

A Monte Carlo Resampling Approach for the Calculation of Hybrid Classical and Quantum Free Energies

Christopher Cave-Ayland, Chris-Kriton Skylaris, and Jonathan W. Essex*

School of Chemistry, University of Southampton, Hampshire, SO17 1BJ, United Kingdom

E-mail: J.W.Essex@soton.ac.uk

Abstract

Hybrid free energy methods allow estimation of free energy differences at the quantum mechanics (QM) level with high efficiency by performing sampling at the classical mechanics (MM) level. Various approaches to allow the calculation of QM corrections to classical free energies have been proposed. The single step free energy perturbation approach starts with a classically generated ensemble, a subset of structures of which are post-processed to obtain QM energies for use with the Zwanzig equation. This gives an estimate of the free energy difference associated with the change from an MM to a QM Hamiltonian. Owing to the poor numerical properties of the Zwanzig equation however, recent developments have produced alternative methods which aim to provide access to the properties of the true QM ensemble. Here we propose an approach based on the resampling of MM structural ensembles and application of a Monte Carlo acceptance test which in principle, can generate the exact QM ensemble or intermediate ensembles between the MM and QM states. We carry out a detailed comparison against the Zwanzig equation and recently proposed non-Boltzmann methods. As a test system we use a set of small molecule hydration free energies for which

hybrid free energy calculations are performed at the semi-empirical Density Functional Tight Binding level. Equivalent ensembles at this level of theory have also been generated allowing the reverse QM to MM perturbations to be performed along with a detailed analysis of the results. Additionally, a previously published nucleotide base pair data set simulated at the QM level using *ab initio* molecular dynamics is also considered. We provide a strong rationale for the use of the Monte Carlo Resampling and non-Boltzmann approaches by showing that configuration space overlaps can be estimated which provide useful diagnostic information regarding the accuracy of these hybrid approaches.

1 Introduction

The prediction of protein-ligand binding affinities and free energies of hydration remains a grand challenge of computational chemistry. Of the wide range of techniques available for these purposes, the most rigorous are based on the theoretical application of statistical mechanics, such as free energy perturbation approaches. A notable drawback of free energy perturbation is the requirement for extensive sampling of relevant thermodynamic states with Monte Carlo (MC) or molecular dynamics (MD). This places a practical limitation on the accuracy of the energy models that can be used to describe the states of interest. Typically this means the use of approximate force field models (often known as molecular mechanics or MM) that omit explicit treatment of electronic structure. Whilst in many cases the MM level of theory can be surprisingly accurate (particularly when exploiting cancellation of errors within relative free energies) there are numerous examples in the literature highlighting its insufficiency in binding affinity predictions¹⁻⁵.

Use of first principles quantum mechanics (QM) methods with free energy techniques would be highly desirable to achieve greater accuracy. However, the computational cost of these is several orders of magnitude greater than for MM. This led to the proposal, that QM correc-

tions to free energy differences at the MM level could be efficiently calculated. Use of these end-point corrections (see figure 1) rely on successful calculation of free energy differences between MM and QM descriptions (thermodynamic states) of the same molecules. This approach has been successfully applied to free energies of hydration⁶⁻⁹; however, calculations achieving well converged results in binding affinity predictions have been considerably more challenging¹⁰⁻¹³.

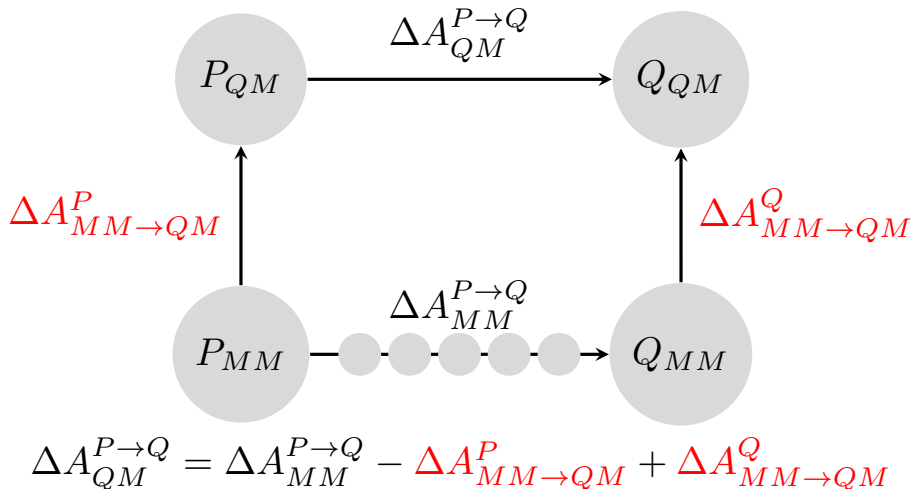


Figure 1: Thermodynamic cycle for correction of free energy difference between system P and Q at the MM to the QM level. The free energy difference between P and Q at the QM level of theory is given by $\Delta A_{QM}^{P \rightarrow Q}$. This can be given as a combination of the rest of the terms in the cycle. The free energy difference at the MM level $\Delta A_{MM}^{P \rightarrow Q}$, can be computed by subdivision into steps along an alchemical or other reaction coordinate. The correction terms $\Delta A_{MM \rightarrow QM}^P$ and $\Delta A_{MM \rightarrow QM}^Q$ give the free energy difference of moving from the MM to the QM level of theory for each state.

Recent developments have provided more sophisticated methods for the calculation of hybrid free energies. Most notably these include the non-Boltzmann approaches of Brooks et al.¹⁴⁻¹⁷, a non-equilibrium work approach from Boresch et al.¹⁸, a multi-scale MC approach from Mulholland et al.¹⁹ and the “stepping stone” approach of Skylaris et al.²⁰.

In previous work, we have considered hybrid free energy calculations using the Zwanzig equation²¹. This study was carried out on a simple test system, an Adenine-Thymine DNA base pair in vacuum. This system allowed us to demonstrate, for the first time convergence

of MM to QM free energy differences by carrying out the reverse QM to MM perturbations. This required the generation of extensive ensembles at the QM level of theory with *ab initio* MD simulations.

In this work we develop and apply alternative approaches to the Zwanzig equation for MM to QM free energy estimation. We present a new technique for the calculation of QM corrections to classical free energy differences, based on resampling an MM ensemble and application of a Metropolis criterion. We demonstrate application of this approach to a set of small molecule hydration free energies using a semi-empirical model, as well as to the DNA base pair system previously discussed. We use this as the basis for a detailed comparison of several state-of-the-art approaches for hybrid free energy calculations, focusing on the comparative accuracy of the techniques, and whether useful diagnostics may be derived to allow for the accuracy of the calculated free energies to be determined.

2 Methods Overview

2.1 Single Step Free Energy Perturbation (SSFEP)

The Zwanzig equation²² was one of the earliest established theoretically rigorous results for the calculation of free energy differences. It can be used to efficiently estimate hybrid free energies as:

$$\Delta A_{MM \rightarrow QM} = -\frac{1}{\beta} \ln \langle \exp[-\beta(U_{QM} - U_{MM})] \rangle_{MM} \quad (1)$$

where U_{QM} and U_{MM} give energies at the QM and MM level of theory and $\langle \dots \rangle_{MM}$ denotes an ensemble average over a simulation conducted at the MM level. This takes advantage of the requirement of the Zwanzig equation to only evaluate a single ensemble average. The MM state is sampled using standard techniques and a subset of configurations are

post-processed with single point energy calculations at the QM level. The MM potential efficiently generates significantly different snapshots and the Zwanzig equation then corrects for the difference between the potentials. This approach avoids the need to generate a QM ensemble and is exact at the limit of infinite sampling with the MM protocol. However, the exponential weighting of configurations within (1) can lead to poor convergence properties for such calculations^{23,24}.

