J Comput Aided Mol Des (2014) 28:187–200
DOI 10.1007/s10822-014-9717-3

Extensive all-atom Monte Carlo sampling and QM/MM corrections in the SAMPL4 hydration free energy challenge

Samuel Genheden · Ana I. Cabedo Martinez · Michael P. Criddle · Jonathan W. Essex

Received: 12 November 2013 / Accepted: 21 January 2014 / Published online: 1 February 2014

Abstract We present our predictions for the SAMPL4 hydration free energy challenge. Extensive all-atom Monte Carlo simulations were employed to sample the compounds in explicit solvent. While the focus of our study was to demonstrate well-converged and reproducible free energies, we attempted to address the deficiencies in the general Amber force field force field with a simple QM/MM correction. We show that by using multiple independent simulations, including different starting configurations, and enhanced sampling with parallel tempering, we can obtain well converged hydration free energies. Additional analysis using dihedral angle distributions, torsion-root mean square deviation plots and thermodynamic cycles support this assertion. We obtain a mean absolute deviation of 1.7 kcal mol^{-1} and a Kendall's τ of 0.65 compared with experiment.

Keywords SAMPL4 · Hydration free energy · Monte Carlo simulation · QM/MM corrections · SMD

Introduction

The hydration free energy, i.e., the transfer free energy of a molecule from vacuum to water, is a central thermodynamic quantity for aqueous systems. Hydration free energies are a major determinant in phenomena such as the binding of small molecules to macromolecules and protein folding [1]. In the context of modeling, the hydration free energy is important in methods to compute binding free energies, as some of these methods are expressed as the difference between a free energy calculation in the receptor and in the solvent [2]. Hence, it is important to be able to accurately determine this quantity.

Computational methods to predict hydration free energies range from continuum models where the solvent is treated as a dielectric continuum to all-atom sampling schemes where the atoms are described by a molecular mechanics force field [3–5]. As such, all-atom sampling predictions of hydration free energy are good tests of the underlying force field. However, it should be noted that in the case of binding free energy calculations where one evaluates a difference between results derived from two independent simulations of the free and bound ligand, one would expect a fair amount of cancellation of the deficiencies in the force field [6].

Herein, we report predictions of the hydration free energy for the 52 molecules in the SAMPL4 challenge [7, 8]. This is the fourth community-wide blind challenge of hydration free energies [9–11]. The blindness of the test allows for a fair benchmarking of the available methodology without introducing bias towards the "correct" result. We present predictions from all-atom Monte Carlo simulations using molecular mechanics, together with a comparatively inexpensive QM/MM polarization correction. As a final benchmark, we report free energies calculated using a very efficient continuum model [12].

Electronic supplementary material The online version of this article (doi:10.1007/s10822-014-9717-3) contains supplementary material, which is available to authorized users.

S. Genheden (✉) · A. I. Cabedo Martinez · M. P. Criddle · J. W. Essex
School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK
e-mail: s.genheden@soton.ac.uk

Method

Compound preparation

Initial three-dimensional structures were created from the provided SMILES strings [8] using UCSF Chimera [13]. These structures were then optimized at the semi-empirical AM1 Hamiltonian level [14] using Gaussian09 [15]. Thereafter, general Amber force field (GAFF) atom types [16] and AM1-BCC charges [17] were assigned using antechamber in Amber12 [18]. As we did not sample any bonds and only some of the angles in the solute molecules, the compounds were further minimized with 100 steps of steepest descent using the sander module in Amber12 [18].

Additional conformations for some compounds were generated. For compounds **1–4**, **25–31**, 10 random conformations were generated using the stochastic sampling algorithm in MOE [19]. For compounds **6–11**, **33–35** and **37**, two conformations were generated; one where the hydroxyl group of the compound made an intramolecular hydrogen bond and one where the hydroxyl group pointed out into solution. For compound **22**, three initial conformations were manually created as discussed in the text. For compounds **23** and **24**, six initial conformations were created either manually or with the stochastic sampling algorithm using MOE. All conformations are available as supplementary material in the form of PDB files.

Each compound was immersed in a cubic box of TIP4P water molecules [20] created by extending a pre-equilibrated box such that it extended a minimum of 10 Å from the solute in each direction and removing water molecules that overlapped with the solute. On average 1,000 water molecules were added.

Free energy simulations

The free energy of hydration (ΔG_{hyd}) was calculated by decoupling the compound from the surrounding water molecules. The decoupling free energy (ΔG_{dec}) was estimated using finite difference thermodynamic integration (FDTI) [21, 22]

$$-\Delta G_{\text{hyd}} = \Delta G_{\text{dec}} = \int_0^1 \left\langle \frac{\partial G}{\partial \lambda} \right\rangle d\lambda$$
$$\frac{\partial G}{\partial \lambda} = \frac{-\text{RT}\left[\ln\langle \exp(-(U_{\lambda+h} - U_\lambda)/RT\rangle_\lambda - \ln\langle \exp(-(U_{\lambda-h} - U_\lambda)/RT\rangle_\lambda\right]}{2h} \quad (1)$$

where R and T is the gas-constant and absolute temperature, respectively. $U_\lambda = (1 - \lambda)U_0 + \lambda U_1$ is the combination of the energy of the initial state (U_0) where the compound is interacting with the surrounding water and the energy of the final state (U_1) where the electrostatics and van der Waals interactions between the compound and all the water molecules have been turned off [5, 23]. λ is a coupling parameter and the brackets indicate an ensemble average at a fixed value of the coupling parameter. Unless otherwise stated, we used 24 values of λ (0.00, 0.03, 0.06, 0.09, 0.12, 0.19, 0.26, 0.33, 0.40, 0.47, 0.54, 0.61, 0.68, 0.72, 0.76, 0.80, 0.84, 0.88, 0.90, 0.92, 0.94, 0.96, 0.98, and 1.0). The free energy gradient was estimated by a second-order finite difference approximation using exponential averaging [24] and $h = 0.0001$. At the end-points ($\lambda = 0$ and $\lambda = 1$), a first order finite-difference approximation was used instead. The integration was calculated using the trapezium rule. Note that we did not perturb any intramolecular interactions in the compounds; only interaction energies were coupled to λ, and as such we do not have to perform additional gas-phase calculations [25].

To improve sampling we use replica-exchange TI (RETI), i.e., simulations where neighboring values of λ were subjected to a Metropolis test and if the test was successful the configurations of the two simulations were swapped [22, 26]. For some compounds we also added a temperature ladder (from 300 to 400 K in steps of 10 K) at the end-points ($\lambda = 0$ and $\lambda = 1$) and performed parallel tempering (PT) [27] moves concurrently with the RETI swaps.

The electrostatics and van der Waals interactions were decoupled simultaneously using a dual-topology approach [28] and soft-core potentials [29, 30]:

$$U_{\text{Coulomb}} = \lambda \frac{q_i q_j}{4\pi\varepsilon_0} \frac{1}{\left(r^6 + 2(1-\lambda)\right)^{1/6}}$$
$$U_{\text{LJ}} = 4\varepsilon\lambda \left[\frac{\sigma^{12}}{\left(r^6 + 0.2(1-\lambda)\sigma^6\right)^2} - \frac{\sigma^6}{r^6 + 0.2(1-\lambda)\sigma^6} \right] \quad (2)$$

where q, ε, and σ are the standard force field charges and van der Waals parameters, r is the distance between two atoms, and ε_0 is the vacuum permitivity.

All simulations were performed with a modified version of the Monte Carlo program ProtoMS2.3 [31].

