i.e. 40–50 mM instead of 20 mM). To check whether our binding affinity results were affected by the buffer concentration employed in our simulations, we conducted additional simulations after the deadline for the SAMPL3 competition. In those simulations, we lowered the buffer concentration to about 20–30 mM by alchemically removing one ion pair from our simulations. The changes of the binding free energy due to this change of buffer concentration ($\Delta\Delta\Delta G_{bind}^{\Delta c_{buffer}}$) are shown in Table 3.

Table 3 Effect of the buffer concentration on the binding affinity ($\Delta\Delta\Delta G_{bind}^{\Delta c_{buffer}}$) as a result from alchemically lowering the sodium chloride concentration in our simulations by approximately 10–20 mM

Guest	$\Delta\Delta\Delta G_{bind}^{\Delta c_{buffer}}$	$\Delta G_{bind}^{20\,\mathrm{mM}\,buffer}$
1	-0.7 ± 3.1	-9.8 ± 3.2
2	0.9 ± 1.7	-6.4 ± 1.8
3	2.4 ± 2.7	-2.9 ± 3.0
4	4.7 ± 2.2	-3.9 ± 2.3
5	-0.7 ± 1.6	-10.0 ± 1.8
6	1.1 ± 3.6	-7.9 ± 3.7
7	5.3 ± 3.0	-1.1 ± 3.1
RMSE		3.8

The resulting absolute binding free energies of the systems with a concentration of 20–30 mM and the standard deviations due to error propagation are shown in the rightmost column. The corresponding RMSE for the deprotonated guests is presented in the last row. All free energy differences are in kcal/mol

The $\Delta\Delta\Delta G_{bind}^{\Delta c_{buffer}}$ results fluctuate between -0.7 and $+5.4$ kcal/mol. This finding is counter-intuitive, given that the buffer competitively binds to the host molecule, and, therefore, $\Delta\Delta\Delta G_{bind}^{\Delta c_{buffer}}$ should be approximately the same for all binding free energy predictions. The high standard deviations of up to 3.6 kcal/mol indicate that the simulations are not converged. Thus, our free energy results strongly depend on the initial positions of the ions. Only the results for **4** and **7** can be considered statistically significant. The absolute binding affinity of **4** lies at -3.9 kcal/mol, which is significantly closer to the experimental values of -4.2 kcal/mol than in the original case. On the other hand the correspondence of **7** with experimental results is lowered. In total, the RMSE of all $\Delta G_{bind}^{20\,mM\,buffer}$ is higher than in our original predictions (3.8 vs. 2.6 kcal/mol), demonstrating that the use of the right ionic strength is more important than reproducing the buffer concentration.

Conclusions

The binding free energies of seven host-guest systems were predicted using relative free energy calculations. For the analysis of the trajectories, three different methods were employed: BAR using 12 λ-points, TI-TR with 12 λ-points, and TI-SI with 11 λ-points. While both BAR and TI-TR with 12 λ-points resulted in a RMSE of 2.6 kcal/mol, the corresponding TI-SI result with 11 λ-points yielded a RMSE of 3.2 kcal/mol. This difference can be traced back to the shape and the uncertainties of the $\partial U/\partial\lambda$ integrand. On the other hand, the performance of BAR and TI-TR can be differentiated in terms of the absolute median error (1.7 kcal/mol for BAR and 2.9 kcal/mol for TI-TR). Thus, BAR performs slightly better than TI when using the same set of data.

Overall, our results show that binding affinities of host-guest systems can be determined with about the same accuracy as solvation free energies of relatively small organic molecules in the previous SAMPL challenges (i.e., with a root mean square error of about 2–3 kcal/mol). This demonstrates that solvation free energies are indeed a valuable benchmark system. However, because of the six protonizable groups that are involved in the binding process, the host-guest systems in SAMPL3 are more challenging. This is reflected by the large errors of the simulations of the unprotonated state (RMSE of 5.2 kcal/mol). Our free energy simulations demonstrate that the free energy result is also very sensitive to the buffer concentration. Lowering the buffer concentration by about 20 mM is accompanied by an increase of the RMSE to 3.8 kcal/mol.