2.2 Monte Carlo Resampling (RSM)

The direct generation of a QM ensemble is very computationally expensive requiring sampling with molecular dynamics or Monte Carlo under the QM Hamiltonian. For this reason an alternative approach has been proposed through the use of multi-stepping Monte Carlo¹⁹ or hybrid molecular dynamics²⁰. Briefly, these techniques perform sampling at the MM level with periodic QM calculations and application of an acceptance test. Those configurations that pass the test are added to the QM ensemble, thereby building a replica of the true QM ensemble. Sampling in this way considerably reduces the cost of generating a QM ensemble, but has the drawback that all QM calculations must be carried out sequentially i.e. as the MM simulation is performed. This restricts these techniques from effectively exploiting massively parallel computing architectures. We present here an algorithm to generate a correct QM ensemble (in the limit of infinite sampling) solely through post-processing of an MM ensemble, allowing the necessary QM calculations to be carried out concurrently for completed MM trajectories. Starting from an already generated MM ensemble the algorithm is as follows:

1. Select a configuration, c_1 , with replacement, from the MM ensemble with equal probability and add this to the QM ensemble.
2. Select a trial configuration, c_2 , with replacement, from the MM ensemble with equal probability.

3. Apply the following acceptance test in an MC procedure:

$$P_{accpt}^{resamp}(c1 \rightarrow c2) = \min(1, \exp[-\beta\Delta\Delta U]) \quad (2)$$

$$\Delta\Delta U = (U_{QM}(c2) - U_{MM}(c2)) - (U_{QM}(c1) - U_{MM}(c1)).$$

4. If test passed, c2 is added to the QM ensemble and becomes c1. Repeat from (2).

5. If test failed, another copy of c1 is added to the QM ensemble. Repeat from (2).

A demonstration of the correctness of this algorithm in an idealized test system is provided in the supporting information.

2.2.1 Acceptance Test Derivation

The general form of the Metropolis-Hastings test is given as:

$$P_{accpt}^{resamp}(c1 \rightarrow c2) = \min \left(1, \frac{P_{QM}(c2)P_g(c2 \rightarrow c1)}{P_{QM}(c1)P_g(c1 \rightarrow c2)} \right) \quad (3)$$

Where P_{QM} is the probability of a particular configuration arising within the QM ensemble we are trying to construct, and $P_g(c1 \rightarrow c2)$ is the probability of generating c2 as a trial move. P_{QM} is known straightforwardly from the Boltzmann distribution i.e. $P_{QM}(c1) = \exp[-\beta U_{QM}(c1)]/Z_{QM}$, whilst P_g can also be intuitively derived. From the above algorithm the probability of proposing c2 is the product of choosing c2 out of the collection of MM configurations and the probability of that configuration being generated within the MM ensemble. Therefore:

$$P_g(c1 \rightarrow c2) = \frac{1}{N_{MM}} \frac{\exp[-\beta U_{MM}(c2)]}{Z_{MM}} \quad (4)$$

Where N_{MM} is the number of structures within the sampled MM ensemble We note from

this result that the selection of c_2 is independent of c_1 . Substituting (4) and P_{QM} into (3) naturally gives (2).

2.3 Non-Boltzmann Reweighting (NBR)

This approach is based around application of the following result, typically associated with umbrella sampling.

$$\langle X \rangle_u = \frac{\langle X \exp[\beta V] \rangle_b}{\langle \exp[\beta V] \rangle_b} \quad (5)$$

here $\langle \dots \rangle_u$ represents an ensemble average on an unbiased system, and b represents a biased system. V is the bias potential by which the biased and unbiased states differ and X is some property of interest. Most commonly V is chosen as a harmonic potential and X is chosen as a potential of mean force. More generally however this result can be seen as recovering an ensemble average of property X from sampling under a different Hamiltonian. Sampling of the biased state is sometimes referred to as non-Boltzmann sampling of the unbiased state, giving rise to the name of this technique. Maintaining the formalism of biases, V can be defined as:

$$V = U_{MM} - U_{QM} \quad (6)$$

i.e. the difference between an MM and QM potential. This definition allows the recovery of QM ensemble average properties from sampling under an MM Hamiltonian:

$$\langle X \rangle_{QM} = \frac{\langle X \exp[\beta V] \rangle_{MM}}{\langle \exp[\beta V] \rangle_{MM}} \quad (7)$$

this can be profitably applied to the calculation of free energies, starting with the Zwanzig

equation:

$$\Delta A_{QM,0 \rightarrow 1} = -\frac{1}{\beta} \ln \langle \exp[-\beta(U_{QM,1} - U_{QM,0})] \rangle_{QM,0} \quad (8)$$

Here we are calculating a free energy difference between two separate thermodynamic states, 0 and 1, both at the QM level of theory. The QM energies for the different states are denoted by $U_{QM,0}$ and $U_{QM,1}$. Combining (7) with (8) gives the non-Boltzmann Zwanzig equation¹⁵:

$$\Delta A_{QM,0 \rightarrow 1} = -\frac{1}{\beta} \ln \frac{\langle \exp[-\beta(U_{QM,1} - U_{QM,0})] \exp[\beta V_0] \rangle_{MM,0}}{\langle \exp[\beta V_0] \rangle_{MM,0}} \quad (9)$$

where $V_0 = U_{MM,0} - U_{QM,0}$. Such a development may equally be applied to the BAR estimator, to give the non-Boltzmann BAR result¹⁵:

$$\Delta A_{QM,0 \rightarrow 1} = \frac{1}{\beta} \ln \left(\frac{\langle F(\beta(U_{QM,0} - U_{QM,1} + C)) \exp[\beta V_1] \rangle_{MM,1} \langle \exp[\beta V_0] \rangle_{MM,0}}{\langle F(\beta(U_{QM,1} - U_{QM,0} - C)) \exp[\beta V_0] \rangle_{MM,0} \langle \exp[\beta V_1] \rangle_{MM,1}} \right) + C \quad (10)$$

$$C = \frac{1}{\beta} \ln \frac{n_1}{n_0} - \Delta A_{QM,0 \rightarrow 1} \quad (11)$$

Where F is the Fermi function, n_1 and n_0 are the number of configurations from states 1 and 0 respectively. As with the Zwanzig equation this result provides a QM free energy difference without sampling the QM ensemble. Equations (9) and (10) suggest an alternative to the use of end point corrections and the thermodynamic cycle in figure 1. In practice this is highly inefficient, likely requiring subdivision of the $\Delta A_{QM,0 \rightarrow 1}$ perturbation and completion of additional QM computations for the intermediate states. The non-Boltzmann BAR result

can be adapted for use with end-point corrections however:

$$\Delta A_{MM \rightarrow QM, \lambda \rightarrow 0} = \frac{1}{\beta} \ln \left(\frac{\langle F(\beta(U_{MM, \lambda} - U_{QM, 0} + C)) \exp[\beta V_0] \rangle_{MM, 0}}{\langle F(\beta(U_{QM, 0} - U_{MM, \lambda} - C)) \rangle_{MM, \lambda} \langle \exp[\beta V_0] \rangle_{MM, 0}} \right) + C \quad (12)$$

$$C = \frac{1}{\beta} \ln \frac{n_0}{n_\lambda} - \Delta A_{MM \rightarrow QM, \lambda \rightarrow 0} \quad (13)$$

Where the subscript λ is used to denote an intermediate MM state between 0 and 1. This can be chosen to be arbitrarily similar to 0, however if equal to 0, (12) and (13) reduce to (1). Applying this result at both ends of a classical perturbation gives:

$$\Delta A_{QM, 0 \rightarrow 1} = \Delta A_{MM, \lambda_1 \rightarrow \lambda_2} - \Delta A_{MM \rightarrow QM, \lambda_1 \rightarrow 0} + \Delta A_{MM \rightarrow QM, \lambda_2 \rightarrow 1} \quad (14)$$

Usage of this scheme requires the generation, at the MM level, and post-processing, at the QM level, of two states at different values of lambda. Of these, the MM calculations at the end-points of λ are reweighted to provide a QM ensemble average. This is then used in the BAR estimator to give the free energy difference with an MM state at an intermediate λ value.