Modifications were only made to the soft-core potentials available. The temperature was kept at 300 K and the pressure at 1 atm using standard Metropolis tests. The non-bonded interactions were truncated at 10 Å using a molecule-based cutoff. Translation and rotation of a water molecule was attempted with 99 % probability, a solute move with 0.9 % and a volume move with 0.1 %. A solute move included sampling of all torsions, excluding aromatic rings, and sampling of all angles, excluding aromatic rings and angles involving non-polar hydrogen atoms. The displacement factors were set to achieve an approximate acceptance ratio of 40 %. Five million (m) Monte Carlo moves were performed at each value of λ before any replica-exchange tests were attempted. Thereafter, 50 m Monte Carlo moves were performed and the first 20 m were discarded as equilibration. RETI moves were attempted every 200 thousand (k) moves and PT moves were attempted every 100 k moves. Snapshots and energies were collected every 100 k moves. The energies used to compute the free energy gradient were averaged over 100 k MC moves. Between 5 and 20 independent sets of simulations were initiated for each compound.

QM/MM corrections

Following a recently suggested approach [32], we performed QM/MM calculations at the initial state ($\lambda = 0$) to correct for the lack of polarization in the force field. From each simulation we extracted 300 snapshots and performed QM/MM energy calculations: the solute was treated at the B3LYP/6-31G* level of theory [33, 34] whereas the water molecules within 10 Å of the solute were represented as point charges. All the other water molecules were omitted from the calculations. The QM/MM energy was calculated using Gaussian09 [15]. The QM/MM correction (ΔG_{corr}) was subsequently computed using the Zwanzig equation (exponential averaging) [24]

$$\Delta G_{\text{corr}} = -RT\ln\left\langle \exp\left(-\frac{U_{\text{QM/MM}} - U_{\text{MM}}}{RT}\right)\right\rangle_{\text{MM}} \quad (3)$$

where R and T are the gas constant and simulation temperature respectively, and the angle brackets indicate an ensemble average over snapshots generated using the molecular mechanics force field. U_{MM} is the electrostatic energy between the solute and the water molecules. $U_{\text{QM/MM}}$ is the difference between the energy of the QM/MM calculation as described above and the gas-phase QM energy of the solute and the self-energy of the charges [32]. Finally, the QM/MM free energy can then be written as the sum of the MM free energy and the correction free energy. Using this formalism, we are incorporating polarization by correcting for the electrostatic interaction energy between the solute and the water molecules.

Continuum solvent predictions

We performed SMD calculations [12] on the snapshots generated by UCSF Chimera [13] from the SMILES strings. The calculations were made at the B3LYP/6-31G* level of theory using Gaussian09 [15].

Quality metrics and bootstrapping

The quality of the predictions compared to experiments was measured using mean absolute deviation (MAD), the correlation coefficient (r^2) and Kendall's τ [35]. As both predictions and experimental data have uncertainties, so have these quality metrics. The uncertainty of the quality metrics was computed using a parametric bootstrap procedure [36, 37]. Both the predictions and experiments were assumed to be normally distributed, centered on the hydration free energy and with standard deviation equal to the simulation standard error in the case of the predictions, or equal to the reported uncertainty [7] in the case of the experimental data. 1,000 bootstrap samples were used.

Results

We computed hydration free energies for 52 compounds in the SAMPL4 challenge using extensive all-atom Monte Carlo simulations. The compounds were divided by the organizers into a blind set and a supplementary set [8]. We performed extensive tests of sampling convergence for the compounds in the blind set as discussed below. These led to a number of broadly defined heuristics that we then used to predict hydration free energies for the supplementary set. Hence, we have chosen to divide this section based on these two sets. After the MM results, we will present the predictions from a simple QM/MM polarization correction for the solutes. Finally, we will present the predictions of a simple continuum approach.

Blind set

We divided the blind set further to focus our effort on more challenging compounds. We considered compounds **5** and **12–21** to be fairly straightforward in terms of sampling and convergence, because they are fairly small, rigid and cannot form intramolecular hydrogen bonds. Hence, a standard protocol was used for these compounds: we started RETI simulations from a single conformation of each compounds, using 24 values of λ and 50 M moves at each λ. Initial test simulations showed that this was appropriate to obtain converged RETI estimates and a smooth curve to

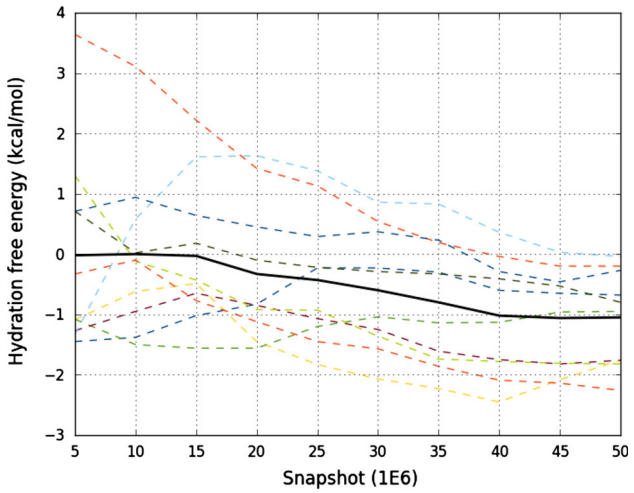

Fig. 1 Convergence of the hydration free energy of compound **13**. The convergence of the free energy is illustrated as a function of the number of snapshots employed at each value of λ. The ten individual repeats are plotted as *dashed lines* whereas the average over the ten repeats are shown as a *thick, continuous line*

integrate. To obtain high precision, we employed ten independent sets of simulations. The reported uncertainties are the standard deviation of the average of these independent simulations. A typical convergence plot is shown in Fig. 1, which illustrates that the average of the independent sets is well converged. The hydration free energies are shown in Table 1. It is clear that the RETI estimates are consistently more positive than the experimental values. The deviations are generally greater than 2 kcal mol^{-1}, with compound **17** as the sole exception.

Next, we considered compounds **1**–**4**, which have long carbon-chains and therefore many possible conformations. To address this potential sampling issue, we generated ten random conformations for each compound and started RETI simulations from each of these conformations. The results for compound **1** are shown in Table 2. Interestingly, we observe free energies ranging from −13 to −21 kcal mol^{-1}, giving an average −17 kcal mol^{-1}. This is a large range and an indication of convergence problems. To investigate further we chose to study torsion RMSD (root mean square deviation) plots of the simulations. The torsion RMSD between two snapshots i and j from a simulation was defined as

$$\mathrm{RMSD}_{ij} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\cos\theta_{ik} - \cos\theta_{jk})} \qquad (4)$$

where θ_{ik} is the angle for dihedral k of the compound at snapshot i (there are N torsions in total) and the sum is over all sampled dihedral angles. By plotting this RMSD for all pairs of snapshot, an indication of sampling efficiency is achieved. A typical example for a RETI simulation of compound **1** is shown in Fig. 2. The plot displays large areas of low RMSD, which shows that the simulation becomes trapped in local conformations and that only a small amount of global sampling is achieved. One solution is to run even longer simulations and hope to see more transitions between the conformations. However, here we chose to employ PT in conjunction with the RETI simulations. We added a temperature ladder at each of the endpoints to increase the sampling at these states. Naturally, it is possible to also add ladders to the other values of λ, but such an approach soon becomes very computationally expensive. Our RETI + PT simulations require 64 cores per compound. The results of the RETI + PT simulations are collected in Table 2. We obtain a narrower range of free energies, from −11 to −16 kcal mol^{-1}, with an average of −14 kcal mol^{-1}. A typical RMSD plot for the RETI + PT simulations is shown in Fig. 2, and it is clear that there is a significant improvement in sampling and less conformational trapping. To increase precision, we ran a second repeat for each of the ten starting conformations, and averaged to obtain the final result in Table 1. The prediction of −15.2 kcal mol^{-1} is 8 kcal mol^{-1} too positive, compared to the experimental data. For the compounds **2**–**4**, we also made ten random, initial conformations, and the individual results are shown in Table 2. In contrast to **1**, the RMSD plots indicate that the RETI simulations are sufficiently converged (see Fig. 3). Therefore, we ran a second repeat of each conformation without employing PT to obtain the final predictions in Table 1. The predictions are between 2 and 5 kcal mol^{-1} more positive than the experimental data.

