

Predicting hydration free energies with a hybrid QM/MM approach: an evaluation of implicit and explicit solvation models in SAMPL4

Gerhard König · Frank C. Pickard IV ·
Ye Mei · Bernard R. Brooks

Received: 15 November 2013 / Accepted: 11 January 2014 / Published online: 7 February 2014
© Springer International Publishing Switzerland (outside the USA) 2014

Abstract The correct representation of solute-water interactions is essential for the accurate simulation of most biological phenomena. Several highly accurate quantum methods are available to deal with solvation by using both implicit and explicit solvents. So far, however, most evaluations of those methods were based on a single conformation, which neglects solute entropy. Here, we present the first test of a novel approach to determine hydration free energies that uses molecular mechanics (MM) to sample phase space and quantum mechanics (QM) to evaluate the potential energies. Free energies are determined by using re-weighting with the Non-Boltzmann Bennett (NBB) method. In this context, the method is referred to as QM-NBB. Based on snapshots from MM sampling and accounting for their correct Boltzmann weight, it is possible to obtain hydration free energies that incorporate the effect of solute entropy. We evaluate the performance of several QM implicit solvent models, as well as explicit solvent QM/MM for the blind subset of the SAMPL4 hydration free energy challenge. While classical free energy simulations with molecular dynamics give root mean square deviations (RMSD) of 2.8 and 2.3 kcal/mol, the hybrid approach yields an improved

RMSD of 1.6 kcal/mol. By selecting an appropriate functional and basis set, the RMSD can be reduced to 1 kcal/mol for calculations based on a single conformation. Results for a selected set of challenging molecules imply that this RMSD can be further reduced by using NBB to reweight MM trajectories with the SMD implicit solvent model.

Keywords Hydration free energy calculations · Non-Boltzmann Bennett · Implicit solvent · Explicit solvent · QM/MM

Introduction

Most biological phenomena take place in aqueous solution. Solvation affects conformational changes and protein folding, as well as binding processes of ligands to macromolecules. In order for a ligand to bind to its target, both the ligand and the binding pocket must be stripped of water molecules at the binding interface. This incurs a desolvation penalty to the total free energy change of the process. If solute-solvent interactions for both ligand and binding pocket are not treated correctly, the resulting binding free energies will suffer from substantial errors. In the calculation of absolute binding free energies, these errors are unlikely to cancel systematically in the thermodynamic cycle. This is because the solute-solvent interactions are only considered in one leg of the cycle and are usually very different from the nature of the ligand-receptor environment. Thus, errors from solvation can probably be considered a kind of baseline error for absolute binding free energy calculations.

The most rigorous way to evaluate solute-water interactions is to determine the corresponding hydration free energy, which is the transfer free energy of the solute from an ideal gas phase reference state into aqueous solution. Recently, the

Electronic supplementary material The online version of this article (doi:10.1007/s10822-014-9708-4) contains supplementary material, which is available to authorized users.

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

Y. Mei
State Key Laboratory of Precision Spectroscopy, Institute of Theoretical and Computational Science, East China Normal University, Shanghai 200062, China

SAMPL blind challenges have been established to compare the accuracy of computational methods to predict such hydration free energies [1–3] as well as binding free energies [4]. Those challenges supply data on the expected accuracies of current force fields and solvent models, in particular with respect to the solute-water interactions. In addition, the competitions highlight which methods yield better results.

In SAMPL0 and SAMPL1, the root mean square deviations (RMSD) of free energy calculations based on molecular dynamics simulations (MD) with explicit solvent ranged between 1.3 [1] and 3.6 kcal/mol, [2, 5] which gives a good picture of the errors that can be expected from MD for small systems. For example, the same approach yielded an RMSD of 2.8 kcal/mol when predicting the solvation free energies of 23 small organic compounds in the SAMPL2 competition [6]. The most problematic aspect of those calculations was the selection of the right parameters (in particular charges) [7, 8]. In addition, sampling was considered one of the main culprits for some deviations from experiment. Most quantum chemical methods in the SAMPL competitions employed some form of density functional theory (DFT) in conjunction with implicit solvent, due its relatively high accuracy for relatively low computational costs. These approaches yielded reasonable results, with the RMSD being about 2.5 kcal/mol for SAMPL1 [9] and between 1.5 and 3.5 kcal/mol for SAMPL2 [10, 11]. Most of the submissions varied the functional, implicit solvent model or the the form of geometry optimization done prior to the solvation free energy calculation. The Achilles' heel of using pure QM however, is the difficulty of incorporating conformational entropy into the computed results. If there are several local minima with similar energies on the potential energy surface, the accuracy of the predicted solvation free energies will not reflect the associated entropy.

A promising approach to compute hydration free energies is the combination of quantum-chemical calculations with MD. Thus, sampling can be performed with fast MD, while the free energy results are further refined by conducting the potential energy evaluations with accurate quantum-chemical methods. Recent applications of such approaches involve the calculation of solvation free energies of ligands [12, 13] or the computation of free energy barriers in enzymatic reactions [14–16]. More applications have been made possible by the development of QM/MM alchemical free energy calculations [17, 18] and improved sampling methods [19]. A novel alternative to the approaches mentioned above is the use of reweighting in the form of Non-Boltzmann Bennett (NBB) [20, 21]. Similar to Bennett's acceptance ratio method (BAR), [22] it employs data from two end states to minimize the variance of the estimate. However, the end states involved in the calculation only exist virtually since they are generated by reweighting simulations of closely related states. For example, the trajectories can be generated with MM, while the end states are evaluated through quantum-chemical calculations.

Notably, this approach has also been used by the Ryde group for the host-guest binding free energy challenge of SAMPL4 [23]. In this paper, we provide the first practical evaluation of the QM NBB method for the calculation of hydration free energies. To explore its advantages and weaknesses, we compare it with implicit and explicit solvent MM calculations. In addition, we contrast it with an alternative method that is based on thermodynamic perturbation (TP) [24] and quantum-chemical implicit solvent calculations that were performed for a single conformation.

In particular, we focus on the blind subset of the SAMPL4 hydration free energy challenge, which encompasses 24 out of the total 43 molecules. To put our data into a broader perspective, we would like to refer the reader to an overview of the experimental data [25] and a performance comparison of the computational results [26] in this special issue as a kind of "required reading." As described in Reference [25], the blind subset originally included 24 molecules, but three molecules (7, 8 and 18) were removed from that list after the competition deadline. The remaining compounds still exhibit a wide chemical diversity and include several challenging properties to model.

The remainder of this paper is organized as follows. First, we describe the employed methods in more detail. We then present the results for all submissions to SAMPL4 and assess their accuracy. Next, the results are compared with each other and with quantum-chemical implicit solvent data that was generated after the deadline for SAMPL4. Finally, we conclude with a short discussion on the influence of sampling on the quantum-chemical implicit solvent results for selected cases and also consider the possible influence of protonation states on some of the free energy results. The Supplemental Material at the end of this manuscript provides an exhaustive overview of all QM single conformation hydration free energy results. The trivial or chemical names for the compounds of the blind subset of the SAMPL4 hydration free energy challenge can be found in Table 5 of the Supplemental Material. In addition, a glossary of abbreviations is provided in Table 6 of the Supplemental Material. A complete description of the underlying QM-NBB methodology is reported elsewhere (König et al. *Multiscale free energy simulations: An efficient method for connecting classical MD simulations to QM or QM/MM free energies using Non-Boltzmann Bennett reweighting schemes*, submitted to J. Chem. Theory Comput.).

Methods

Free energy methods

One of the most basic methods to calculate the free energy difference ΔA between an initial state 0 and a final state 1 is

TP [24]. This method is also known in the literature as Free Energy Perturbation, the exponential formula, or Zwanzig's equation.

$$\Delta A^{0 \rightarrow 1} = -\beta^{-1} \ln \langle \exp(-\beta \Delta U_{fw}) \rangle_0 \quad (1)$$

Here, β is the inverse thermodynamic temperature ($\frac{1}{k_B T}$), and ΔU_{fw} is the forward perturbation, consisting of the potential energy difference ($U_1 - U_0$) between the state 0 and state 1 for the coordinates of interest. The angular brackets $\langle \rangle_0$ indicate that this equation is evaluated for an ensemble average of state 0 (e.g., by calculating a simple average over a trajectory of state 0). If state 0 has a potential energy function that is computed with classical mechanics and state 1 is computed with quantum mechanics, it is possible to use Eq. 1 to calculate the free energy difference between those two states based on a MD trajectory. Such free energy differences can then be used to correct errors that arise from the force-field representation in molecular dynamics based free energy calculations [12–16].