2.4 Configuration Space Overlap Calculations

Following the quantity given by Bennett²⁵, we define the overlap between two thermodynamic states as:

$$O_{0 \rightarrow 1} = \int \frac{2\rho_0\rho_1}{\rho_0 + \rho_1} d\mathbf{q}^N \quad (15)$$

where ρ_0 and ρ_1 are the normalized probability densities of the N system degrees of freedom \mathbf{q} for the two states respectively. That is, $\rho_0(\mathbf{q}^N) = \exp[-\beta U_0(\mathbf{q}^N)]/Z_0$, where Z has the usual definition of the configuration integral. It should be noted that under this definition,

the measure of overlap is invariant with respect to which state is chosen as 0 and which is 1, i.e. $O_{0 \rightarrow 1} = O_{1 \rightarrow 0}$.

Owing to the high dimensionality of the integral, (15) is impossible to evaluate directly for realistically sized biological systems. In the following we demonstrate how (15) can be expressed in terms of ensemble averages from the two states of interest. We begin by noting some results also provided by Bennett²⁵. Starting with the definition of the variance for BAR, σ^2 :

$$\sigma^2 = \frac{\langle F^2 \rangle_0 - \langle F \rangle_0^2}{n_0 \langle F \rangle_0^2} + \frac{\langle F^2 \rangle_1 - \langle F \rangle_1^2}{n_1 \langle F \rangle_1^2} \quad (16)$$

$$= \left(\int \frac{2n_0 n_1 \rho_0 \rho_1}{n_0 \rho_0 + n_1 \rho_1} d\mathbf{q}^N \right)^{-1} - \frac{n_0 + n_1}{n_0 n_1} \quad (17)$$

where for clarity, arguments to the Fermi function have been omitted but correspond to those for the relevant ensemble averages as in the BAR equations. Here C takes the value given from self-consistent solution. Equation (16) differs from Bennett's²⁵ due to a typo in the original publication that erroneously gave the subscripts of the denominator of the second term as 0 instead of 1. Bennett also notes that for some value, \bar{n} , lying between n_0 and n_1 :

$$\sigma^2 = \frac{2}{\bar{n}} \left[\left(\int \frac{2\rho_0 \rho_1}{\rho_0 + \rho_1} d\mathbf{q}^N \right)^{-1} - 1 \right]. \quad (18)$$

The following result is based on use of the simplifying assumption that $n_0 = n_1$ ensuring the corollary that $\langle F \rangle_0 = \langle F \rangle_1$ (see Supporting Information). This allows the use of (18) and results in an exact expression for the overlap. This assumption is defensible as free choice of these values is rarely prevented. A derivation lacking this assumption is presented in the Supporting Information although in such a case the result is limited to providing an upper and lower bound for the overlap. Using (16) and (18) and providing the full arguments for

the Fermi functions:

$$O_{0 \rightarrow 1} = \frac{2 \langle F(\beta(U_0 - U_1 - C)) \rangle_0^2}{\langle F(\beta(U_0 - U_1 - C))^2 \rangle_0 + \langle F(\beta(U_1 - U_0 + C))^2 \rangle_1} \quad (19)$$

Using (19) we are able to calculate the overlap between two thermodynamic states where we have the corresponding ensemble averages. This technique may also be combined with an ensemble generated using NBR or RSM however. In such a case, the overlap between an MM and a QM state may be estimated without the need to directly sample the QM state.

2.5 Interaction Energies

As established in our previous work^{11,21} the use of interaction energies in the place of total energies is required to achieve convergence with the use of free energy techniques. This approach has also been used by other groups^{8,16}. In the case of the absolute hydration free energy calculations carried out in this work, the use of interaction energies also provides the benefit that only the solvent phase perturbation end point need be corrected to the QM level. The correction for the vacuum perturbation end point will necessarily be zero as there are no interaction energies in the gas phase.

2.6 Simulation Protocol

To explore the free energy calculation methods, simulations on two sets of systems were carried out. In the first, absolute hydration free energy calculations were performed for a set of 25 small molecules (listed in figure S5) using the AMBER 16 simulation package²⁶. Both MM and SQM/MM ensembles were generated using molecular dynamics simulations. Initial structures were solvated in TIP3P water²⁷ such that a minimum distance of 12 Å between the solute and the system edge was maintained. A cubic simulation cell was used

with periodic boundary conditions. Water molecules were kept rigid whilst solutes were fully flexible. A cutoff distance of 10 Å was used for non-bonded interactions whilst long range electrostatics were treated with Particle Mesh Ewald. All simulations were carried out at 300 K using a Langevin thermostat with a collision frequency of 3.0 ps⁻¹. Where relevant, the system pressure was regulated using a Monte Carlo barostat with volume moves attempted once every 100 time steps.

For the classical free energy of hydration calculations solutes were described using the GAFF force field²⁸ with AM1/bcc partial charges^{29,30}. Systems were first equilibrated for 1 ns in the NPT ensemble. Production simulations were then run in the NVT ensemble for 5 ns. Configurations for post processing were captured every 5 ps. Absolute hydration free energies were calculated by perturbing solutes from solvent into the gas phase. Lennard-Jones and Coulombic terms were scaled simultaneously using soft core potentials with the default functional forms and parameters provided in AMBER 16. Thermodynamic integration with the trapezium rule was employed to calculate free energy differences over 12 evenly interspaced λ values.

Density Function Tight Binding^{31,32} (DFTB) ensembles were generated using the implementation available in the SANDER engine³³. An SQM/MM treatment¹³ was used such that the solute was simulated at the DFTB level of theory, electrostatically embedded among classical TIP3P solvent molecules. DFTB simulations were started with the structures resulting from the classical equilibration runs. DFTB equilibration was carried out for 1 ns in the NVT ensemble followed by 5 ns production runs. Again, configurations for post processing were recorded every 5 ps. Post processing calculations were carried out with aid of the GNU parallel tool³⁴. The 3ob parameter set^{35,36}, was used for all DFTB calculations.

The second set of simulations were performed on a nucleotide base pair, and have been previously described elsewhere. We present here a brief summary of the DNA data set; see ref 21 for full details. Ensembles were generated directly using molecular dynamics with four

different Hamiltonians - the GAFF and ff99SB force fields, along with DFT using both the LDA and PBE functionals - with a total of around 100 ps of trajectory data. SSFEP was used to calculate all possible transformations between these four states giving twelve individual free energy differences due to the directionality of the Zwanzig equation. Convergence of free energy differences was considered by comparing the forward and reverse calculations for each pair of Hamiltonians and was found to be generally good. In this work, we re-use this data set to explore the RSM and NBR free energy estimators.