The next group we considered contained compounds **6**–**11**, as these have the possibility to make both intra- and intermolecular hydrogen bonds. This is a natural test of sampling convergence: if we start from either configuration of the hydroxyl group, we should obtain identical estimates of the hydration free energy. Consequently, we started RETI simulations from the two different conformations of each compound, but otherwise used the same protocol as for the simpler compounds. To obtain adequate precision, we employed three independent sets of simulations for each starting conformation. The results are collected in Table 3. For compounds **6**, **9** and **10**, the difference between the two starting conformations is smaller than 1 kcal mol^{-1} and not statistically significant at the 95 % confidence level. However for the other compounds, we see much larger conformation-dependent differences of between 2 and 3 kcal mol^{-1}. To investigate this, we plotted the distribution of the torsions that describes the rotation around the hydroxyl and methoxy groups. As a typical example we have, in Fig. 4, plotted the hydroxyl group torsion from a simulation of compound **11**. The simulations starting from the intermolecular conformation readily

Table 1 Prediction of the hydration free energy by various methods

Compound	MM	QM/MM	SMD	Exp
Blind set				
1	−15.2 ± 0.2	−22.7 ± 0.3	−22.0	−23.6
2	−0.4 ± 0.2	−1.0 ± 0.2	−1.1	−2.5
3	0.5 ± 0.2	−3.1 ± 0.2	−4.1	−4.8
4	−0.1 ± 0.2	−3.7 ± 0.2	−4.2	−4.5
5	−2.6 ± 0.2	−3.4 ± 0.3	−1.8	−5.3
6	−2.2 ± 0.2	−3.6 ± 0.3	−3.5	−5.3
7	−2.0 ± 0.2	−4.6 ± 0.3	−4.4	
8	−4.0 ± 0.2	−5.4 ± 0.3	−5.9	
9	−7.0 ± 0.3	−10.0 ± 0.4	−6.7	−8.2
10	−4.3 ± 0.3	−6.5 ± 0.3	−4.2	−6.2
11	−7.7 ± 0.2	−10.1 ± 0.3	−6.1	−7.8
12	−0.9 ± 0.3	−3.1 ± 0.3	−3.1	−3.8
13	−1.8 ± 0.3	−5.0 ± 0.3	−3.7	−4.4
14	−0.8 ± 0.3	−2.6 ± 0.3	−2.6	−4.1
15	−1.8 ± 0.3	−3.2 ± 0.3	−2.9	−4.5
16	−0.7 ± 0.3	−3.8 ± 0.4	−2.7	−3.2
17	−2.3 ± 0.3	−4.7 ± 0.3	−2.5	−2.5
18	−11.3 ± 0.3	−14.4 ± 0.4	−12.8	
19	−1.1 ± 0.3	−0.3 ± 0.3	−2.2	−3.8
20	−0.5 ± 0.3	0.2 ± 0.3	−0.9	−2.8
21	−5.9 ± 0.2	−8.9 ± 0.3	−8.4	−7.6
22	−9.5 ± 0.4	−7.3 ± 0.5	−5.6	−6.8
23	−6.9 ± 0.3	−5.5 ± 0.5	−6.3	−9.3
24	−3.0 ± 0.4	−2.2 ± 0.5	−4.4	−7.4
r^2	0.76 ± 0.02	0.83 ± 0.02	0.94 ± 0.01	
τ	0.63 ± 0.04	0.53 ± 0.04	0.69 ± 0.04	
MAD	3.0 ± 0.1	1.8 ± 0.1	1.5 ± 0.1	
Supplementary set				
25	−2.4 ± 0.2	−6.0 ± 0.2	−5.6	−5.7
26	−3.5 ± 0.2	−6.6 ± 0.3	−8.0	−5.3
27	−4.1 ± 0.2	−7.0 ± 0.2	−1.9	−4.8
28	−3.6 ± 0.3	−6.6 ± 0.3	−1.3	−4.3
29	0.6 ± 0.2	−1.2 ± 0.2	0.8	−1.7
30	−0.4 ± 0.2	−1.8 ± 0.2	−2.3	−2.3
31	−2.3 ± 0.2	−6.1 ± 0.3	−9.6	
32	−2.6 ± 0.2	−5.3 ± 0.3	−5.9	−7.3
33	−4.9 ± 0.3	−7.2 ± 0.4	−6.2	−6.9
34	−1.9 ± 0.2	−4.1 ± 0.3	−3.8	−5.8
35	−6.4 ± 0.3	−7.0 ± 0.7	−8.3	−4.7
36	−2.1 ± 0.3	−3.1 ± 0.3	−4.6	−5.7
37	−4.3 ± 0.2	−5.6 ± 0.5	−5.3	−5.9
38	−2.9 ± 0.3	−3.4 ± 0.4	−3.6	−3.9
39	1.0 ± 0.3	1.1 ± 0.3	0.1	−0.9
40	−0.7 ± 0.2	−3.6 ± 0.3	−4.5	
41	−2.4 ± 0.2	−5.5 ± 0.2	−3.9	−5.1
42	−0.5 ± 0.2	−3.4 ± 0.3	−2.0	−3.1
43	3.1 ± 0.2	1.9 ± 0.2	0.7	0.1
44	−3.3 ± 0.2	−7.2 ± 0.2	−5.3	−5.1

Table 1 continued

Compound	MM	QM/MM	SMD	Exp
45	−12.1 ± 0.4	−13.9 ± 0.5	−10.8	−11.5
46	−10.1 ± 0.5	−11.9 ± 0.6	−7.6	−9.4
47	−9.5 ± 0.5	−11.5 ± 0.6	−12.5	−14.2
48	−10.3 ± 0.4	−10.9 ± 0.4	−9.2	−11.9
49	−1.4 ± 0.4	−1.5 ± 0.4	−1.4	−3.2
50	−3.1 ± 0.4	−1.3 ± 0.4	−2.1	−4.1
51	−8.1 ± 0.5	−11.1 ± 0.8	−15.2	−9.5
52	1.1 ± 0.3	1.0 ± 0.4	−1.8	−2.9
r^2	0.83 ± 0.03	0.80 ± 0.03	0.76 ± 0.01	
τ	0.65 ± 0.02	0.65 ± 0.02	0.70 ± 0.02	
MAD	2.2 ± 0.1	1.6 ± 0.1	1.6 ± 0.1	
All compounds				
r^2	0.77 ± 0.02	0.81 ± 0.02	0.84 ± 0.01	
τ	0.69 ± 0.02	0.64 ± 0.02	0.72 ± 0.01	
MAD	2.6 ± 0.1	1.7 ± 0.1	1.6 ± 0.1	