An alternative strategy consists of using re-weighting with NBB [20] to account for the differences between MM and QM (we refer to this approach as QM-NBB) [21]. This involves calculating weights for each frame of a MD trajectory by evaluating the biasing potential V^b . In the present case, V^b is given by potential energy differences between molecular mechanics (U^{MM}) and quantum mechanics (U^{QM}).

$$V^b = U^{MM} - U^{QM} \quad (2)$$

The biasing potential is then used in the NBB equation [20]. Generally, this method requires trajectories for both state 0 and state 1.

$$\Delta A^{0 \rightarrow 1} = \beta^{-1} \ln \left(\frac{\langle f(\Delta U_{bw} + C) \exp(\beta V_1^b) \rangle_{1,MM} \langle \exp(\beta V_0^b) \rangle_{0,MM}}{\langle f(\Delta U_{fw} - C) \exp(\beta V_0^b) \rangle_{0,MM} \langle \exp(\beta V_1^b) \rangle_{1,MM}} \right) + C \quad (3)$$

The free energy difference is calculated using Eq. 3, where ΔU_{bw} is the backward perturbation ($U_0 - U_1$), f denotes the Fermi function ($f(x) = \frac{1}{1 + \exp(\beta x)}$) and C is solved for iteratively. We use the notation $\langle \rangle_{x,MM}$ to indicate which end state is evaluated ($x = 0$ or 1), and to indicate that the ensemble averages are evaluated from the MD trajectories.

MD simulations

All MD simulations were conducted with CHARMM, [27, 28] using the PERT module and the CHARMM General Force Field for organic molecules [29]. Hydration free energies were calculated by turning off all non-bonded

interactions of the solute both in gas phase and solution (we will refer to this approach as “MM-TIP3P”). The alchemical mutation was done in two steps: First, all charges of the solute were set to zero (we refer to this process as the “uncharging” process). Second, all Lennard-Jones interactions of the solute were set to zero (the “vanishing” process). In total, the thermodynamic cycle invoked consists in (1) gas phase \rightarrow uncharged solute in gas phase (2) uncharged solute in gas phase \rightarrow solute without any non-bonded interactions in gas phase (3) solute without any non-bonded interactions in gas phase \rightarrow solute without any non-bonded interactions in solution (4) solute without any non-bonded interactions in solution \rightarrow solute without charges in solution (5) solute without charges in solution \rightarrow solute in solution. Step 1 (uncharging in gas phase) was subdivided into six λ points ($\lambda = 0.00, 0.05, 0.15, 0.40, 0.80$ and 1.00), whereas step 2 (vanishing in gas phase) used seven values ($\lambda = 0.00, 0.15, 0.35, 0.65, 0.80, 0.90$ and 1.00). The free energy change of step 3 is zero since the solute is not able to interact with the solvent. Step 4 (negative of the vanishing process in solution) required thirteen λ points ($\lambda = 0.00, 0.05, 0.10, 0.20, \dots, 0.90$ and 1.00) and step 5 (negative of the uncharging process) employed twelve λ points ($\lambda = 0.00, 0.05, 0.10, 0.20, \dots, 0.90, 0.95$ and 1.00).

To ensure proper sampling over all relevant degrees of freedom, λ -Hamiltonian Replica Exchange [30] was employed to exchange structures between neighboring λ points. Exchanges were attempted every 1,000 steps. In solution, exchange rates varied between 1 and 82 % with an average exchange rate of 24% over all molecules and all λ states. In gas phase, the average exchange rate was 43 %. Since the last λ point involves an ideal gas state of the solute that only incorporates bonded terms in the potential energy calculation (i.e., no Lennard-Jones interactions and no electrostatic interactions are considered), energy barriers are considerably lower. This allows fast sampling if the conformations are propagated through frequent exchanges.

MD simulations in gas phase were performed with Langevin dynamics at a temperature of 300 K and using time step of 1 fs with a friction coefficient of 5 ps^{-1} on all atoms. No cutoffs were used. In solution phase, we used 1,492 water molecules in a truncated octahedron box that was cut from a cube with a side length of 38.60 Å. A Nosé-Hoover thermostat was used to keep the temperature at 300 K. Long range electrostatic interactions were computed with the Particle Mesh Ewald method [31] and Lennard-Jones interactions were switched off between 10 and 12 Å. Soft cores, as implemented in the PERT module of CHARMM, were used to avoid the end point problem. All molecules were first equilibrated for 0.1 ns using constant pressure and each λ point was further equilibrated for 0.1 ns using constant volume. For expedience, we employed constant volume for production (thus neglecting

the $P\Delta V$ work, which is very small). Simulations of 50 ns were used in gas phase, saving frame coordinates every 1,000 steps. In solution, the simulations were 0.5–1.0 ns long and trajectories were written every 20 steps. All simulations were repeated three times with different random seeds to compute standard deviations.

For the implicit solvent MD simulations (MM-GB), the hydration free energies were computed without λ intermediate states. For gas phase, $\lambda = 0.00$ of the uncharging process was employed (see description above). Solvation phase simulations were carried out by employing the “Generalized Born using Molecular Volume” (GBMV) implicit solvation model [32]. As in gas phase, the GBMV simulations were performed with Langevin dynamics at a temperature of 300 K and using a friction coefficient of 5 ps^{-1} on all atoms. All non-bonded interactions were switched off between 16 and 18 Å. The simulation length was 10 ns, using a time step of 1 fs. Frames were saved every 200 steps. The simulations were repeated three times with different random seeds to compute standard deviations.

QM hydration free energy calculations

In this work, DFT was used extensively to calculate solvation free energies directly using standard QM techniques, and to supplement free energy calculations based upon classical MD simulations and free energy perturbation techniques.

In Gaussian 09, [33] the B3LYP [34] and M06-2X [35, 36] functionals were used in conjunction with a variety of double- ζ Pople [37] and Dunning [38] style basis sets to perform geometry optimizations of the structures supplied by Peter Guthrie. Next, analytic Hessians were computed to verify that geometries had been optimized to a true minimum on the potential energy surface, and to compute zero-point energy (ZPE) corrections. Finally, a triple- ζ basis set was used to perform a single-point energy calculation, to correct for basis set deficiencies. All B3LYP calculations used the default Gaussian numerical quadrature, while M06-2X calculations used the “UltraFine” grid. All geometries and electronic densities were optimized using their respective “Tight” convergence criteria. To mimic the bulk effects of aqueous solvent, both the default implicit solvent [39] and the SMD implicit solvent [40] models were used. Solvation free energies were computed in four different ways. (a) The vertical energy was computed by taking the optimized gas geometry and performing a single point energy calculation in the polarizable continuum reaction field. (b) The relaxed energy was calculated by taking the differences of energies computed using both optimized gas and solvation geometries. (c) The relaxed energy was then corrected to account for ZPE

effects. (d) Finally, the total energy was computed by taking the relaxed energy and including additional corrections for ZPE and vibrational and rotational corrections to 298.15 K.



We also calculated $\text{p}K_{\text{a}}$ values for molecules 22, 23 and 24 in an effort to guide the QM computation of hydration free energies for these potentially problematic molecules. In addition to hydration free energies for molecules 22, 23 and 24, the $\text{p}K_{\text{a}}$ calculations also required analogous computations to be performed on the conjugate acids of the respective neutral species. Molecule 22 and its conjugate acid (22h) had multiple low lying conformers, and thus necessitated the optimization of multiple geometries. We used the relationship $\text{p}K_{\text{a}} = \Delta G^\circ / 2.303RT$, when calculating $\text{p}K_{\text{a}}$ values. Here R is the gas constant, T the absolute temperature and $\Delta G^\circ = \Delta G_{\text{aq}}$, where ΔG_{aq} is defined by Eq. 4. Values for AH_{aq} and A_{aq}^- are computed directly using the pure QM protocol outlined above. A value for H_{aq}^+ is also necessary to calculate a $\text{p}K_{\text{a}}$ value. This is obtained starting from $\text{H}_{\text{gas}}^+ = -6.28 \text{ kcal/mol}$ at 298 K, and then correcting it to aqueous phase via a value for $\Delta G_{\text{solv}}(\text{H}^+)$. [41] Finally, a standard state correction of -1.89 kcal/mol is applied to account for moving from a gas-phase pressure of 1 atm to a liquid-phase concentration of 1 M. Unfortunately, there are a wide variety of commonly used values for $\Delta G_{\text{solv}}(\text{H}^+)$ values in the literature, ranging from -259 to -264 kcal/mol . Because an error of 1.36 kcal/mol in ΔG° yields an error of 1 $\text{p}K_{\text{a}}$, it is difficult to draw quantitative conclusions from our predicted $\text{p}K_{\text{a}}$ values. However, qualitative predictions should be valid. In this work we used a value of $\Delta G_{\text{solv}}(\text{H}^+) = -264.0 \text{ kcal/mol}$, as recommended by the authors of SMD.