3 Results and Discussion

The purpose of this work is to explore the relative efficiencies of the SSFEP, NBR, and RSM methods to calculate end-point MM to QM corrections. Each of these methods requires only an MM ensemble to be explicitly generated, but RSM and NBR use this ensemble to build a QM ensemble which is subsequently used in the free energy calculation. SSFEP does not build a QM ensemble and therefore requires fewer QM calculations, and is hence less computationally demanding. To validate our results, we have independently generated a QM ensemble using DFTB (for the hydration free energies) and DFT (for the nucleotide base pair). This allows us to calculate the MM to QM free energy directly using BAR, and we consider this to be the reference free energy against which our other, more approximate, methods are judged. In a real-world application, one would like to avoid explicitly simulating with the QM Hamiltonian for reasons of computational cost. We have therefore also sought diagnostics which can be used to indicate under what circumstances SSFEP, NBR and RSM become unreliable.

As noted previously, any numerical advantages of applying the non-Boltzmann and MC resampling techniques over SSFEP are not obvious *a priori*. The ability to make use of the more statistically efficient BAR estimator is in theory advantageous. This requires evaluation

of double the number of QM energies however due to the requirement to post process an additional MM ensemble, whilst introducing an additional reweighting or ensemble building step potentially provides another source of error.

To explore the behavior of the different techniques we make use of a set of hydration free energy calculations corrected to a semi-empirical DFTB Hamiltonian. For each solute, an absolute free energy of hydration calculation was performed at the MM level. Additionally a solvent phase DFTB ensemble was generated for each compound, a more computationally demanding step that would not typically be carried out for the calculation of hybrid free energies. With the direct DFTB ensemble available we are able to use BAR to provide high quality estimates of the free energy difference between the MM and SQM/MM Hamiltonians giving reliable reference data for comparison of SSFEP, NBR and RSM, which require only MM simulation ensembles. In principle, the NBR and RSM techniques can be applied in a number of ways to such a data set, however we will here focus on their use in the calculation of endpoint corrections as shown in figure 2a. As validation of the employed simulation protocols, the purely classical hydration free energies compared with experiment are presented in figure S4.

The majority of our observations and conclusions are drawn from the above calculations: however, we also make use of the DNA base pair data set developed and validated in previous work²¹. This provides trajectories for a nucleotide base pair in vacuum with two different classical Hamiltonians and at the DFT level with the LDA and PBE functionals. We present the use of NBR and RSM to this data set in a form analogous to the hydration free energy data.

Validation of the employed data sets is considered through the calculation of closures for relevant thermodynamic cycles (see figures S5 and S6). This allows us to assess the quality of the reference data against which SSFEP, NBR and RSM are compared. For the hydration data a cycle can be constructed for each solute between the DFTB Hamiltonian and the MM

Hamiltonian at the end point λ values, denoted λ_1 and λ_2 . In the base pair data, multiple three and four state cycles can be constructed between the four different Hamiltonians. The mean unsigned closures are 0.17 and 0.54 kcal.mol⁻¹ respectively for these data sets. The poorer closures shown for the base pair data set are most likely due to the short trajectory lengths necessitated for the generation of ensembles at the DFT level. We will primarily focus on the hydration data therefore, but the base pair data are included to demonstrate the generality of the results with respect to different Hamiltonians and work in a poorly convergent regime.

Our interest here is to validate our proposed RSM approach by comparison with the performance of other hybrid free energy methodologies. We investigate the relative practical advantages of the different techniques and aim to establish clearly any differences in convergence or accuracy. Previous work has compared the performance of NBR and SSFEP against experimental free energy differences^{9,14,15}. Whilst recreating experimental data is perhaps ultimate goal of computational work, this is subject to confounding factors such as experimental error or deficiencies in the level of theory used at the QM state. This may be related to conflicting results regarding whether any significant benefits are offered by NBR^{15,37}. In some cases comparisons have also failed to make use of comparable protocols¹².

3.1 Accuracy

Comparison of the techniques can be achieved in this data set by considering the use of the different methods in deriving end point corrections to classical free energy differences. Comparative use of SSFEP, RSM and NBR is shown in figures 2a. Application of SSFEP is straightforward, configurations derived from the molecular dynamics simulation performed at $\lambda = 1.0$ are simply post processed using the Zwanzig equation. For RSM and NBR however configurations from the $\lambda = 1.0$ window are used derive the DFTB ensemble that is subsequently used with BAR to evaluate the end-point correction free energy.