The uncertainties of the predictions are standard errors on the average of several independent sets of simulation; the uncertainties for the performance metrics were computed by bootstrapping using both the experimental and prediction uncertainty. The performance metrics are the correlation coefficient (r^2), Kendall's rank correlation coefficient (τ) and the mean unsigned deviation (MAD)

Table 2 Hydration free energies for various compounds starting from different conformations in kcal mol^{-1}

Compound	**1**		**2**	**3**	**4**	**23**		**24**	
Conformation	RETI	RETI + PT	RETI	RETI	RETI	RETI	RETI + PT	RETI	RETI + PT
1	−16.2	−10.7	0.9	−0.7	−1.0	−7.9	−6.5	−4.3	−4.7
2	−12.7	−13.3	−2.5	0.5	2.1	−7.1	−5.2	−4.8	−3.7
3	−16.5	−14.4	0.6	−0.7	1.0	−7.7	−5.9	−4.5	−2.7
4	−18.1	−12.4	−0.3	1.0	0.3	−5.7	−8.1	−3.5	−0.9
5	−21.0	−14.7	0.8	2.5	−0.8	−4.5	−6.7	−3.0	−2.3
6	−18.0	−14.2	−1.3	1.2	2.9	−8.4	−7.6	−2.3	−0.9
7	−14.8	−12.8	2.9	0.7	−0.7				
8	−15.0	−14.0	−1.2	−0.2	−1.5				
9	−16.6	−15.5	−1.4	1.0	1.6				
10	−18.6	−15.2	−1.9	−3.2	0.5				

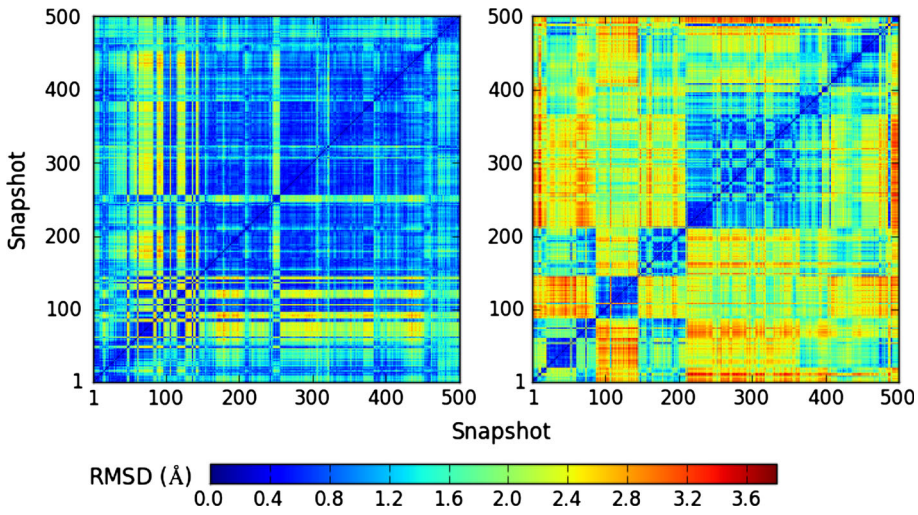

Fig. 2 Torsion RMSD between pairs of snapshots in a RETI (*left*) and a RETI + PT (*right*) simulation of compound **1** at $\lambda = 0$

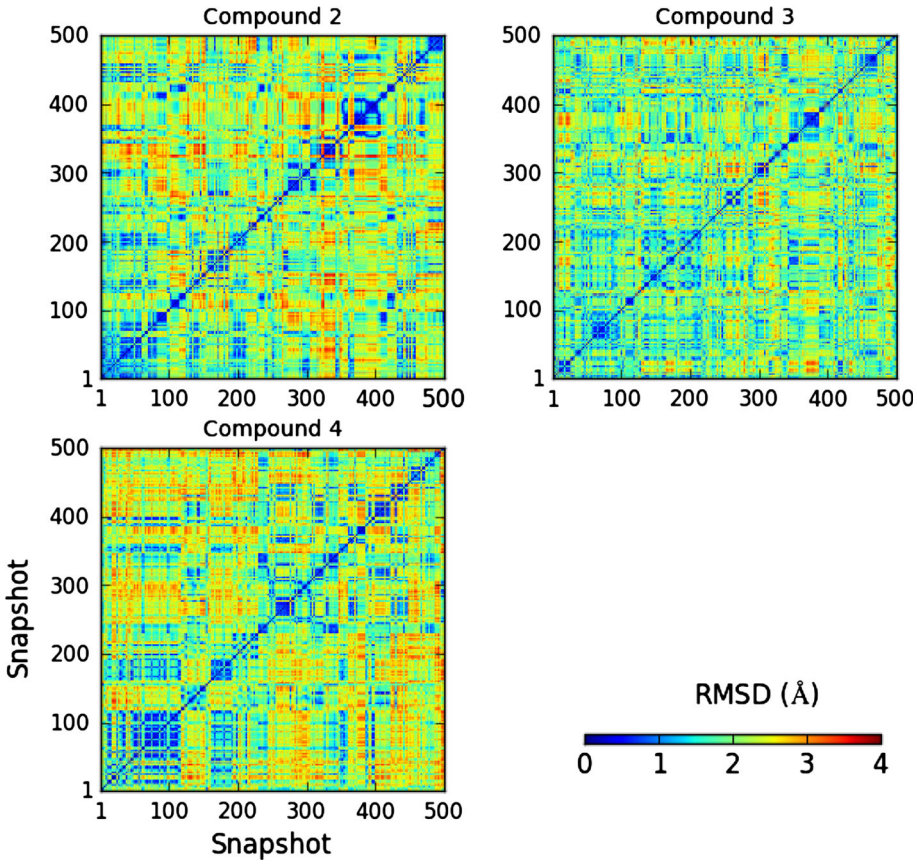

Fig. 3 Torsion RMSD between pairs of snapshots in RETI simulations of compounds **2, 3** and **4** at $\lambda = 0$

Table 3 Hydration free energies for compounds **6–11** starting from different conformations in kcal mol^{-1}

Compound	RETI		RETI + PT	
	Intramolecular	Intermolecular	Intramolecular	Intermolecular
6	−1.8 ± 0.5	−2.6 ± 0.5	−2.0 ± 0.5	−2.9 ± 0.4
7	−0.8 ± 0.3	−2.6 ± 0.2	−1.2 ± 0.4	−2.5 ± 0.4
8	−2.2 ± 0.3	−4.0 ± 0.6	−4.3 ± 0.3	−4.3 ± 0.4
9	−7.1 ± 0.8	−7.1 ± 0.7	−8.1 ± 0.5	−8.7 ± 0.4
10	−4.2 ± 0.6	−4.4 ± 0.4	−4.6 ± 0.4	−4.2 ± 0.3
11	−6.4 ± 0.5	−9.4 ± 0.6	−7.4 ± 0.4	−7.5 ± 0.5