QM and QM/MM simulation post-processing

To reduce computational costs, we use an indirect free energy approach [18, 21]. This entails that, instead of performing NBB for every λ step, we only employ it to correct the first λ step of the MM alchemical uncharging transformation ($\lambda = 0.00 \rightarrow \lambda = 0.05$) in both gas phase and solution (steps 1 and 5 in “MD simulation” section). Thus, QM calculations only have to be performed on those four trajectories instead of all 36 of them (saving about 90 % of the costs of the direct approach). Since the free energy is a state function and the end points involved are the same, the results do not change or depend on the number of λ steps in between. Translated to the notation of Eq. 3, $\lambda = 0.00$ is state 0 and $\lambda = 0.05$ serves as end state 1. Consequently, the fully charged system (state 0) is turned into a system with solute partial charges scaled by a factor of 0.95

($\lambda = 0.05$ or rather state 1). The remaining free energy steps are then conducted with pure MD using BAR (which is equivalent to setting all V^b in Eq. 3 to zero). Since scaling partial charges is difficult to achieve in most QM methods, we used MM for state 1. Thus, we use NBB to calculate the free energy difference between a fully charged QM state and an alchemical MM state with scaled charges for the solute. Therefore, U_0 is a QM potential energy for the current conformation and V_0^b is the energy difference between MM and QM for that fully charged system. However, U_1 is a MM energy for a system with scaled charges. Since we actually generate state 1 in our MD simulations, it is not a virtual end state and, therefore, does not require re-weighting (i.e., $V_1^b = 0$). This approach minimizes the number of required QM calculations. In addition, only every hundredth or thousandth MD step is used for free energy calculations, which reduces the effects of auto-correlation between consecutive frames. Therefore, the expensive QM calculations are performed for only a small fraction of the total number of MD steps.

Trajectories were reweighted using single point QM or QM/MM potential energy calculations. In this work, we considered three approaches. a) We used TP to calculate corrections for gas phase and solution free energy differences of MM-TIP3P based on QM/MM calculations of $\lambda = 0.00$ (we refer to this approach as QM/MM-TP). b) We corrected MM-TIP3P by using NBB based on QM/MM calculations of $\lambda = 0.00$ and $\lambda = 0.05$ in gas phase and solution (QM/MM-NBB). c) For selected examples (molecules 1, 22, 23, 24), we used NBB calculations with SMD implicit solvent to directly compute the hydration free energy based on $\lambda = 0.00$ in gas phase and $\lambda = 0.00$ in solution (SMD-NBB).

Gaussian 09 was used to perform reweighting with the implicit SMD solvation model in SMD-NBB. We performed a series of calculations to assess the cost to benefit ratio of various DFT techniques. After conducting this analysis we used the B3LYP and M06-2X functionals with the 6-31G(d) basis in conjunction with loose self consistent field convergence criteria and the default quadrature. While these options are inappropriate for tight, accurate geometry optimizations, their usage here with respect to the options employed in our geometry optimizations incurred minimal errors (absolute errors lower than ~ 200 μ H, and solvation free energy errors less than ~ 0.005 kcal/mol). Using these options resulted in an about five to ten times speedup. For the selected examples, only 1000 frames could be evaluated from each trajectory in the gas phase and the solution phase due to time constraints. This represents only 2% of the total number of frames in gas phase; and 10% of the frames in solution, these small numbers of computations reduce the statistical precision of this approach.

Q-Chem, [42] driven by the CHARMM/Q-Chem interface, [43] was used to perform reweighting with explicit solvent with QM/MM. The QM region consisted of the solute and was modeled at the B3LYP/6-31G(d) level, while the solvent was modeled classically with the TIP3P water model; [44] a default quadrature and an SCF convergence criterion of 10 were used. Lennard Jones interactions between the solute and the solvent were calculated with CHARMM, while all electrostatic interactions were calculated with Q-Chem. Furthermore, the partial charges from the solvent molecules were able to polarize the electron density during the self consistent optimization process, allowing the solute to respond to the effects of its electrostatic environment. Since periodic boundary conditions and Particle Mesh Ewald are currently not supported by Q-Chem, the explicit solvent QM/MM calculations used a single box of water molecules that were centered around the solute for each frame of the trajectory. The biasing potential was calculated by performing corresponding potential energy evaluations without periodic boundary conditions and Particle Mesh Ewald calculations with MM (c.f. Ref. [21]).

Results and discussion

In Table 1 we present an overview of our submissions to SAMPL4. The methods are ordered from left to right based on their RMSD with respect to experiment (RMSD shown in the last row). Starting with the least accurate method: MD based free energy simulations with implicit solvent (MM-GB, first column), MD based free energy simulations with TIP3P explicit solvent (MM-TIP3P, second column), MM-TIP3P results corrected with TP to QM/MM states (QM/MM-TP) and finally, the NBB QM/MM hybrid free energy approach with explicit solvent (QM/MM-NBB, fourth column). The experimental results are shown in the rightmost column. Except for MM-GB, all methods are based on the same set of trajectories and the use of replica exchange harmonizes the sampling among λ points (especially for neighboring λ points). Thus the sampled parts of phase space are the same for the three methods and any errors that arise from missing important regions are expected to be consistent. This provides a fair test to compare the accuracy and precision of these methods.

In terms of accuracy, all RMSDs are in the range of data from previous SAMPL competitions (between 1.3 and 2.6 kcal/mol for SAMPL0, [45] between 2.4 and 3.6 kcal/mol for SAMPL1, [2, 9] between 1.5 and 2.8 kcal/mol for SAMPL2 [3, 6, 10] and between 1.2 and 4.3 kcal/mol in SAMPL3) [7, 46–49]. For the SAMPL4 blind subset, the best prediction has an RMSD of 1 kcal/mol and the average RMSD (after the removal of clear outliers with

Table 1 Overview of results that were submitted to the SAMPL4 hydration free energy challenge

Molecule ^a	MM-GB ^b	MM-TIP3P ^c	QM/MM-TP ^d	QM/MM-NBB ^e	Exp. ^f
1	−19.8 ± 0.7	−17.5 ± 0.4	−20.9 ± 2.1	−21.5 ± 0.7	−23.6 ± 0.3
2	−4.0 ± 0.1	−4.7 ± 0.4	−1.5 ± 2.2	−3.6 ± 1.0	−2.5 ± 0.9
3	−3.4 ± 0.2	−4.3 ± 0.1	−5.1 ± 0.4	−5.7 ± 0.3	−4.8 ± 0.3
4	−3.7 ± 0.1	−4.5 ± 0.1	−5.8 ± 1.4	−5.6 ± 0.8	−4.5 ± 0.2
5	−4.6 ± 0.2	−6.2 ± 0.2	−5.4 ± 0.8	−5.7 ± 0.4	−5.3 ± 0.1
6	−3.7 ± 0.0	−3.4 ± 0.2	−4.2 ± 0.4	−5.0 ± 0.4	−5.3 ± 0.2
9	−14.1 ± 0.1	−12.6 ± 0.0	−8.9 ± 3.0	−10.0 ± 1.3	−8.2 ± 0.8
10	−8.2 ± 0.0	−8.3 ± 0.1	−6.5 ± 1.1	−6.9 ± 0.8	−6.2 ± 0.4
11	−11.5 ± 0.0	−10.1 ± 0.2	−6.8 ± 1.2	−7.7 ± 0.9	−7.8 ± 0.8
12	−4.6 ± 0.0	−4.9 ± 0.1	−4.3 ± 0.4	−4.6 ± 0.3	−4.4 ± 0.4
13	−3.2 ± 0.1	−4.7 ± 0.1	−5.6 ± 0.9	−5.9 ± 0.5	−5.0 ± 0.4
14	−8.1 ± 0.0	−6.7 ± 0.1	−5.3 ± 0.7	−5.5 ± 0.3	−4.1 ± 0.2
15	−3.2 ± 0.0	−4.3 ± 0.2	−7.5 ± 1.2	−6.4 ± 0.3	−4.5 ± 0.1
16	−2.3 ± 0.1	−3.5 ± 0.2	−4.4 ± 0.4	−4.4 ± 0.3	−3.2 ± 0.3
17	−3.6 ± 0.1	−4.6 ± 0.1	−3.8 ± 0.3	−4.8 ± 0.6	−2.5 ± 0.3
19	−0.7 ± 0.0	−2.8 ± 0.0	−3.9 ± 0.3	−4.4 ± 0.6	−3.8 ± 0.1
20	−1.5 ± 0.0	−1.9 ± 0.0	−2.7 ± 0.3	−2.6 ± 0.1	−2.8 ± 0.1
21	−6.5 ± 0.1	−7.8 ± 0.0	−9.5 ± 0.5	−9.3 ± 0.4	−7.6 ± 0.1
22	−7.2 ± 0.1	−6.2 ± 0.2	−0.9 ± 1.7	−3.5 ± 2.1	−6.8 ± 0.1
23	−3.8 ± 0.1	−6.4 ± 0.2	−5.9 ± 1.0	−5.9 ± 0.5	−9.3 ± 0.6
24	−2.1 ± 0.1	−4.6 ± 0.2	−5.1 ± 0.7	−5.5 ± 0.7	−7.4 ± 0.6
$\langle\sigma\rangle^g$	0.1	0.2	1.0	0.6	0.3
RMSD ^h	2.8	2.3	2.0	1.6	