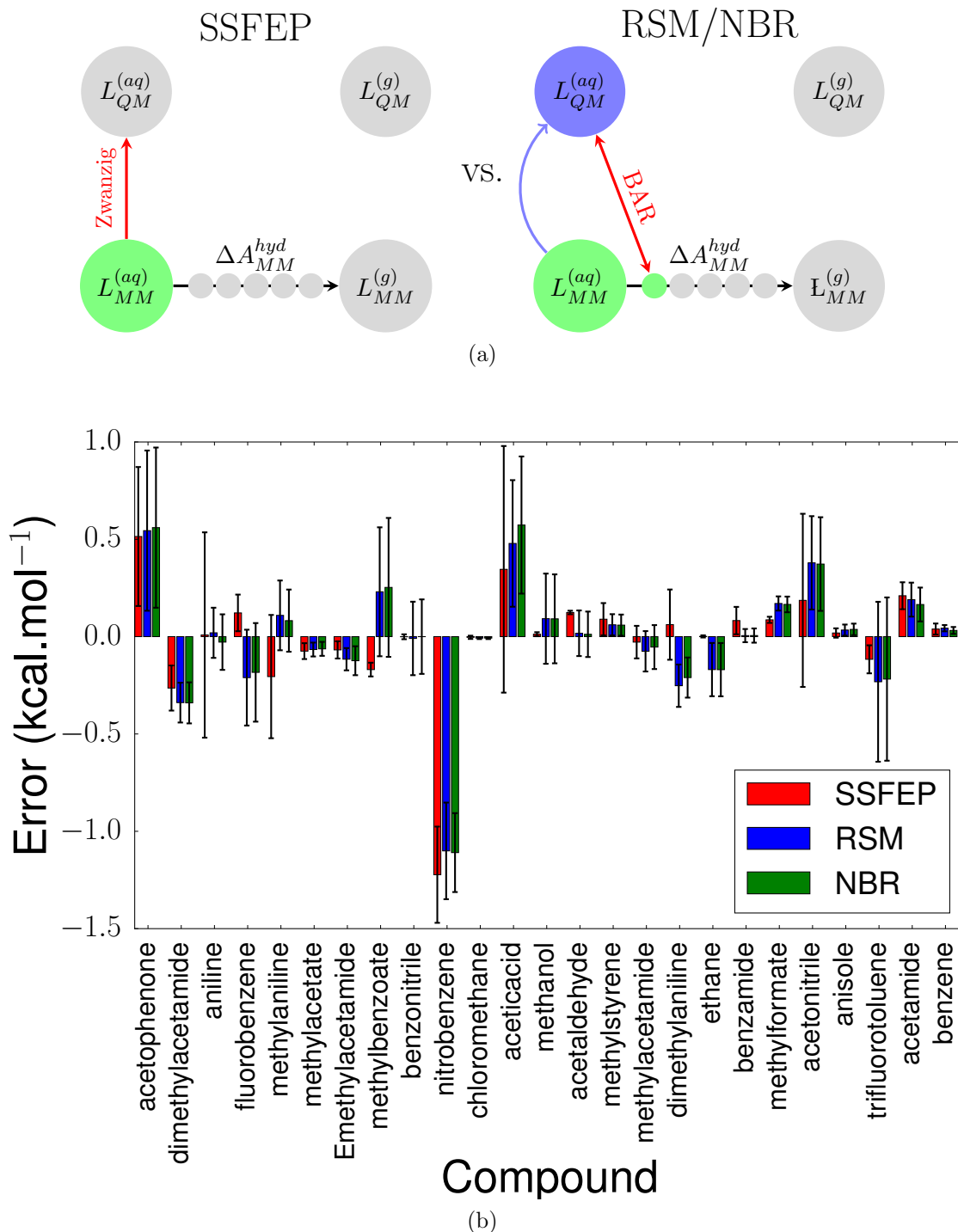


Figure 2: Accuracy comparison between techniques for DFTB free energy calculations. a) Application of end point corrections to hydration free energy scheme. Green denotes states that are sampled directly with MD and post-processed. The blue QM state is derived indirectly from the use of RSM or NBR. Red arrows indicate the relevant free energy differences calculated. b) Errors for individual perturbations with different techniques compared to reference BAR data. Error bars show one standard error, calculated by combining the variance of three independent repeats from both the reference BAR data and perturbations with each technique.

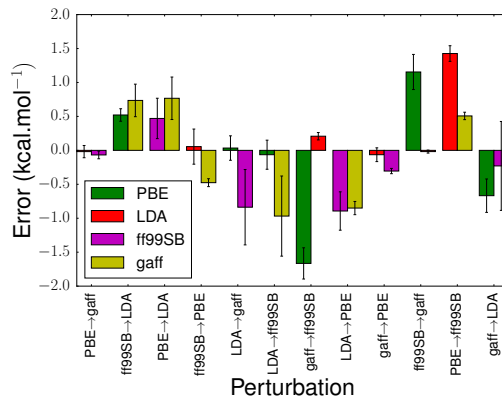
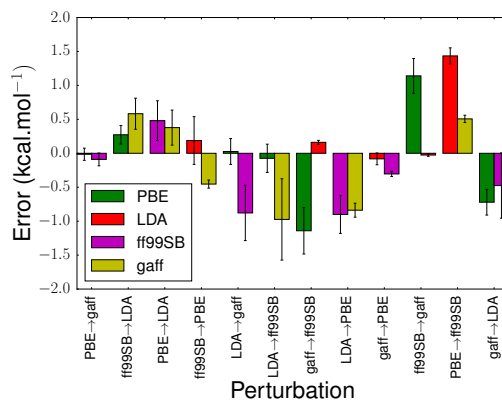
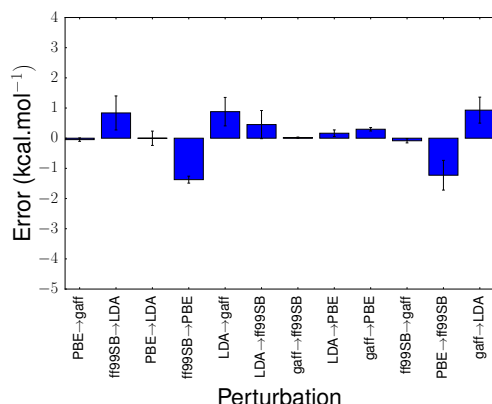
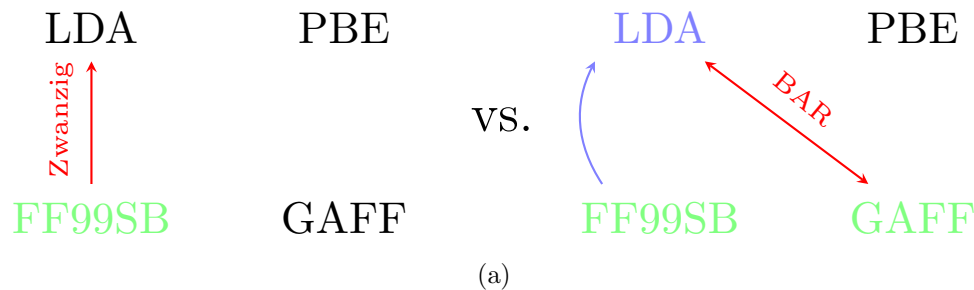


Figure 3: Accuracy comparison between techniques for DNA base pair. a) Analogous application of hybrid free energy techniques for the base pair data set. The meaning of colors is as described for figure 2. b-d) show the errors with respect to BAR associated with the techniques SSFEP, NBR and RSM respectively. For NBR and RSM the perturbation target state been produced indirectly with the source ensemble used given by the color. Error bars show one standard error, calculated by combining the variance of five independent repeats from both the reference BAR data and perturbations with each technique.

Errors in the calculated free energies of hydration for each compound derived using SSFEP, NBR and RSM with respect to direct application of BAR are given in figure 2. The overall mean unsigned error (MUE) over all compounds for the different techniques is 0.16, 0.20 and 0.23 kcal.mol⁻¹, for SSFEP, NBB and RSM respectively, indicating comparable performance. This is disappointing overall for RSM and NBR, as these techniques require twice the number of QM computations as SSFEP with little apparent gain in accuracy. The majority of compounds are close to one standard error away from zero suggesting a lack of systematic errors. The notable exception is nitrobenzene which is a significant outlier. Comparison of the GAFF and DFTB ensembles of this compound suggest that the significant error in this case is due to a difference in sampling of the nitro group dihedral angle with respect to the benzene ring (see figure S7).

The results for the base pair data set are provided in figure 3b-d. Use of these data to perform a comparable analysis is more complex. The directionality of the Zwanzig equation gives SSFEP twelve unique free energy differences between the four states, the errors with respect to direct application of BAR are given in 3b. Application of NBR (figure 3c and RSM (figure 3d) can be carried out by working with the various possible subsets of three states. For example, for the red colored PBE→ff99SB result, the PBE ensemble is taken directly from MD whilst the ff99SB ensemble is produced by reweighting/building from the LDA ensemble. Similarly to the hydration data, the techniques give an MUE with respect to direct calculation with BAR of 0.53, 0.50 and 0.54 kcal.mol⁻¹ for SSFEP, NBR and RSM respectively. The lower accuracy results for this data set are most likely due to the small ensembles that were generated at the full DFT level of theory implying that the BAR reference data is not satisfactorily converged.