Intramolecular means that in the starting conformation there is hydrogen bond between the hydroxyl group and the methoxy group of the compound, intermolecular means that the hydroxyl group forms a hydrogen bond with the solvent

sample two conformations during the simulations, whereas the simulations starting from the intramolecular conformation remain trapped in the starting conformation. For compounds **7** and **8** this means that for simulations starting with an intermolecular hydrogen bond, both intermolecular and intramolecular hydrogen bonds will be sampled during the simulations, whereas simulations starting from an intramolecular hydrogen bond retain this conformation during the simulations. For compound **11**, however, the hydroxyl group only interacts with one of the methoxy groups if the simulation is starting from an intramolecular conformation. For compounds **9–11**, the hydroxyl group rarely interacts with the surrounding water molecules. This analysis shows incomplete sampling of an important dihedral. Therefore, we decided to run RETI + PT simulations. The free energy results are shown in Table 3, and it is clear that for all the compounds the difference between simulations run with the two different starting conformations is equal to or less than 1 kcal mol^{-1}, with the exception of compound **7**. For this compound we observe a statistically significant difference of 1.3 kcal mol^{-1} at the 95 % confidence level. When we plot the distribution of the

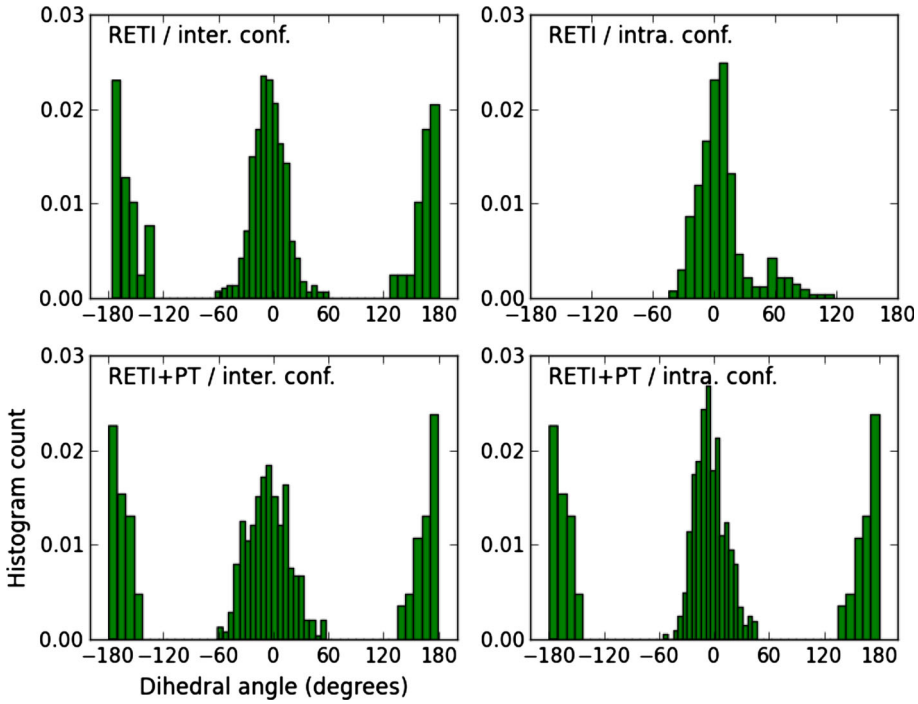

Fig. 4 Histograms of the hydroxyl group torsion in compound **11** using four simulation protocols ($\lambda = 0$). The label in the *upper-left* corner indicates if the simulation was a RETI or RETI + PT simulation and from which conformation of the hydroxyl group the simulation was initiated. The histograms have been normalized to unity

hydroxyl group torsion (see for instance Fig. 4), we see that in many simulations starting from the intramolecular conformation, the RETI + PT approach allows the hydroxyl group to visit two conformations, i.e., we have succeeded in increasing the sampling. With the exception of compound **7**, it is clear that using RETI + PT yields more converged results than when using RETI alone. To improve the precision, further simulations were run resulting in a total of twelve independent repeats. However, to reduce computational cost, we chose to use only RETI for compounds **6**, **9**, and **10**. Interestingly, for compound **7**, we observe a free energy of hydration of 2.2 and 1.9 kcal mol^{-1} starting from the intramolecular and intermolecular hydrogen bond, respectively. Hence, with more repeats, we obtain converged results for all compounds. The final predictions for compounds **6**–**11** are shown in Table 1. Unfortunately, the experimental data for compounds **7** and **8** were judged to be too uncertain and these compounds were therefore removed from the set. The predictions for the other compounds are consistently more positive than the experimental data, by between 0.1 and 3 kcal mol^{-1}.

For compound **22**, we created three initial conformations manually. In conformation 1, the hydroxyl group of the carboxylic acid was hydrogen bonded to the amino-group. In conformation 2, the carbonyl group made that interaction, and in conformation 3, we rotated the entire ring containing the acid-group such that neither the hydroxyl group nor the carbonyl group could interact with the amino-group. All the conformations are sketched in 2D in Fig. 5 and are available in 3D as supplementary material. We then started RETI simulations from each of these conformations, and obtained the results in Table 4. Conformations 1 and 2 give similar free energies whereas

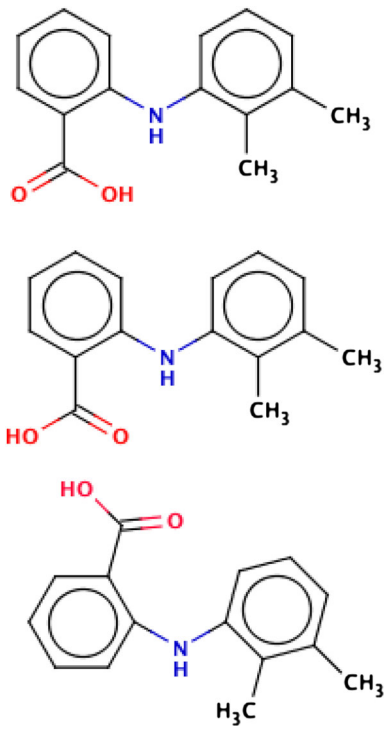

Fig. 5 Simulated conformations of compounds 22. The groups that can form intramolecular hydrogen bonds are marked in *color* and *bold face*. The conformations 1–3 are shown *top* to *bottom*

Table 4 Free energies for compound **22** in kcal mol^{-1}

	RETI	RETI + PT
Hydration free energies		
Conformation 1	−5.4	−2.7
Conformation 2	−4.1	−1.1
Conformation 3	−7.4	−9.2
Relative free energies		
1 → 2	1.1	
1 → 3	−5.0	
2 → 3	−5.5	

conformation 3 gives values that are 2–3 kcal mol^{-1} more negative. By looking at dihedral distributions, it is clear that the initial conformation was retained during the entire simulation (data not shown). It is unsurprising that by starting from conformation 3 we never sample conformation 1 and 2, or vice verse. Such sampling would require a large move with significant solvent reorganisation, a global move for which Monte Carlo is particularly unsuitable. We therefore attempted to run RETI + PT simulations, but the free energies shown in Table 4 and the dihedral distribution showed that this did not resolve the sampling problems. Indeed, the free energy difference between conformations 1 and 2 and conformation 3 actually increased in the RETI + PT simulations. We then performed RETI simulations to compute the free energy difference between conformations to check internal consistency. These results are also shown in Table 4. Forming the thermodynamic cycle for transformations 1 → 2 → 3 → 1, we obtained a sum of free energies as 0.7 ± 0.8 kcal mol^{-1} indicating that these simulations are converged (in the event of perfect convergence, we would obtain a sum of 0 kcal mol^{-1}). Taking these results into consideration, we decided to only use conformation 3, simply because this free energy estimate would dominate a Boltzmann average of the three estimates. Thus ten repeats of the RETI simulation of conformation 3 were run to obtain the result in Table 1. The prediction is 3 kcal mol^{-1} more negative than the experimental value.