A striking feature of our results is the large variance. The standard deviations range from 0.4 to 2.4 kcal/mol, or, relative to the absolute results, between 5 and 120%. About 30% of the RMSE can be explained in terms those uncertainties. Considering that the simulation lengths were short (3 ns for each λ-point), these standard deviations mostly signify the uncertainties due to the sampling of a local energy minimum. This means that our free energy results are insufficiently converged to be used as an absolute measure of the quality of current force fields. Nevertheless, they demonstrate the errors that can be expected from relative free energy calculations with current computational resources, also accounting for the effects of protonation state and buffer concentration. They also indicate where more methodological development in the field of free energy simulations is required. For example, there is still room for improvement in the area of sampling and the treatment of groups with unknown protonation states. We, therefore, look forward to future free energy prediction challenges.

Acknowledgments The authors would like to thank R. Pastor, F. Pickard and A. Okur for helpful comments on the manuscript, as well as A. Damjanović and R. Venable for stimulating discussions on the potential protonation states of the host and guest molecules.

References

1. Oostenbrink C, van Gunsteren W (2005) PNAS 102(19):6750
2. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA (2007) J Mol Biol 371(4):1118
3. Kästner J, Senn H, Thiel S, Otte N, Thiel W (2006) J Chem Theory Comput 2(2):452
4. Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA (2009) J Chem Theory Comput 5(2):350
5. Seeliger D, de Groot B (2010) Biophys J 98(10):2309

6. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS (2008) J Med Chem 51:769
7. Guthrie JP (2009) J Phys Chem B 113(14):4501
8. Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ (2010) J Comput-Aided Mol Des 24(4, SI):259
9. Mobley DL, Bayly CI, Cooper MD, Dill KA (2009) J Phys Chem B 113(14):4533
10. Klimovich PV, Mobley DL (2010) J Comput-Aided Mol Des 24(4, Sp. Iss. SI):307
11. Moghaddam S, Yang C, Rekharsky M, Ko YH, Kim K, Inoue Y, Gilson MK (2011) J Am Chem Soc 133(10):3570
12. Ma D, Zavalij PY, Isaacs L (2010) J Org Chem 75(14):4786
13. Merz KM (2010) J Chem Theory Comput 6:1018
14. Bennett CH (1976) J Comp Phys 22:245
15. Kirkwood JG (1935) J Chem Phys 3:300
16. Bruckner S, Boresch S (2011) J Comp Chem 32(7):1303
17. Brooks B, Brooks C III, Mackerell A Jr, Nilsson L, Petrella R, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner A, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor R, Post C, Pu J, Schaefer M, Tidor B, Venable R, Woodcock H, Wu X, Yang W, York D, Karplus M (2009) J Comp Chem 30(10, Sp. Iss. SI):1545
18. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187
19. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD Jr (2010) J Comp Chem 31(4):671
20. Bruckner S, Boresch S (2011) J Comp Chem 32(7):1320
21. Shirts MR, Bair E, Hooker G, Pande VS (2003) Phys Rev Lett 91:140601
22. Jorgensen WL, Chandrasekhar H, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926
23. Neria E, Fischer S, Karplus M (1996) J Chem Phys 105:1902
24. Lamoureux G, Roux B (2003) J Chem Phys 119(6):3025
25. Hoover WG (1985) Phys Rev A 31:1695
26. Van Gunsteren WF, Berendsen HJC (1977) Mol Phys 34:1311
27. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) J Chem Phys 103:8577
28. Lee MS, Feig M, Salsbury FR, Brooks CL III (2003) J Comput Chem 23:1348
29. König G, Boresch S (2009) J Phys Chem B 113(26):8967