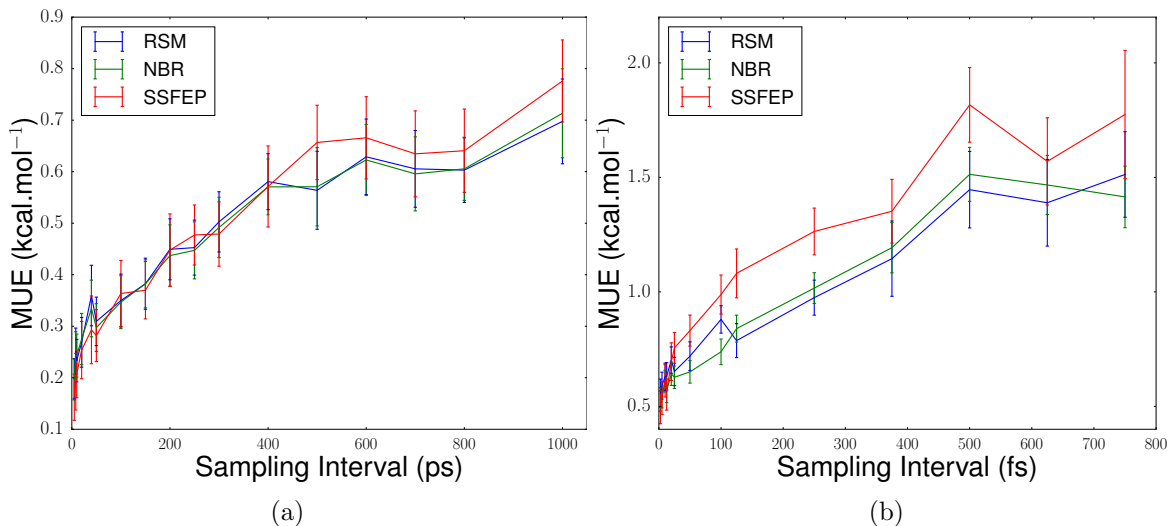


Figure 4: MUEs of different hybrid free energy techniques with increasing sampling interval of MD trajectories for a) hydration free energies and b) base pair data set. Total length of MD trajectories is the same for all free energy differences, however fewer individual configurations contribute to each data point at high sampling intervals. Errors bars show one standard error, calculated by taking the mean of the combined variances in the error for all perturbations.

3.2 Convergence

The results of the previous section demonstrate approximate parity between the techniques under consideration. It should be noted however that these results are generally very well converged even with respect to use of the Zwanzig equation (SSFEP). The use of the BAR estimator with NBR and RSM however may offer improved efficiency, requiring sub-sampling at less frequent intervals to achieve convergence. It is therefore pertinent to consider the relative efficiency of the different techniques, i.e. how rapidly their answer converges with increasing numbers of configurations. To consider this question we have plotted the mean error across all possible perturbations as a function of increasing sampling interval between post-processed configurations (figure 4). These results at best suggest only a very mild advantage with the use of RSM/NBR over SSFEP. This is also most apparent in the less well converged base pair data set.

3.3 Quality Diagnostics

The data so far suggest that the use of RSM and NBR does not offer significant improvements in accuracy or convergence of hybrid free energy calculations over SSFEP. We now consider whether indirect access to QM ensemble properties can provide useful diagnostic information about the quality of computed perturbations. We consider two approaches to this end: an analysis of the relative importance of individual snapshots within a perturbation and the calculation of configuration phase space overlaps.

Consider the RSM ensemble building algorithm: in the limit that the target and source ensembles are the same, individual configurations will occur in both with the same probability. For increasingly less similar ensembles however, configuration frequencies will differ significantly. This simple line of reasoning suggests that it may be possible to identify problematic calculations by considering the distribution of configurations from the source ensemble that are included in the built target ensemble. In cases where the source and target ensembles are very different, the free energies calculated by all techniques would be expected to be of low accuracy. Figure 5a-b show the populations of the source classical ensemble configurations accepted into the target DFTB ensemble built using RSM for chloromethane and nitrobenzene respectively. As expected, the well behaved chloromethane perturbation shows an almost uniform distribution of configurations whereas the poorly converged nitrobenzene result shows a small number of configurations dominating the target ensemble.

Deviations from uniformity of the distribution can be quantified through the χ^2 statistic. Here we are not attempting to establish statistical significance through comparison with the χ^2 distribution but are simply using the computed χ^2 value as a diagnostic. When plotted against unsigned error values for the hydration data set, nitrobenzene is successfully identified as having the largest χ^2 statistic (figure 5c). An analogous analysis can also be carried out for NBR by consider the exponential bias values associated with each configuration (that is $\exp[\beta V_0(\mathbf{x})]$ from (12)).

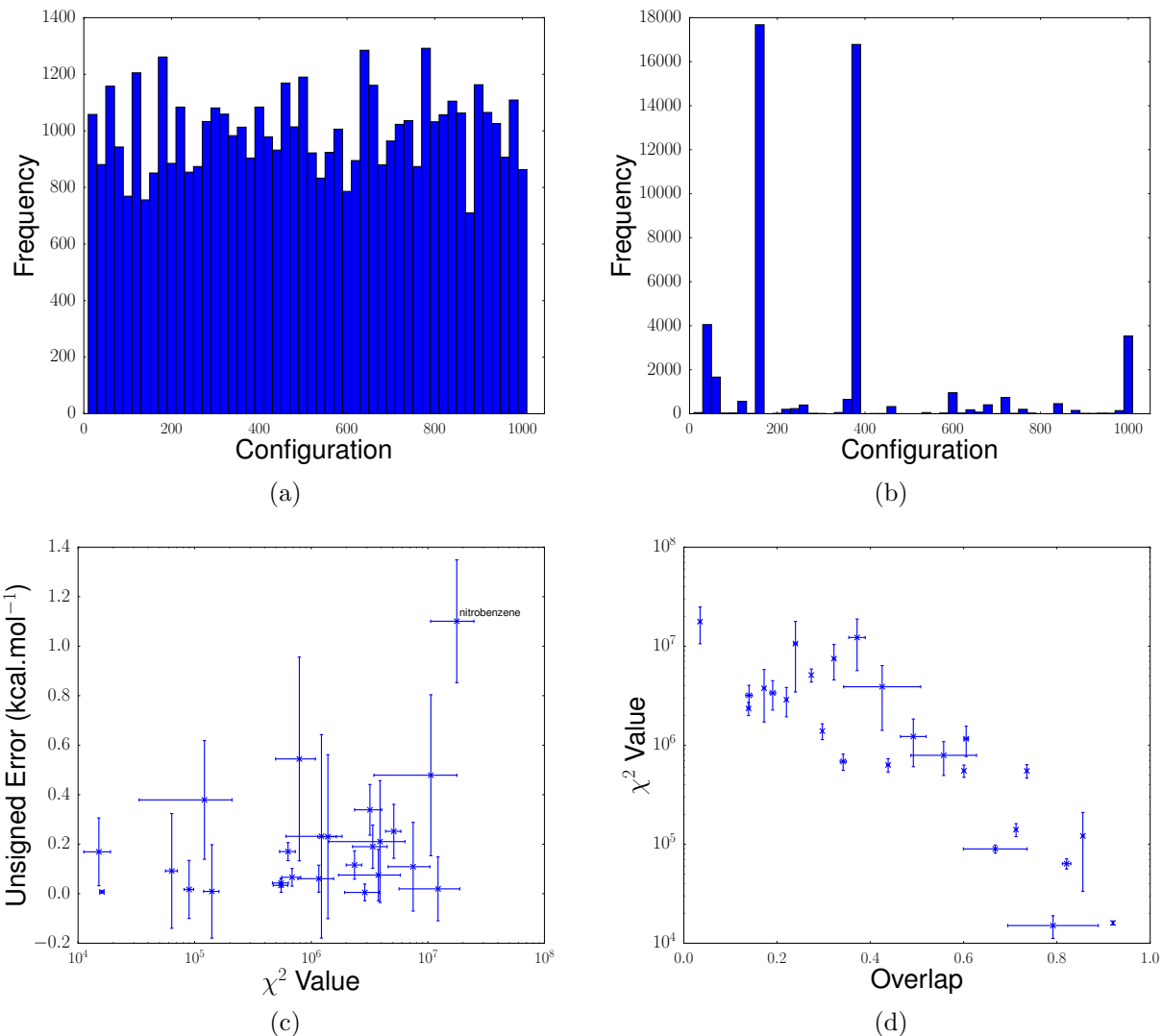


Figure 5: RSM target distribution data. a-b) shows the populations of source ensemble configurations within the target ensemble built by RSM for chloromethane and nitrobenzene respectively. c) shows the relationship between unsigned error values for individual perturbations along with the χ^2 statistic of the built ensembles. d) shows the relationship between the χ^2 statistic for built ensembles and the configuration space overlap between source and target ensembles.