Experimental results are shown on the right, while the computational results are sorted from the method with worst RMSD from experimental results on the left to the best predictions in terms of RMSD on the right. All energies are in kcal/mol

^a For the trivial or chemical names of each compound, please see Table 5 in the Supplemental material

^b MD simulations based free energy calculations with the Generalized Born-based implicit solvent model GBMV [32]

^c MD simulations based free energy calculations with explicit solvent

^d Thermodynamic Perturbation (Zwanzig equation [24]) from MM to B3LYP/6-31G(d) QM/MM end states

^e QM/MM hybrid approach with Non-Boltzmann Bennett using B3LYP/6-31G(d) on the solute

^f Experimental hydration free energy results

^g Average standard deviation of results

^h Root mean square deviation from experimental results

RMSD > 10 kcal/mol) is 2.2 kcal/mol. We conclude that the performance of MM-GB is below average, whereas MM-TIP3P and QM/MM-TP can be considered average. Only the QM/MM-NBB approach is better than the median RMSD of 1.9 kcal/mol. Nevertheless, we think that it is very instructive to compare the relative performances of the four methods considered here, since all methods were treated in a consistent way.

The first two columns show that implicit solvent (MM-GB, RMSD = 2.8 kcal/mol) is not as accurate as explicit solvent (MM-TIP3P, RMSD = 2.3 kcal/mol), at least when using MD simulations. Given that implicit solvent employs several approximations and neglects the molecular nature of the solvent, the qualitative ranking is not surprising. Quantitatively, it is interesting to observe such a small difference

between both methods (≈ 0.5 kcal/mol), indicating that the GBMV algorithm is effective at mimicking bulk solvent. The relative performance of the two methods also agrees well with what has been previously observed for solvation free energies of amino acids (0.3 kcal/mol) [50, 51]. For explicit solvent, the RMSD of MM-TIP3P (2.4 kcal/mol) is similar to the RMSD that we have obtained for host-guest binding free energies in SAMPL3 (2.6 kcal/mol), [52] using the same setup in terms of force-field, cutoffs and number of water molecules. This is an indicator that the errors from this hydration free energy challenge are useful in estimating the expected error of other applications of free energy simulations, which makes it a valuable benchmark system.

The relative performance of MM-TIP3P and QM/MM-TP is mostly determined by the different treatments of the

solute. QM/MM-TP is based on the MM-TIP3P results, but performs TP correction steps at the end points of the alchemical uncharging processes in both the gas and solution phases to account for solute polarization. This is done by using QM/MM calculations with B3LYP/6-31G(d) for each frame of the MM trajectories. Since both methods use TIP3P water, errors that arise from the classical treatment of water are probably the same for both data sets. Overall, the correction steps lead to an increase of accuracy by 0.3 kcal/mol, which is slightly less than the step from implicit solvent to explicit solvent. A striking feature is the decreased precision. In the penultimate row of Table 1 we list the average standard deviation ($\langle\sigma\rangle$) over all molecules for each method. While the precision of MM-TIP3P (second column) is on the same order as the experiments ($\langle\sigma\rangle = 0.2$ and 0.3 kcal/mol for simulation and experiment), the average standard deviation of QM/MM-TP is significantly higher ($\langle\sigma\rangle = 1.0$ kcal/mol). This can be attributed to the large differences between the QM and MM potential energy surfaces, which leads to greater fluctuations of the perturbations.

One way to mitigate the effect of such fluctuations in free energy simulations is to use BAR instead of TP. BAR employs data from two trajectories to minimize the variance of the estimator and is more efficient than the one-sided TP method used in the example above. To post-process MD trajectories with potential energies that were obtained with QM/MM, it is necessary to use NBB instead of BAR as outlined in the Methods section. The results for this approach are shown in column four (QM/MM-NBB). Compared to QM/MM-TP, QM/MM-NBB employs twice the number of QM/MM data points, but it increases the accuracy by 0.4 kcal/mol to an RMSD of 1.6 kcal/mol and reduces $\langle\sigma\rangle$ from 1.0 to 0.6 kcal/mol. Compared to the underlying MD simulations (MM-TIP3P), the accuracy increases by 0.7 kcal/mol, which exceeds the improvement from implicit solvent to explicit solvent (0.6 kcal/mol). While the QM/MM-NBB approach is the most accurate of all methods considered here, the standard deviations are still slightly too high to be acceptable. However, this problem is easily addressed by running longer simulations.

To summarize, the MM simulations (MM-GB and MM-TIP3P) exhibit very low standard deviations of about 0.1 to 0.2 kcal/mol, which can probably be attributed to the use of enhanced sampling methods in the form of Hamiltonian Replica Exchange. However, they suffer from RMSD of 2.3–2.8 kcal/mol, which are most likely the effect of an imperfect force field. For the QM/MM approaches, the situation is the other way around: Accuracy is improved (RMSD of 1.6 to 2.0 kcal/mol), but the precision is inadequate ($\langle\sigma\rangle = 0.6$ to 1.0 kcal/mol). This demonstrates that more sampling is required for converged results with both the QM/MM-TP and QM/MM-NBB approach.

Estimating the reliability of predictions

Accuracy and precision are not the only criteria for determining the utility of a method, as it should also be possible to assess the reliability of a prediction a priori. In Table 2, we list the deviations from experimental results (ϵ) for MM-TIP3P (first column) and QM/MM-NBB (fifth column) as well as some measures that might influence accuracy (columns two to four pertain to MM-TIP3P, the two rightmost columns are for MM-NBB). In the last row, we supply correlation coefficients (R) for the respective column with ϵ . The correlation coefficients serve to determine whether the errors can be explained by the respective measure: $R = 0$ implies no correlation (most likely no influence on accuracy), while $R = -1$ and $R = +1$ means negative and positive correlation, respectively. In addition, molecules for which the computational method exhibits deviations of more than 2 kcal/mol from experiment are underlined in Table 2 to identify the most problematic cases.

One of the most intuitive indicators for unreliable results is the standard deviation, σ (shown in the second column for MM-TIP3P and the sixth column for QM/MM-NBB). High standard deviations can be a sign of poor sampling or convergence. For MM-TIP3P, the average standard deviation is quite low ($\langle\sigma\rangle = 0.2$ kcal/mol, see Table 1), so it is unlikely that the errors are caused by sampling problems. This is also indicated by the low correlation between σ with ϵ , as $R = 0.1$, which is nearly uncorrelated. For QM/MM-NBB, however, the correlation coefficient between σ and ϵ is 0.4, which indicates that high standard deviations are a warning sign for poor results in the QM/MM hybrid free energy calculations. The below-average standard deviations of molecules 17 and 23 demonstrate that a low standard deviation does not automatically imply high quality results.

The MM-TIP3P results are generated with the CGenFF force field, which assigns parameters and charges based on the similarity to existing parameters. To estimate the quality of the force field representation, CGenFF provides a penalty score [53] for both the bonded parameters and charges (shown in Table 2). Penalties that are lower than 10 indicate a high confidence in the quality of the generated parameter. Penalties between 10 and 50 mean that some basic validation is recommended prior to performing production simulations, while penalties higher than 50 indicate poor availability of analogous parameters. Notably, 9 out of the 24 molecules have parameter scores above 50, highlighting the fact that obtaining reliable parameters is still a challenging problem for MM.