Finally, for compounds **23** and **24**, we created six initial conformations, both manually and by the stochastic algorithm in MOE. The results from the RETI simulations for all six conformations are shown in Table 2. The range of the predictions is rather narrow; 4 and 3 kcal mol^{-1} for **23** and **24**, respectively. We also attempted RETI + PT simulations giving results that were very similar to those of the RETI simulations. We therefore consider the RETI simulations to be converged. However, to improve the precision, we ran a second repeat of each conformation giving a total of 12 values, averaged to obtain the predictions in Table 1. The predictions for both **23** and **24** are more positive than the experimental values.

Comparing all the predictions of the blind set with the experiments we obtain a Kendall's τ of 0.65, an r^2 of 0.77 and a MAD of 2.8 kcal mol^{-1}. Thus the molecular mechanics predictions are able to rank the compounds rather accurately, although the absolute estimates deviate from experiment, and are in general too positive.

Supplementary set

The simulation protocol for the supplementary set was based on the knowledge acquired from the blind set. Therefore, for compounds **25–31**, we ran RETI + PT simulations on ten random starting conformations as these compounds contain long carbon-chains. For compounds **33–35** and **37**, we ran three repeats of RETI + PT simulations of a conformation where the hydroxyl group made an intramolecular hydrogen bond and three repeats of a conformation where the hydroxyl group made an intermolecular hydrogen bond. For the remaining compounds we ran five repeats of RETI simulations on a single starting conformation. All the results are collected in Table 1.

Most of the predictions are more positive than the experimental values, by between 0.5 and 5 kcal mol^{-1}. The predictions for compounds **45** and **46** and are slightly negative, but the differences are not significant. For compound **35**, we obtain a prediction that is 2 kcal mol^{-1} more negative than the experimental value. Compared to experiments we obtain a τ of 0.65, an r^2 of 0.83 and a MAD of 2.2 kcal mol^{-1}. This is slightly better than for the blind set. Looking at all 47 compounds with experimental data that were judged to be reliable, we obtain a τ of 0.70, an r^2 of 0.78 and a MAD of 2.5 kcal mol^{-1}.

QM/MM corrections

To correct for deficiencies in the force field, we computed a simple QM/MM correction by performing QM/MM calculations on the snapshots generated at the initial state ($\lambda = 0$). The results for the QM/MM predictions are shown in Table 1. The correction amounts to between −7 and 2 kcal mol^{-1}, and −2 kcal mol^{-1} on average. For four of the compounds (**35**, **38**, **49** and **52**) the correction is not significantly different from zero at 95 % confidence. For seven compounds the correction was positive and for the reminder it was negative. Comparing the QM/MM corrected free energies to experiments for both data sets, we obtain a τ of 0.65, an r^2 of 0.81 and a MAD of 1.7 kcal mol^{-1}. Although the ranking is very similar to that obtained with MM, QM/MM gives better absolute predictions.

Predictions with a continuum method

A final set of predictions was made with the SMD model that is a combination of IEFPCM (integral equation formalism polarizable continuum model) and a non-polar estimation from Marenich et al. [12]. The hydration free energy was calculated for the single conformation UCSF Chimera created from the SMILES string. No attempt was made to find a low-energy conformation. These free energies were computed initially to confirm that the RETI estimates were reasonable, and as such they were not submitted to the challenge. The SMD predictions are shown in Table 1. The differences compared to experiment range from 6 kcal mol^{-1} more positive to 7 kcal mol^{-1} more negative. The τ is 0.73, r^2 is 0.84 and a MAD is 1.7 kcal mol^{-1}. Interestingly, this is on par with the QM/MM results.

Discussion

Is QM/MM better than MM?

We performed a paired t test with the null hypothesis that the absolute difference between the MM predictions and experiment is equal to the absolute difference between the QM/MM predictions and experiment. The p value of this hypothesis is 0.04, showing that there is a significant difference between the MM and QM/MM predictions at the 95 % confidence level. However, one should also take into consideration whether the difference is of practical significance. To this end, Cohen's effect size, d was computed [38]. When comparing the MM and QM/MM predictions $d = 0.61$ which is considered to be a medium effect. This shows that although there is a significant difference between the MM and QM/MM prediction, the practical importance is not large. If we repeat the tests on only the blind set, $p = 0.03$ and $d = 0.72$, indicating that for this subset the significant difference between the predictions is also of medium effect. Finally, repeating the test on the supplementary set gave $p = 0.06$ and $d = 0.52$, i.e., for the compounds in this subset, the difference between the MM and QM/MM results is not statistically significant at the 95 % confidence level.

These calculations show that there is an advantage of applying the QM/MM corrections on average. Furthermore, the ranking metrics (τ and r^2) are very similar. There is a small difference of 0.05 for τ, which is significant considering the bootstrapped uncertainty, but is negligible from a practical perspective. The difference in r^2 is 0.03, and is not significant. Again, if we just look at the blind set, the differences between the performance of MM and QM/MM is increased. However, for the supplementary set, the difference is negligible.

Can we observe trends in the prediction errors?

If both methods rank the compounds equally well, is there any utility for the more expensive QM/MM approach? Here, we have to look at individual predictions, and investigate whether QM/MM is a more reliable method than pure MM. Therefore we plot the distribution of the absolute errors (see Fig. 6) and list the seven worst predictions. This roughly corresponds to the compounds outside the quartile range but does not include compounds that are not significantly different from the quartile. For MM, this list includes compounds **1**, **3**, **4**, **24**, **32**, **47**, and **52**, and for QM/MM compounds **19**, **20**, **23**, **24**, **47**, **50**, and **52**. Among these compounds we can observe some weak trends.

For the apolar and bulky compounds **19**, **20**, and **50**, the positive QM/MM correction makes the predictions worse. This is also true for the compounds **23** and **24** that are also bulky but contain some polar groups. It should be noted that compound **24** is also problematic for MM. For the polar and bulky 9,10-anthraquinones, compounds **45**–**48** and **51**, the results are varying. The QM/MM results for compounds **45** and **46** are among the 25 % worse performing compounds (although they are not significantly different from the quartile), but the MM results are excellent. The prediction of **47** is poor for both MM and QM/MM. The QM/MM correction improves the prediction

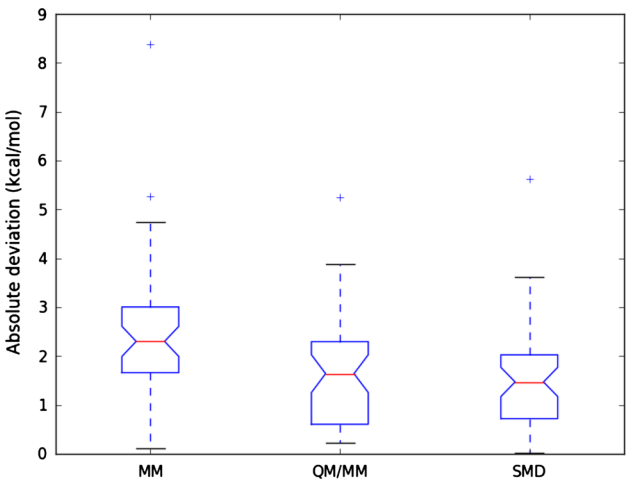

Fig. 6 *Box plot* of absolute deviation between hydration free energy predictions and the experimental data. The *horizontal line* in the middle of the box shows the median and the box covers the interquartile range. The whiskers extend to 1.5 of the interquartile range and the crosses outside are considered to be outlier predictions

only partially. A possible reason for the bad MM prediction is that the long carbon-chain becomes trapped in local conformations, something that could be solved by more extensive sampling.