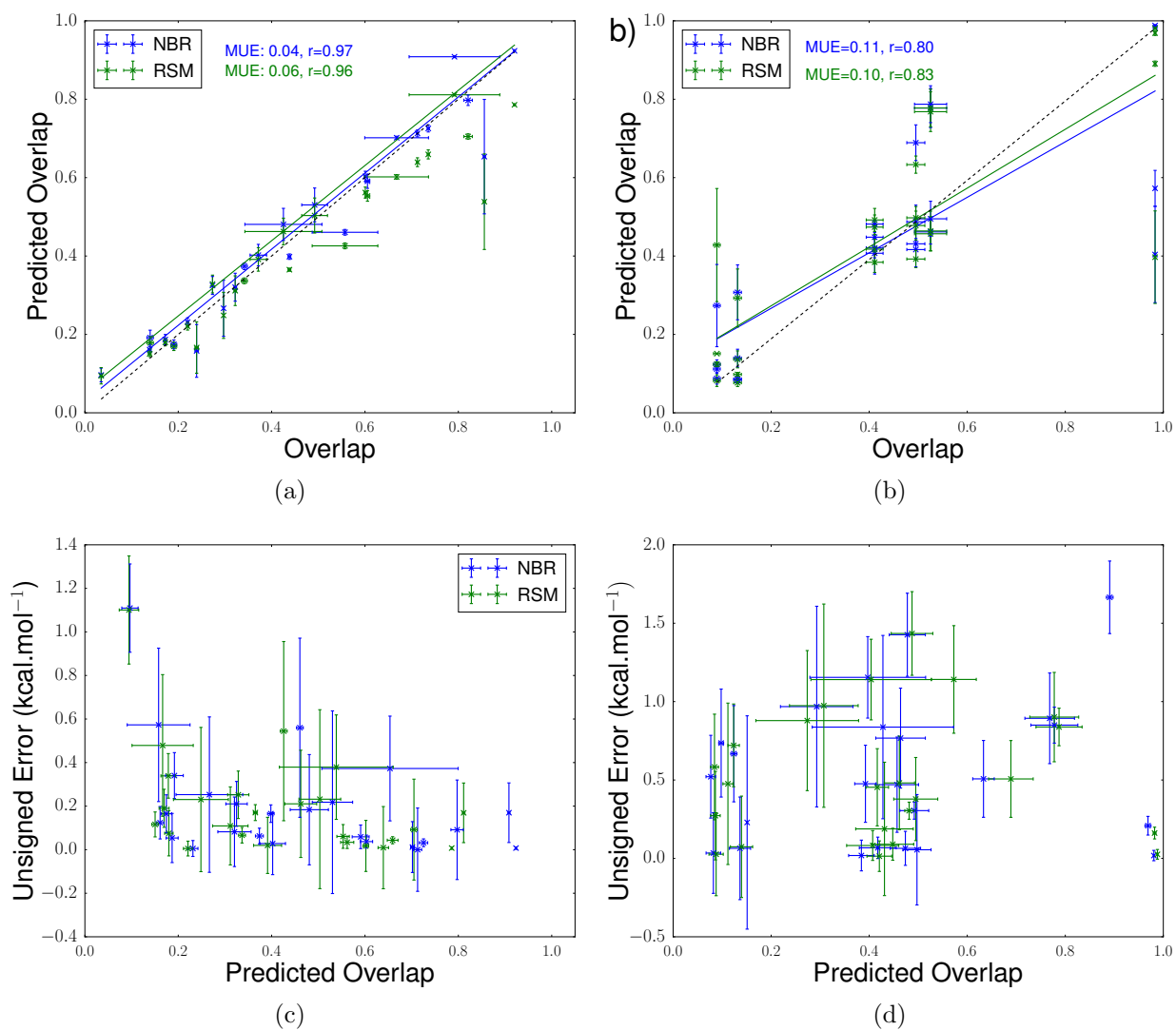


Figure 6: Comparison of predicted configuration space overlap using hybrid free energy techniques for a) hydration data and b) base pair data. Colored lines are respective lines of best fit, black dashed lines give $x = y$. c-d) show predicted overlap values plotted against unsigned errors for individual perturbations.

Whilst a large χ^2 value appears to be a useful warning sign for individual perturbations, clear interpretation in the absence of known errors is hindered as the diagnostic is unnormalized and gives values spread over several orders of magnitude. For this reason we also consider the use of configuration space overlap measures. The impact of configuration space overlap on the convergence of free energy calculations and their utility has been previously considered^{38,39}, in particular in the context of NBR¹⁷. As developed in Methods, we present a simple to compute and intuitive measure for quantification of the configuration overlap between the endpoints of a free energy calculation. The result, (15) requires ensemble averages from the two states of interest. Therefore RSM and NBR can be used to compute this overlap measure without the need to directly sample the target state, something not possible with SSFEP. We here explore the possibility of predicting overlap values in this way and their potential ability to highlight poorly performing perturbations.

Results of this analysis are given in figure 6. Overlap values were acquired using (15) with the directly sampled ensembles of the data set. Predicted overlap values replace one of the directly sampled ensemble averages in (15) with one built by RSM or in the case of reweighting with NBR:

$$O_{0,1}^{NBR} = \frac{2 \langle f(\beta(U_0 - U_1) - C) \rangle_0^2}{\langle f(\beta(U_0 - U_1) - C)^2 \rangle_0 + \frac{\langle f(\beta(U_1 - U_0) + C)^2 \exp[\beta V_{2,1}] \rangle_2}{\langle \exp[\beta V_{2,1}] \rangle_2}} \quad (20)$$

$O_{0,1}^{NBR}$ is the predicted overlap between states 0 and 1, where state 2 has been reweighted to give state 1 and $V_{2,1} = U_2 - U_1$. Equivalently in the hydration data set we obtain the overlap between the DFTB Hamiltonian and the classical Hamiltonian at λ_2 , where the classical Hamiltonian at λ_1 has been reweighted to give the DFTB ensemble. For the base pair use of different permutations of the four states of the data set with (20) again provides 24 unique data points. Predicted overlaps using both the RSM and NBR approaches are shown for the hydration (figure 6a) and base pair data figure (6b) compared with values calculated from independently derived ensembles using (15).

Both data sets show that predicting the true overlap between states is quite accurate (low MUE values and good correlations). This is particularly apparent for the hydration data where better numerical convergence provides a clearer signal. Demonstration of the utility of these values is shown in 6c) where nitrobenzene, the worst performing perturbation, also has the lowest predicted overlap for both NBR and RSM. The same cannot be argued for the base pair data set (6d). Given the relatively poorer numerical convergence and less realistic application of the proposed techniques we do not consider this to be particularly problematic.

The relationship between overlaps and the chi squared metric is direct and shown in figure 5d. As overlap decrease, we see an increase in χ^2 , clearly showing the the MM ensemble is struggling to generate a representative QM ensemble. Overlap values are more readily interpretable than χ^2 however, as they are normalized and linearly distributed. Low predicted overlap values may be useful to highlight “high risk” perturbations where anomalous behaviour is possible and errors may be large. Filtering calculations in this way provides a potential route to allow more consistent performance when applying hybrid free energy techniques.