There is mild correlation between penalty scores and errors in predicted solvation free energies ($R = 0.4$ for the bonded parameters and $R = 0.3$ for charges). This indicates that the quality of the predictions can likely be increased by improving the quality of the parameterization, and that bad

Table 2 Correlation between deviation from experimental results (ϵ) and potential measures for the reliability of the prediction

Molecule	MM-TIP3P				QM/MM-NBB		
	ϵ^a	σ^b	Param ^c	Charge ^c	ϵ^a	σ^b	ovl ^e
1	3.8	0.4	4.0	0.3	2.2	0.7	3.7
2	1.5	0.4	33.0	17.5	1.1	1.0	3.3
3	1.4	0.1	33.0	0.7	1.0	0.3	6.6
4	0.7	0.1	33.0	0.7	1.2	0.8	8.7
5	0.7	0.2	40.2	0.0	0.3	0.4	11.7
6	1.5	0.2	40.2	22.0	0.2	0.4	8.6
9	5.9	0.0	53.0	31.1	1.7	1.3	1.2
10	1.9	0.1	53.0	31.1	0.6	0.8	4.2
11	3.7	0.2	53.0	31.1	0.0	0.9	2.4
12	0.2	0.1	16.0	9.8	0.2	0.3	12.5
13	1.8	0.1	46.0	20.0	0.8	0.5	4.2
14	4.0	0.1	108.0	20.0	1.4	0.3	15.2
15	1.3	0.2	26.5	3.5	1.9	0.3	13.8
16	0.9	0.2	0.9	0.5	1.2	0.3	10.0
17	1.1	0.1	4.9	2.8	2.3	0.6	11.4
19	3.0	0.0	1.8	0.0	0.7	0.6	20.3
20	1.3	0.0	94.4	5.0	0.2	0.1	31.2
21	1.1	0.0	95.5	39.9	1.7	0.4	22.8
22	0.4	0.2	82.4	43.8	3.2	2.1	0.8
23	5.6	0.2	121.5	23.9	3.4	0.5	4.0
24	5.3	0.2	121.5	29.7	1.9	0.7	4.4
R		0.1	0.4	0.3		0.4	−0.3

Only the MM-TIP3P(left) and QM/MM-NBB results(right) are considered. Molecules with deviations of more than 2 kcal/mol from experiment are italicized. All energies given in kcal/mol

^a Deviation from experimental results

^b Standard deviation of computational result

^c Paramchem penalty score for the assigned CGenFF parameters

^d Paramchem penalty score for the assigned CGenFF charges

^e Overlap between forward and backward perturbations (as a percentage) between MM and QM/MM in gas phase (see Eq. 5)

scores may lead to inaccurate results, but not necessarily (e.g. molecule 22 has a small error of just 0.4 kcal/mol even though the parameter and charge penalties are very large). Conversely, a small penalty does not guarantee good agreement, as demonstrated by molecule 1, which exhibits one of the lowest penalties both in terms of bonded parameters and charges, but turns out to have one of the highest deviations from experimental results. This might be attributable to its highly polar nature (M06-2X/cc-pVTZ predicts a dipole of nearly 3 D). Thus, the force field parameters are very good, but the lack of polarization terms in the MM Hamiltonian prevents an accurate potential energy prediction. This argument is corroborated by the high accuracy of the QM with respect to experiment at all levels of theory. This does not mean that the provided penalties are useless, it

indicates that there are more sources of error than just the parameterization (e.g. shortcomings in the electrostatic description of the water model, or a lack of polarization as discussed above).

No parameter scores are available for QM/MM-NBB, as the solute in the QM/MM calculations is parameter-free except for the van der Waals interactions with the classical water molecules. However, the hybrid approach might be affected by errors of the underlying MM trajectories. For example, if the MM simulation never visits the parts of phase space that are low lying on the QM/MM potential energy surface, the re-weighting procedure will not correct for this error. However, if the relevant parts of the QM/MM surface are visited by the simulation, but not generated with the correct target Boltzmann probabilities, the re-weighting can account for the resulting deviations. The fraction of time spent in regions that are relevant for QM/MM can be checked by calculating the overlap (ovl) between the forward and backward perturbations of the end states involved.

$$ovl = \int_{-\infty}^{+\infty} 2 \frac{p_0(\Delta U_{fw}) p_1(-\Delta U_{bw})}{p_0(\Delta U_{fw}) + p_1(-\Delta U_{bw})} \quad (5)$$

Here, p denotes a normalized density function and the subscript specifies the respective end state. In practice, ovl is approximated by calculating histograms for ΔU_{fw} and ΔU_{bw} for the trajectories of states 0 and 1, and using Eq. 5 on those histograms.

The overlap is inversely related to the expected variance of the free energy estimate in BAR (c.f. equation 11 in Ref. [22]). Therefore, a low overlap is expected to lead to larger errors. This effect can be seen in the rightmost column of Table 2, where the overlap has a correlation coefficient of −0.3 with the error, ϵ . However, this is less than the correlation between σ and ϵ . Since the correlation between ovl and the standard deviation σ is −0.6, it is likely that a low overlap has an indirect influence on accuracy by leading to higher standard deviations. This, in turn, increases the probability of errors. In addition, a high overlap is not a guarantee of good results, as demonstrated by molecule 21, where the overlap is 22.8% (the second highest overlap in Table 2), but the error is still 1.7 kcal/mol. Still, as a rule of thumb, we recommend keeping the overlap higher than 1% by optimizing the MM parameters and adding significantly more data points in cases where the overlap is lower than 5%.

Improving the accuracy of hydration free energy calculations

The results from the prior section demonstrate that there is no clear metric to predict improvements to the QM-NBB free energy results. Despite this, there are multiple

straightforward ways to increase the accuracy of the reweighting calculations. For example, more accurate functionals and larger basis sets may be readily deployed in the QM/MM calculations. Due to the sizable computational costs of reweighting trajectories, an exhaustive test of different combinations of functionals and basis sets would go well beyond the scope of this paper. We therefore resorted to employing pure QM hydration free energy calculations based on a single low lying conformation for each molecule. Since the same conformation is used in all cases, the effects from conformational entropy are probably very similar. Thus, the relative ranking of the methods can still be used to evaluate the optimal combination of affordable functionals and basis sets for reweighting.

The B3LYP/6-311++G(3df,3pd)//6-31+G(d,p) level of DFT in conjunction with IEF-PCM implicit solvation was initially used to calculate pK_a values for molecules 22, 23 and 24. As the computations required to calculate the pK_a are a superset of the calculations required to compute the hydration free energy, we decided to determine hydration free energies for all members of the blind subset. This level of theory gives a pragmatic baseline measurement of the capabilities of a generic and very affordable quantum chemical methodology. While these initial QM results enjoyed only fair agreement (RMSD = 3.0 kcal/mol) with experiment, there is a very clear path for obtaining increasingly accurate results, even without a priori knowledge of the experimental solvation free energies. First, the SMD implicit solvation method is purpose built for obtaining accurate solvation thermochemical properties, this method has been employed successfully in previous SAMPL competitions [9, 10]. Indeed, by switching from the Gaussian 09 default implicit solvation method to the SMD method, dramatic improvements in computed energies are realized. B3LYP with a large Pople triple- ζ and the IEF-PCM implicit solvation method gives an RMSD of 2.90 kcal/mol with respect to experiment. When SMD is used instead, the RMSD shrinks to 1.52 kcal/mol. The best M06-2X calculations are also improved by SMD, from an RMSD of 2.82 kcal/mol to an RMSD of 1.19 kcal/mol. Second, while the success of the B3LYP functional is well known to the quantum chemistry community, the M06-2X functional has emerged as a worthy successor, especially for first row atoms that are critical in biochemistry. Table 3 shows that when switching from B3LYP/6-31G(d) to M06-2X/6-31G(d) we reduce errors with respect to experiment from an RMSD of 1.71 kcal/mol to an RMSD of 1.42 kcal/mol when using the SMD solvation model. This trend is also present with analogous triple- ζ quality calculations, as errors with respect to experiment decrease from an RMSD of 1.52 kcal/mol to an RMSD of 1.19 kcal/mol (see SI for more information). This increased accuracy is consistent with expectations.