Along with **24** and **47**, compound **52** is poorly predicted by both MM and QM/MM. MM predicts this compound to have a positive hydration free energy, and because the QM/MM correction is more or less zero, it does not improve that prediction.

The predictions for compounds **1**, **3** and **4** are poor with MM, but are improved by the QM/MM correction. The MM predictions of **2**–**4** are almost indistinguishable, but the QM/MM prediction correctly identifies **3** and **4** as having a more negative hydration free energy. However, it is hard to explain why the MM predictions were poor. It is also difficult to explain why the MM prediction of compound **32** is bad. The sampling should be sufficient for this rather rigid compound, and other similar compounds (e.g., **7** and **10**) are predicted well.

To conclude, it is challenging to find any rational trends that capture the failures of the MM and QM/MM predictions.

Is QM/MM better than a simple approach?

The QM/MM approach to hydration free energy prediction is a comparatively expensive method. The most expensive aspect of the calculation is the sampling at the MM level. An MM simulation of 50 m steps takes roughly 15 h on a single core, and if we need 24 λ-values the MM part requires 360 CPU h. If we also add temperature ladders as we did here for some compounds, the computational cost is 960 CPU h. Each single-point QM/MM energy calculation takes about 15 min on average but can be extensively parallelized. Therefore it is pertinent to see how well our methods fair when compared with something much simpler and cheaper, i.e., a null model. We chose the SMD approach on a single snapshot, generated from the SMILES as our null model. These predictions took about a 15 min per compound, i.e., more than 60 times cheaper than the MM or QM/MM approach. Naturally, this is very naïve approach, as we have not taken any care to find a representative conformation for the calculation (apart from checking that it was chemically correct).

Performing a paired t test between the QM/MM and the SMD predictions (similar to the one above), gives a $p = 0.57$ and $d = 0.10$. This shows that there is no statistically significant difference between the two approaches. Furthermore, for both the blind and supplementary set, the difference is not statistically significant. Also if we look at the rank metrics, τ and r^2, there is a very small difference between the two approaches.

Therefore, we have to look at the individual predictions to see whether the SMD predictions are at least as reliable as the QM/MM results. The distribution of the absolute errors is plotted in Fig. 6, and it is clear that SMD has one outlier just as QM/MM. The seven worst predictions are for compounds **5**, **23**, **24**, **27**, **28**, **35** and **51**. Only compounds **23** and **24** are common with the corresponding list for QM/MM. Of course this means that SMD was able to make more accurate predictions than MM or QM/MM for some compounds. Compounds **27** and **28** have long tails, for which the SMD prediction necessarily is very dependent on the single conformation, but there are similar compounds where the SMD prediction was good. It is clear that by just using a single random conformation we can obtain both very good and very bad predictions. However, for SMD we also have bad predictions that are not as easily explained, as was the case with MM. Hence it seems that we cannot easily reject our null model and therefore the SMD approach is reasonable.

How good are the predictions compared to similar submissions?

There were several submissions to SAMPL4 that used explicit-solvent simulations to make the predictions [8]. Therefore, it is of interest to compare different simulation protocols with each other. Here, we limit the comparison to submissions 004 and 544 that used force fields similar to what we used and standard simulation protocols [39]. This selection limits the possible sources for discrepancies. The two submissions in question used standard molecular dynamics (MD) using GAFF [16] with AM1-BCC charges [17] to describe the ligands in conjunction with the TIP3P model [20] for water. Furthermore, they employed long-range electrostatics using the Ewald summation [39]. Using this protocol, submission 544 obtained a τ of 0.77, an r^2 of 0.90 and a MAD of 1.0 kcal mol^{-1} for all compounds. Submission 004 performed very similarly [39]. All of these metrics are better than those reported for our MM predictions. We can envisage three major sources for this discrepancy. First, the water model is different. We used a TIP4P model whereas the MD submissions 004 and 544 used TIP3P model. It has previously been shown that the TIP4P-Ew water model performs slightly worse than the TIP3P model in predicting hydration free energies, with an average error that is 0.3 kcal mol^{-1} worse when the AM1-BCC charge model was employed [5]. Second, we ignore long-range electrostatics and truncate the interaction at 10 Å. Although this is a crude approximation for some applications such as free-energy calculations with charged species [40], the effect on the hydration free energy of a small, neutral compound (such as the ones in the SAMPL4 set) is probably small when an appropriate cut-off is used

[41]. We have unpublished data that support this finding when it comes to ranking the compounds; we computed hydration free energies for 44 small, neutral compounds previously studied by Mobley et al. [5]. We employed a very similar MD protocol as in the original study. Using an atom-based cut-off at 10 Å, we obtain a τ of 0.89, an r^2 of 0.94 and a MAD of 1.55 kcal mol^{-1}, and when using long-range electrostatics we obtained a τ of 0.86, an r^2 of 0.94 and a MAD of 0.90 (which is similar to the original study [5]). Transferring this knowledge to the SAMPL4 predictions, we might expect a deterioration in the absolute values of the predictions by about 0.7 kcal mol^{-1} on average, but that the ranking should not be affected. Third, the discrepancy between our predictions and the MD submissions could be due to the sampling, i.e., MD versus Monte Carlo. This is the hardest to judge with the results in hand, because the MD submissions do not discuss sampling. For instance, we have shown that for some of the compounds, different starting geometries can produce widely different results. However, the MD submissions only use a single starting conformation [39]. Furthermore, it is not possible to pinpoint the major discrepancies to any particular class of compounds. For instance there is a 1.7 kcal mol^{-1} discrepancy for compound **15** but only a 0.7 kcal mol^{-1} discrepancy for compound **17**, despite showing structural similarity. This suggests that the discrepancy is statistical rather than systematic.

How good a set of results can we expect?

Considering that all methods performed rather equally, at least regarding their ranking abilities, it is of interest to question the quality of the ranking performance. Another way to phrase this is to ask what is the optimum performance that could be obtained for these experimental data. To investigate this we performed bootstrap simulations on the distribution of the experimental data. We computed the average τ and r^2 when both the experimental and predicted free energies were drawn from the experimental distribution, allowing for normally distributed noise [42]. The predicted free energies are drawn from the experimental distribution since we are making the best-case assumption that the predictions are free of systematic error. For the resamples representing the predictions, we allow for a standard deviation of 0.3 kcal mol^{-1}, which is roughly the average standard error of our MM and QM/MM predictions. This is a judicious choice based on our observed prediction uncertainty. For more precise methods, this parameter should be adjusted according to the lower uncertainty. Furthermore, for the resamples representing the experiments, we allow for a variable standard deviation in the range from 0.0 to 2.0 kcal mol^{-1}. This noise represents the uncertainty of the experimental data. The