4 Conclusions

We have presented a rigorous evaluation of some of the most well established techniques for the calculation of MM to QM free energy differences. Through comparison with computational data where converged free energy differences are available we avoid the complications of using experimental data. In addition we present our own technique based on resampling a previously generated MM distribution and application of a Monte Carlo Metropolis test to build an ensemble of structures with the correct Boltzmann weights for the QM Hamiltonian. This technique is theoretically rigorous and consists only of a simple algorithm.

Comparable accuracy is seen for SSFEP, NBR and RSM at a range of sampling frequencies despite the relatively greater computational expense for NBR and RSM. We demonstrate the value of these latter techniques by showing that MC resampling and non-Boltzmann reweighting can be used to predict the configuration space overlap between MM and QM states without the need to directly sample at the QM level. Comparable performance is seen between the techniques in this regard and the useful application of the predicted overlap values to detect notable sources of error is demonstrated. Although we consider these techniques in the context of hybrid free energy calculations we note their generality in the efficient calculation of free energy differences between any Hamiltonians. Thus, despite their extra computational cost, the ability to detect poorly converged calculations using the overlap metric supports the use of NBR or RSM over SSFEP.

Acknowledgement

This work was supported by an EPSRC Doctoral Training Centre grant (EP/G03690X/1). The authors would like to thank the Complex Systems Simulation Doctoral Training Centre. Calculations in this work made use of the Iridis3 and Iridis4 Supercomputers from the University of Southampton. Additional calculations were performed on the UK National Supercomputing Service, using both HECToR and ARCHER. Jonathan Essex and Chris-Kriton Skylaris would like to acknowledge support from EPSRC grant number EP/K039156/13. The authors would like to thank one of the reviewers for noting a typo in equation 10a of reference 25.

Supporting Information Available

Demonstration of Monte Carlo resampling algorithm in a simple one dimensional test sys-

tem. Alternative derivation of resampling Metropolis acceptance test. Full derivation of configuration overlap result. Free energies of hydration compared with experimental data. Reference free energy data calculated using BAR for both the hydration and base pair data. Distributions of nitrobenzene dihedral angle under different Hamiltonians. MM to QM free energies for both data sets. Analysis of statistical properties of convergence behaviour for hydration data set. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

References

- (1) Tjörnhammar, R.; Edholm, O. *J. Chem. Phys.* **2010**, *132*, 205101.
- (2) Martin, D. P.; Blachly, P. G.; McCammon, J. A.; Cohen, S. M. *J. Med. Chem.* **2014**, *57*, 7126–7135.
- (3) Bollini, M.; Domaoal, R. A.; Thakur, V. V.; Gallardo-Macias, R.; Spasov, K. A.; Anderson, K. S.; Jorgensen, W. L. *J. Med. Chem.* **2011**, *54*, 8582–8591.
- (4) Fanfrlík, J.; Kolář, M.; Kamlar, M.; Hurný, D.; Ruiz, F. X.; Cousido-Siah, A.; Mitschler, A.; Řezáč, J.; Munusamy, E.; Lepšák, M.; Matějček, P.; Veselý, J.; Podjarný, A.; Hobza, P. *ACS Chem. Biol.* **2013**, *8*, 2484–2492.
- (5) Dziedzic, P.; Cisneros, J. A.; Robertson, M. J.; Hare, A. A.; Danford, N. E.; Baxter, R. H. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2015**, *137*, 2996–3003.
- (6) Shaw, K. E.; Woods, C. J.; Mulholland, A. J. *J. Phys. Chem. Lett.* **2010**, *1*, 219–223.
- (7) Wesolowski, T.; Warshel, A. *J. Phys. Chem.* **1994**, *98*, 5183–5187.
- (8) Wood, R. H.; Yezdimer, E. M.; Sakane, S.; Barriocanal, J. A.; Doren, D. J. *J. Chem. Phys.* **1999**, *110*, 1329–1337.

- (9) Genheden, S.; Cabedo Martinez, A.; Criddle, M.; Essex, J. W. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 187–200.
- (10) Genheden, S.; Ryde, U.; Söderhjelm, P. *J. Comput. Chem.* **2015**, *36*, 2114–2124.
- (11) Beierlein, F. R.; Michel, J.; Essex, J. W. *J. Phys. Chem. B* **2011**, *115*, 4911–4926.
- (12) Mikulskis, P.; Cioloboc, D.; Andrejić, M.; Khare, S.; Brorsson, J.; Genheden, S.; Mata, R. A.; Söderhjelm, P.; Ryde, U. *J. Comput.-Aid. Mol. Des.* **2014**, *28*, 375–400.
- (13) Olsson, M. A.; Söderhjelm, P.; Ryde, U. *J. Comp. Chem.* **2016**, *37*, 1589–1600.
- (14) König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L. *J. Chem. Theory Comput.* **2014**, *10*, 1406–1419.
- (15) König, G.; Pickard, F. C.; Mei, Y.; Brooks, B. R. *J. Comput.-Aid. Mol. Des.* **2014**, *28*, 245–257.
- (16) Dybeck, E. C.; König, G.; Brooks, B. R.; Shirts, M. R. *J. Chem. Theory Comput.* **2016**, *12*, 1466–1480.
- (17) König, G.; Mei, Y.; Pickard IV, F. C.; Simmonett, A. C.; Miller, B. T.; Herbert, J. M.; Woodcock, H. L.; Brooks, B. R.; Shao, Y. *J. Chem. Theory Comput.* **2016**, *12*, 332–344.
- (18) Hudson, P. S.; Woodcock, H. L.; Boresch, S. *J. Phys. Chem. Lett.* **2015**, *6*, 4850–4856.
- (19) Woods, C. J.; Manby, F. R.; Mulholland, A. J. *J. Chem. Phys.* **2008**, *128*, 014109.
- (20) Sampson, C.; Fox, T.; Tautermann, C. S.; Woods, C.; Skylaris, C.-K. *J. Phys. Chem. B* **2015**, *119*, 7030–7040.
- (21) Cave-Ayland, C.; Skylaris, C.-K.; Essex, J. W. *J. Phys. Chem. B* **2015**, *119*, 1017–1025.
- (22) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (23) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 144107.

- (24) Fox, S. J.; Pittock, C.; Tautermann, C. S.; Fox, T.; Christ, C.; Malcolm, N. O. J.; Essex, J. W.; Skylaris, C.-K. *J. Phys. Chem. B* **2013**, *117*, 9478–9485.
- (25) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (26) Case, D. A.; Betz, R. M.; Botello-Smith, W.; Cerutti, D. S.; Cheatham, T. E.; III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. **2016**, *AMBER 16*, University of California, San Francisco.
- (27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (28) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (29) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132–146.
- (30) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.
- (31) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.
- (32) Krüger, T.; Elstner, M.; Schiffels, P.; Frauenheim, T. *J. Chem. Phys.* **2005**, *122*.
- (33) Seabra, G. d. M.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E. *J. Phys. Chem. A* **2007**, *111*, 5655–5664.
- (34) Tange, O. ;login: *The USENIX Magazine* **2011**, *36*, 42–47.
- (35) Gaus, M.; Goez, A.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.

- (36) Kubillus, M.; Kuba, T.; Gaus, M.; ez, J.; Elstner, M. *J. Chem. Theory Comput.* **2015**, *11*, 332–342.
- (37) Jia, X.; Wang, M.; Shao, Y.; Knig, G.; Brooks, B. R.; Zhang, J. Z. H.; Mei, Y. *J. Chem. Theory Comput.* **2016**, *12*, 499–511.
- (38) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 054103.
- (39) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 084109.

Graphical TOC Entry