Increasingly large basis sets may also be used to reduce errors. Indeed when moving from a double- ζ quality basis set in the B3LYP/SMD calculations to a triple- ζ quality basis set, the RMSD with respect to experiment reduces from 1.75 kcal/mol to 1.52 kcal/mol. The improvements for moving from cc-pVDZ to cc-pVTZ with M06-2X/SMD are also dramatic. Table 11 shows the RMSD decreases from 1.45 kcal/mol to 1.19 kcal/mol. Unfortunately, these larger basis sets incur a substantial computational cost, and are too expensive to use for trajectory reweighting. The other suggested improvements to the QM reweighting calculations are affordable, and their effectiveness was tested.

Using our pure QM calculations as a guide, we tried a few alternative QM protocols for trajectory reweighting. The combination of M06-2X and the SMD implicit

Table 3 Comparison of pure QM single conformation hydration free energy calculations

Molecule	B3LYP 6-31G(d)	M06-2X 6-31G(d)	M06-2X cc-pVDZ	M06-2X cc-pVTZ	Exp.
	6-31G(d)	6-31G(d)	cc-pVDZ	cc-pVTZ	
1	−24.97	−25.88	−22.04	−24.47	−23.62
2	−2.18	−3.02	−2.90	−3.79	−2.49
3	−4.21	−5.88	−4.93	−5.76	−4.78
4	−4.41	−4.61	−4.42	−5.44	−4.45
5	−3.87	−4.54	−4.34	−4.67	−5.33
6	−6.85	−7.40	−6.99	−7.35	−5.26
9	−7.99	−8.24	−7.10	−8.29	−8.24
10	−5.09	−6.16	−5.16	−5.28	−6.24
11	−6.49	−7.28	−6.19	−7.02	−7.78
12	−3.47	−3.61	−3.41	−4.80	−4.44
13	−4.53	−5.12	−4.31	−5.06	−5.03
14	−2.79	−3.46	−2.99	−4.29	−4.09
15	−4.18	−4.27	−3.93	−5.25	−4.51
16	−3.14	−3.85	−3.02	−3.69	−3.20
17	−2.96	−3.31	−2.73	−4.05	−2.53
19	−2.19	−3.26	−3.37	−3.53	−3.78
20	−0.78	−1.89	−1.75	−1.92	−2.78
21	−6.16	−8.04	−8.26	−9.07	−7.63
22	−6.28	−7.47	−5.98	−6.46	−6.78
23	−4.28	−4.69	−6.65	−7.45	−9.34
24	−3.56	−4.57	−5.60	−6.38	−7.43
RMSD	1.71	1.42	1.15	1.01	

The first row indicates the functional, the second row the basis set used in the geometry optimization and the third row indicates the basis set used to compute the energy at a frozen geometry. All solvation calculations used the SMD implicit solvent model. The B3LYP calculations include corrections for ZPE and Gibbs free energy to 298.15 K, while the M06-2X calculations include ZPE corrections. Root mean square deviations from experimental results are shown in the last row. All data are in kcal/mol

solvation was very effective. This protocol was carried forward into NBB trajectory reweighting calculations for molecules 1, 22, 23 and 24. These four molecules were chosen because we felt they are representative of the disparate challenges one faces when modeling free energies of hydration. In addition, they represent some of the most pathological cases for QM/MM-NBB. Molecule 1 is exceptionally polar, and may have several local energy minima with comparable energies. Molecule 22 has multiple protonizable sites and low lying conformations in both its acidic and conjugate base forms. Molecules 23 and 24 are both relatively strong bases. Table 4 shows our implicit solvent NBB results (columns 4 and 5) and the analogous single conformation QM hydration free energies (columns 2 and 3). For comparison, the explicit solvent QM/MM-NBB results are listed on the left.

Both types of SMD-NBB exhibit significantly lower RMSD than the single conformation QM hydration free energies or the QM/MM-NBB scheme. In this context, the extremely high RMSD of 4.3 kcal/mol for QM/MM-NBB might seem surprising, but this is mostly a result of selecting its most “pathologic” cases. In addition, the TIP3P water model in our MD simulations was not optimized for use in QM/MM calculations, so its employment can probably be regarded as mixing disparate methods that were not necessarily intended to work well together. In this light, the results of QM/MM-NBB in Table 4 stand to reason.

As anticipated, the best results are from M06-2X/SMD with the relatively modest 6-31G(d) basis set. Fortuitously, the predicted hydration free energy of molecule 1 is now in virtually perfect agreement with experiment. Molecule 22 is also in very strong agreement with experiment as well (<1 kcal/mol error). Molecules 23 and 24 are also improved, however errors greater than 1 kcal/mol persist, as our predicted free energies are still not negative enough. This effect is likely attributable, at least in part, to the basic nature of both 23 and 24. Due to the possible errors associated with calculating the free energy differences of deprotonation (due to the errors in both the DFT calculations and uncertainties of the free energy of the solvated proton

as discussed in the methods section), and the extreme sensitivity of predicted pK_a values to free energy values, our predicted pK_a value may have errors of 2–3 pK_a units. While M06-2X/cc-pVTZ predicted the acidic form of molecule 23 to have a pK_a of 9.2 and the acidic form of molecule 24 to have a pK_a 10.1, we can only conclude that both acid and conjugate are likely present at a pH of 7. Thus, our predicted free energies for 23 and 24 are not negative enough, as they do not include contributions from their protonized states 23h and 24h. If these states are present, their contributions would be large and negative. This result is consistent with our pK_a calculations for molecule 22, which predict the conjugate acid of 22 to have a $pK_a \leq -3$. Thus, the contribution from 22h should be negligible. Quantitatively characterizing this effect on computed solvation free energies will be the subject of a future study. Despite the persistent errors of 1 kcal/mol for molecules 23 and 24, we feel that the SMD-NBB method is quite robust, successfully combining the accurate potential energy calculations of QM with the entropic contributions from MM. We look forward to studying this methods ability to accurately model a wider range of ligands in the near future.

Conclusions

The results demonstrate that it is possible to improve free energy simulations by using post-processing with QM/MM. When using explicit solvent trajectories, the RMSD drops from 2.3 kcal/mol (MM-TIP3P in Table 1) to 1.6 kcal/mol (QM/MM-NBB). This improvement is actually larger than the one seen from implicit to explicit solvent (RMSD of 2.8 and 2.3 kcal/mol), at least for MD based free energy simulations. Given that the solvent is still treated classically in our QM/MM potential energy evaluations, this is quite an impressive increase of accuracy, especially considering that the quantum-chemical region only encompasses the solute and thus only accounts for the solute polarization.

Table 4 Evaluation of the effect of conformational sampling for some of the most pathological molecules

Molecule	Explicit solvent	Single point		SMD-NBB		Exp.
	QM/MM-NBB	B3LYP	M06-2X	B3LYP _{rw}	M06-2X _{rw}	
1	−21.5 ± 0.7	−25.0	−25.9	−22.4 ± 1.4	−23.6 ± 1.4	−23.6
22	−3.5 ± 2.1	−4.7	−6.3	−6.2 ± 0.4	−7.5 ± 0.4	−6.8
23	−5.9 ± 0.5	−4.3	−4.7	−6.3 ± 0.7	−7.8 ± 0.7	−9.3
24	−5.5 ± 0.7	−3.6	−4.6	−4.1 ± 0.1	−5.7 ± 0.3	−7.4
RMSD	4.3	3.4	3.0	2.3	1.2	

Explicit solvent results (left) are compared with hydration free energies from single point calculations with implicit solvent (middle) reweighted trajectories using implicit solvent (SMD-NBB, right). All methods are using the 6-31G(d) basis set and results are given in kcal/mol

Admittedly, this improved performance comes at a very high price. The computational costs of computing QM/MM energies are orders of magnitude higher than computing MM gradients (calculating a single QM/MM energy requires minutes, while a MM energy and gradient takes a fraction of a second). The total computational time, if run on a single processor of an Intel Xeon E5-2630 2.30GHz machine, is ~ 1 computer year for MM-TIP3P, ~ 25 computer years for QM/MM-TP and ~ 50 computer years for QM/MM-NBB. The high CPU costs of the QM/MM calculations is balanced by their embarrassingly parallel nature. Thus, we were able to perform all QM/MM calculations in less than a month of real time on the Biowulf and LoBoS computer clusters at the NIH. With the proliferation of multi-core computing, the clock time to perform QM-NBB free energy calculations will continue to decrease, while the required time to perform MD simulations is not likely to decrease significantly due to their inherently serial nature. Even if more computationally demanding QM methods are employed in the future, QM and QM/MM post-processing will remain attractive, because it can easily take advantage of continued progress in computer hardware development, while most other simulation techniques struggle to do so.