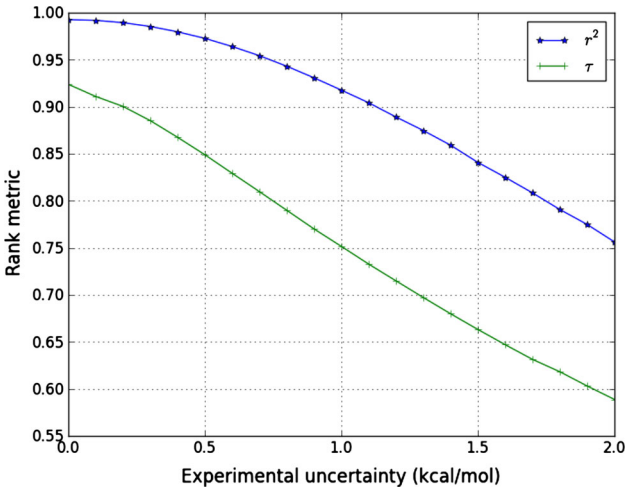

Fig. 7 Estimated optimum r^2 and Kendall's τ as a function of experimental uncertainty. The average rank metric was estimated from 10,000 bootstrap re-samples of the experimental free energies, assuming a 0.3 kcal mol^{-1} uncertainty in predicted free energies and a variable experimental uncertainty

average τ and r^2 over 10,000 bootstrap samples are shown in Fig. 7. We see that given an experimental uncertainty of 0 kcal mol^{-1}, we cannot expect a better τ than 0.92 because of the imprecision in our simulated free energies, i.e., it is unrealistic to strive for a perfect ranking. The reported uncertainty of the experimental data is between 0.1 and 1 kcal mol^{-1}, with an average of 0.3 kcal mol^{-1}. Introducing this uncertainty into the experimental data indicates that we can at most expect a τ of $\sim$0.85. In the literature, it is more common to expect an experimental uncertainty of between 1 and 2 kcal mol^{-1}, and at those levels, we should not expect a τ of any more than between 0.60 and 0.70 to be achievable. This also seems to be the case with the current data set. During the course of this challenge the experimental data of several compounds have either been retracted or revised by several kcal mol^{-1}. This was mostly for compounds where the predictions by all of the participants were outliers, and as such raised the question of experimental validity. Although an average uncertainty of the experimental data of 0.3 kcal mol^{-1} was provided, this should probably be regarded as a lower limit given the problems of determining the experimental free energies of hydration.

Conclusion

We have presented predictions of hydration free energies for 52 compounds in the SAMPL4 challenge. Our predictions were chiefly based on extensive all-atom Monte Carlo simulations. To incorporate explicit polarization in our

simulations we applied an inexpensive QM/MM correction. Overall we observe a MAD of 1.7 kcal mol^{-1} and a Kendall's τ of 0.65 with the QM/MM approach, which is what one would expect given the complexity of the compounds and the experimental uncertainty.

However, we have shown that you would obtain almost identical rankings and correlation with the pure MM approach. We should also note that the QM/MM results are based on the MM sampled configurations, so any error in the MM sampling may very well be propagated to the QM/MM predictions. Furthermore, the QM/MM correction is approximate and consequently it will not necessarily improve the MM predictions. Unfortunately, it seems hard to describe any trends among the prediction failures. We furthermore showed that we obtain an identical overall performance to QM/MM with a very simple approach based on continuum electrostatics (SMD). Future studies have to determine if this is a general conclusion or not. However, it should be noted that although the SMD model seems to work surprisingly well on hydration free energies, one would most likely have to resort to all-atom sampling methods, e.g., Monte Carlo, if one is interested in computing binding free energies.

Finally, we would like to emphasize that in a blind challenge like this, our primary objective was to produce converged, reproducible results. Here, we have taken considerable care to accomplish this and have shown that tools such as multiple independent simulations, multiple starting conformations, dihedral distributions, torsion-RMSD plots and thermodynamic cycles are valuable when assessing the convergence criterion.

Acknowledgments For financial support we acknowledge AstraZeneca Pharmaceuticals and the Wenner-Gren foundations (SG) and Astex Pharmaceuticals (AICM). We acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton.

References

1. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS (2008) J Med Chem 51:769–779
2. Shirts M, Mobley DL, Chodera JD (2007) Annu Rep Comput Chem 3:41–59
3. Sharp KA, Honig B (1990) Ann Rev Biophys Biophys Chem 19:301–332
4. Hirata F (2004) Molecular theory of solvation. Springer, Dordrecht
5. Mobley DL, Dummon E, Chodera JD, Dill KA (2007) J Phys Chem B 111:2242–2254
6. Rocklin GJ, Mobley DL, Dill KA (2013) J Chem Theory Comput 9:3072–3083
7. Guthrie JP (2014) SAMPL4, a blind challenge for computational solvation free energies: the compounds considered. ibid
8. Mobley DL, Wymer K, Lim NM (2014) Blind prediction of solvation free energies from the SAMPL4-challenge. ibid
9. Guthrie JP (2009) J Phys Chem B 113:4501–4507
10. Geballe MT, Skillman AG, Nichools A, Guthrie JP, Taylor PJ (2010) J Comput Aided Mol Des 24:259–279
11. Geballe MT, Guthrie JP (2012) J Comput Aided Mol Des 26:489–496
12. Marenich AV, Cramer CJ, Truhlar DG (2009) J Phys Chem B 113:6378–6396
13. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) J Comput Chem 25:1605–1612
14. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902–3909
15. Gaussian 09 Revision A1 Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery Jr JA, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels A D, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ Gaussian Inc Wallingford CT 2009
16. Wang JM, Wolf RM, Caldwell KW, Kollman PA, Case DA (2004) J Comput Chem 25:1157–1174
17. Jakalian A, Jack DB, Bayly CI (2002) J Comput Chem 23:1623–1641
18. Case DA, Cheatham T, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods R (2005) J Comput Chem 26:1668–1688
19. Molecular Operating Environment (MOE) (2012) 10; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7
20. Jorgensen WL, Chandrasekhar J, Madura JD, Impley RW, Klein ML (1983) J Chem Phys 79:926–935
21. Kirkwood JG (1935) J Chem Phys 3:300–313
22. Woods CJ, King MA, Essex JW (2003) J Phys Chem B 107:13703–13710
23. Shirts MR, Pande VS (2005) J Chem Phys 122:134508
24. Zwanzig RW (1954) J Chem Phys 22:1420–1427
25. Shirts MR, Pitera JW, Swope WC, Pande VS (2003) J Chem Phys 119:5740
26. Woods CJ, King MA, Essex JW (2003) J Phys Chem B 107:13711–13718
27. Swendsen RH, Wang JS (1986) Phys Rev Lett 57:2607–2609
28. Michel J, Verdonk ML, Essex JW (2007) J Chem Theory Comput 3:1645–1655
29. Beutler TC, Mark AE, van Schaik RC, Gerber PR, van Gunsteren WF (1994) Chem Phys Lett 222:529–539
30. Zacharias M, Straatsma TP, McCammon JA (1994) J Chem Phys 100:9025–9031
31. Genheden S, Bodnarchuk M, Michel J, Woods CJ ProtoMS 2.3. http://protoms.org/
32. Beierlein FR, Michel J, Essex JW (2011) J Phys Chem B 115:4911–4926
33. Becke AD (1993) J Chem Phys 98:5648
34. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785–789
35. Kendall M (1938) Biomet 30:81–89
36. Efron B (1979) Anal Stat 7:1–26
37. Genheden S, Ryde U (2010) J Comput Chem 31:837–846

38. Cohen J (2009) Statistical power analysis for the behavioral sciences, 2nd edn. NYU Press, USA
39. Muddana HS, Sapra NV, Fenley AT, Gilson MK (2014) The SAMPL4 hydration challenge: evaluation of partial charge sets with explicit-water molecular dynamics simulations. ibid
40. Bergdorf M, Peter C, Hünenberger PH (2003) J Chem Phys 119:9129
41. Brunsteiner M, Boresch S (2000) J Chem Phys 112:6953
42. Brown SP, Muchmore SW, Hajduk PJ (2009) Drug Discov Today 14:420–427