The attempt to predict unreliable results based on standard deviations, force field penalty scores and phase space overlap can only be considered marginally successful. All of our reliability metrics weakly correlate with our predictions' errors ($R < 0.5$). This supports the view that errors in free energy calculations have many sources, and there is no simple measure to test them all easily. Given the complexity of free energy simulations, most likely there is nothing that can absolve the user from judging every prediction by itself based on experience and chemical intuition.

The quantum-chemical implicit solvent calculations that we conducted after the deadline exhibit RMSD between 1.0 and 1.7 kcal/mol. Notably, the M06-2X results with the cc-pVTZ basis set are on a par with the best prediction in SAMPL4 for the blind subset (submission #561 by OpenEye, which yielded an RMSD of 1.0 kcal/mol) [26]. Given that the results from both cc-pVTZ and #561 are based on a single conformation and, therefore, do not fully incorporate effects from conformational entropy, this is remarkably accurate. For example, the best submission for the blind subset based on explicit solvent and molecular dynamics simulations (#005 by the Mobley group) yielded an RMSD of 1.3 kcal/mol. The relative performance demonstrates the high sophistication of current implicit solvent models. For the four pathological molecules that required more scrutiny (1, 22, 23, 24), we were able to reduce the RMSD from 3.0 to 1.2 kcal/mol by using reweighting of the MM trajectories with NBB. To put this into perspective, if the single

conformation results for molecules 1, 22, 23, 24 of M06-2X/6-31G(d) in Table 3 are replaced by the more accurate results from SMD-NBB in Table 4, the RMSD drops from 1.15 to 0.89 kcal/mol (and further improvements can be expected for the other molecules). This outperforms the RMSD of the more accurate cc-pVTZ basis set.

Our results highlight that both QM and MM calculations can profit from each other if they are used in a synergistic way. On the one hand, the QM-NBB results show that MM free energy simulations improve by accounting for solute polarization through QM/MM calculations. On the other hand, the SMD-NBB data in Table 4 demonstrates that even the highly developed M06-2X functional in combination with SMD can yield better results by incorporating solute entropy through reweighted MD trajectories. Of course reweighting a whole trajectory requires more computational resources than a computation for a single conformation, but due to the high parallelization of the post-processing this does not necessarily have an impact on the clock time. The only factors for the clock time are the required time for a single point potential energy calculation and the available number of computer nodes. If the number of nodes is the same as the number of frames, the clock time is the same as for a single calculation. We are, therefore, optimistic that the QM-NBB reweighting procedure will deliver even better results in the near future.

Acknowledgments The authors would like to thank Tim Miller, Richard Venable and John Legato for technical assistance with the parallelization of the QM/MM calculations. The support by Yihan Shao was invaluable during the setup of the Q-Chem scripts and we also would like to thank Florentina Tofoleanu, Tim Miller and Juyong Lee for carefully reading and commenting on the manuscript as well as Stefan Boresch and Lee Woodcock for fruitful discussions on the optimal performance of NBB. This work was supported by the intramural research program of the National Heart, Lung and Blood Institute of the National Institutes of Health and utilized the high-performance computational capabilities of the LoBoS and Biowulf Linux clusters at the National Institutes of Health. (<http://www.lobos.nih.gov> and <http://biowulf.nih.gov>).

References

1. 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* 51:769–779
2. Guthrie JP (2009) A blind challenge for computational solvation free energies: introduction and overview. *J Phys Chem B* 113(14):4501–4507. doi:10.1021/jp806724u
3. Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ (2010) The SAMPL2 blind prediction challenge: introduction and overview. *J Comput Aid Mol Des* 24(4, SI):259–279. doi:10.1007/s10822-010-9350-8
4. Muddana HS, Varnado CD, Bielawski CW, Urbach AR, Isaacs L, Geballe MT, Gilson MK (2012) Blind prediction of host–guest

- binding affinities: a new SAMPL3 challenge. *J Comput Aid Mol Des* 26(5):475–487. doi:[10.1007/s10822-012-9554-1](https://doi.org/10.1007/s10822-012-9554-1)
5. Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA (2009) Small molecule hydration free energies in explicit solvent: an extensive test of fixed-charge atomistic simulations. *J Chem Theory Comput* 5(2):350–358. doi:[10.1021/ct800409d](https://doi.org/10.1021/ct800409d)
 6. Klimovich PV, Mobley DL (2010) Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. *J Comput Aid Mol Des* 24(4, SI):307–316. doi:[10.1007/s10822-010-9343-7](https://doi.org/10.1007/s10822-010-9343-7)
 7. Mobley DL, Liu S, Cerutti DS, Swope WC, Rice JE (2012) Alchemical prediction of hydration free energies for SAMPL. *J Comput Aid Mol Des* 26(5, SI):551–562. doi:[10.1007/s10822-011-9528-8](https://doi.org/10.1007/s10822-011-9528-8)
 8. Jambeck JPM, Mocci F, Lyubartsev AP, Laaksonen A (2013) Partial atomic charges and their impact on the free energy of solvation. *J Comput Chem* 34(3):187–197. doi:[10.1002/jcc.23117](https://doi.org/10.1002/jcc.23117)
 9. Marenich AV, Cramer CJ, Truhlar DG (2009) Performance of SM6, SM8, and SMD on the SAMPL1 test set for the prediction of small-molecule solvation free energies. *J Phys Chem B* 113(14):4538–4543
 10. Ribeiro R, Marenich A, Cramer C, Truhlar D (2010) Prediction of sampl2 aqueous solvation free energies and tautomeric ratios using the sm8, sm8ad, and smd solvation models. *J Comput Aid Mol Des* 24(4):317–333
 11. Klamt A, Diederhofen M (2010) Blind prediction test of free energies of hydration with COSMO-RS. *J Comput Aid Mol Des* 24(4, SI):357–360. doi:[10.1007/s10822-010-9354-4](https://doi.org/10.1007/s10822-010-9354-4)
 12. Beierlein FR, Michel J, Essex JW (2011) A simple QM/MM approach for capturing polarization effects in protein-ligand binding free energy calculations. *J Phys Chem B* 115(17):4911–4926. doi:[10.1021/jp109054j](https://doi.org/10.1021/jp109054j)
 13. Fox SJ, Pittock C, Tautermann CS, Fox T, Christ C, Malcolm NOJ, Essex JW, Skylaris CK (2013) Free energies of binding from large-scale first-principles quantum mechanical calculations: application to ligand hydration energies. *J Phys Chem B* 117(32):9478–9485. doi:[10.1021/jp404518r](https://doi.org/10.1021/jp404518r)
 14. Rod TH, Ryde U (2005) Quantum mechanical free energy barrier for an enzymatic reaction. *Phys Rev Lett* 94(13):138–302. doi:[10.1103/PhysRevLett.94.138302](https://doi.org/10.1103/PhysRevLett.94.138302)
 15. Rod TH, Ryde U (2005) Accurate QM/MM free energy calculations of enzyme reactions: methylation by catechol O-methyltransferase. *J Chem Theory Comput* 1(6):1240–1251. doi:[10.1021/ct0501102](https://doi.org/10.1021/ct0501102)
 16. Heimdal J, Ryde U (2012) Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations. *Phys Chem Chem Phys* 14:12,592–12,604. doi:[10.1039/c2cp41005b](https://doi.org/10.1039/c2cp41005b)
 17. Min D, Zheng L, Harris W, Chen M, Lv C, Yang W (2010) Practically efficient QM/MM alchemical free energy simulations: the orthogonal space random walk strategy. *J Chem Theory Comput* 6(8):2253–2266. doi:[10.1021/ct100033s](https://doi.org/10.1021/ct100033s)
 18. Yang W, Cui Q, Min D, Li H (2010) Chapter 4—QM/MM alchemical free energy simulations: challenges and recent developments. *Annu Rep Comput Chem* 6:51–62. doi:[10.1016/S1574-1400\(10\)06004-4](https://doi.org/10.1016/S1574-1400(10)06004-4)
 19. Li H, Yang W (2007) Sampling enhancement for the quantum mechanical potential based molecular dynamics simulations: a general algorithm and its extension for free energy calculation on rugged energy surface. *J Chem Phys* 126(11) doi:[10.1063/1.2710790](https://doi.org/10.1063/1.2710790)
 20. König G, Boresch S (2011) Non-Boltzmann sampling and Bennett's acceptance ratio method: how to profit from bending the rules. *J Comput Chem* 32(6):1082–1090. doi:[10.1002/jcc.21687](https://doi.org/10.1002/jcc.21687)
 21. König G, Hudson P, Boresch S, Woodcock HL (2014) Multiscale free energy simulations: an efficient method for connecting classical MD simulations to QM or QM/MM free energies using Non-Boltzmann Bennett reweighting schemes. *J Chem Theory Comput* (in press)
 22. Bennett CH (1976) Efficient estimation of free energy differences from Monte Carlo data. *J Comp Phys* 22:245–268
 23. Mikulskis P, Cioloboc D, Andrejčić M, Khare S, Brorsson J, Genheden S, Mata RA, Söderhjelm P, Ryde U (2014) Free energy perturbation and quantum mechanical study of SAMPL4 octa-acid host-guest binding energies. *J Comput Aided Mol Des* (in press)
 24. Zwanzig RW (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J Chem Phys* 22:1420
 25. Guthrie JP (2014) SAMPL4, a blind challenge for computational solvation free energies: the compounds considered. *J Comput Aided Mol Des* (in press)
 26. Mobley DL, Wymer KL, Lim NM, Guthrie JP (2014) Blind prediction of solvation free energies from the SAMPL4 challenge. *J Comput Aided Mol Des*. doi:[10.1007/s10822-014-9718-2](https://doi.org/10.1007/s10822-014-9718-2)
 27. Brooks B, Brooks C III, Mackerell A Jr., Nilsson L, Petrella R, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caffisch A, Caves L, Cui Q, Dinner A, Feig M, Fischer S, Gao J, Hodosscek 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) CHARMM: the biomolecular simulation program. *J Comput Chem* 30(10, Sp. Iss. SI):1545–1614. doi:[10.1002/jcc.21287](https://doi.org/10.1002/jcc.21287)
 28. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J Comput Chem* 4:187–217
 29. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell Jr AD (2010) CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J Comp Chem* 31(4):671–690. doi:[10.1002/jcc.21367](https://doi.org/10.1002/jcc.21367)
 30. Sugita Y, Kitao A, Okamoto Y (2000) Multidimensional replica-exchange method for free-energy calculations. *J Chem Phys* 113:6042
 31. Darden T, York D, Pedersen L (1993) Particle mesh ewald—an n.log(n) method for ewald sums in large systems. *J Chem Phys* 98:10089–10092
 32. Lee MS, Feig M, Salsbury FR, Brooks CL III (2003) New analytic approximation to the standard molecular volume definition and its application to generalized born calculations. *J Comput Chem* 23:1348–1356
 33. 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, Keith, T, 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, AD, Farkas, O, Foresman, JB, Ortiz, JV, Cioslowski, J, Fox, DJ (2010) Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford, CT
 34. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98(7):5648–5652. doi:[10.1063/1.464913](https://doi.org/10.1063/1.464913)
 35. Zhao Y, Truhlar DG (2007) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics,

- noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor Chem Acc* 120:215–241
36. Zhao Y, Truhlar DG (2008) Density functionals with broad applicability in chemistry. *Acc Chem Res* 41:157–167
 37. Hariharan PC, Pople JA (1974) Accuracy of ah n equilibrium geometries by single determinant molecular orbital theory. *Mol Phys* 27(1):209–214. doi:[10.1080/00268977400100171](https://doi.org/10.1080/00268977400100171)
 38. Dunning TH Jr (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J Chem Phys* 90(2):1007–1023
 39. Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. *Chem Rev* 105(8):2999–3094. doi:[10.1021/cr9904009](https://doi.org/10.1021/cr9904009)
 40. Marenich AV, Cramer CJ, Truhlar DG (2009) Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J Phys Chem B* 113(18):6378–6396
 41. Liptak M, Shields G (2001) Accurate pK(a) calculations for carboxylic acids using complete basis set and gaussian-n models combined with CPCM continuum solvation methods. *J Am Chem Soc* 123(30):7314–7319. doi:[10.1021/ja010534f](https://doi.org/10.1021/ja010534f)
 42. Shao Y, Molnar LF, Jung Y, Kussmann J, Ochsenfeld C, Brown ST, Gilbert ATB, Slipchenko LV, Levchenko SV, O'Neill DP, DiStasio Jr RA, Lochan RC, Wang T, Beran GJO, Besley NA, Herbert JM, Lin CY, Van Voorhis T, Chien SH, Sodt A, Steele RP, Rassolov VA, Maslen PE, Korambath PP, Adamson RD, Austin B, Baker J, Byrd EFC, Dachsel H, Doerksen RJ, Dreuw A, Dunietz BD, Dutoi AD, Furlani TR, Gwaltney SR, Heyden A, Hirata S, Hsu CP, Kedziora G, Khalliulin RZ, Klunzinger P, Lee AM, Lee MS, Liang W, Lotan I, Nair N, Peters B, Proynov EI, Pieniazek PA, Rhee YM, Ritchie, J, Rosta, E, Sherrill, CD, Simmonett, AC, Subotnik JE, Woodcock III HL, Zhang W, Bell AT, Chakraborty AK, Chipman DM, Keil FJ, Warshel A, Hehre WJ, Schaefer III HF, Kong J, Krylov AI, Gill PMW, Head-Gordon M (2006) Advances in methods and algorithms in a modern quantum chemistry program package. *Phys Chem Chem Phys* 8(27):3172–3191 doi:[10.1039/b517914a](https://doi.org/10.1039/b517914a)
 43. Woodcock HL III, Hodoscek M, Gilbert ATB, Gill PMW, Schaefer HF III, Brooks BR (2007) Interfacing Q-chem and CHARMM to perform QM/MM reaction path calculations. *J Comp Chem* 28(9):1485–1502. doi:[10.1002/jcc.20587](https://doi.org/10.1002/jcc.20587)
 44. 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. doi:[10.1063/1.445869](https://doi.org/10.1063/1.445869)
 45. 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. doi:[10.1021/jm070549+](https://doi.org/10.1021/jm070549+)
 46. Geballe MT, Guthrie JP (2012) The SAMPL3 blind prediction challenge: transfer energy overview. *J Comput Aid Mol Des* 26(5, SI):489–496. doi:[10.1007/s10822-012-9568-8](https://doi.org/10.1007/s10822-012-9568-8)
 47. Beckstein O, Iorga BI (2012) Prediction of hydration free energies for aliphatic and aromatic chloro derivatives using molecular dynamics simulations with the OPLS-AA force field. *J Comput Aid Mol Des* 26(5, SI):635–645. doi:[10.1007/s10822-011-9527-9](https://doi.org/10.1007/s10822-011-9527-9)
 48. Reinisch J, Klamt A, Diedenhofen M (2012) Prediction of free energies of hydration with COSMO-RS on the SAMPL3 data set. *J Comput Aid Mol Des* 26(5, SI):669–673. doi:[10.1007/s10822-012-9576-8](https://doi.org/10.1007/s10822-012-9576-8)
 49. Kehoe CW, Fennell CJ, Dill KA (2012) Testing the semi-explicit assembly solvation model in the SAMPL3 community blind test. *J Comput Aid Mol Des* 26(5, SI):563–568. doi:[10.1007/s10822-011-9536-8](https://doi.org/10.1007/s10822-011-9536-8)
 50. König G, Bruckner S, Boresch S (2013) Absolute hydration free energies of blocked amino acids: implications for protein solvation and stability. *Biophys J* 104(2):453–462. doi:[10.1016/j.bpj.2012.12.008](https://doi.org/10.1016/j.bpj.2012.12.008)
 51. König G, Boresch S (2009) Hydration free energies of amino acids: why side chain analog data are not enough. *J Phys Chem B* 113(26):8967–8974. doi:[10.1021/jp902638y](https://doi.org/10.1021/jp902638y)
 52. König G, Brooks BR (2012) Predicting binding affinities of host-guest systems in the SAMPL3 blind challenge: the performance of relative free energy calculations. *J Comput Aid Mol Des* 26(5):543–550. doi:[10.1007/s10822-011-9525-y](https://doi.org/10.1007/s10822-011-9525-y)
 53. Vanommeslaeghe K, Raman EP, MacKerell AD Jr (2012) Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J Chem Inf Model* 52(12):3155–3168. doi:[10.1021/ci3003649](https://doi.org/10.1021/ci3003649)