

## On the convergence of multi-scale free energy simulations

Gerhard König, Bernard R. Brooks, Walter Thiel & Darrin M. York

To cite this article: Gerhard König, Bernard R. Brooks, Walter Thiel & Darrin M. York (2018) On the convergence of multi-scale free energy simulations, Molecular Simulation, 44:13-14, 1062-1081, DOI: [10.1080/08927022.2018.1475741](https://doi.org/10.1080/08927022.2018.1475741)

To link to this article: <https://doi.org/10.1080/08927022.2018.1475741>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 30 May 2018.



Submit your article to this journal



Article views: 1128



View related articles



View Crossmark data

## On the convergence of multi-scale free energy simulations

Gerhard König  <sup>a,b,c</sup>, Bernard R. Brooks<sup>c</sup>, Walter Thiel  <sup>a</sup> and Darrin M. York<sup>b</sup>

<sup>a</sup>Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, Germany; <sup>b</sup>Laboratory for Biomolecular Simulation Research, Center for Integrative Proteomics Research, and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ, USA; <sup>c</sup>Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD, USA

### ABSTRACT

In this work, we employ simple model systems to evaluate the relative performance of two of the most important free energy methods: The Zwanzig equation (also known as ‘Free energy perturbation’) and Bennett’s acceptance ratio method (BAR). Although our examples should be transferable to other kinds of free energy simulations, we focus on applications of multi-scale free energy simulations. Such calculations are especially complex, since they connect two different levels of theory with very different requirements in terms of speed, accuracy, sampling and parallelisability. We try to reconcile all those different factors by developing some simple criteria to guide the early stages of the development of a free energy protocol. This is accomplished by quantifying how many  $\lambda$  intermediate steps and how many potential energy evaluations are necessary in order to reach a certain level of convergence.

### ARTICLE HISTORY

Received 11 April 2018

Accepted 4 May 2018

### KEYWORDS

Free energy protocol design;  
multi-scale simulations;  
convergence properties

Free energy calculations are among the most important methods in the toolbox of computational chemistry. They provide a direct connection between molecular simulation and macroscopic thermodynamic observables such as binding constants, solubilities or the equilibrium of reactions. Moreover they are rigorous from a statistical mechanics point of view and, overall, provide superior results compared to more approximate methods. Unfortunately, they are also computationally expensive and not necessarily straightforward for the non-expert (or sometimes even the expert) to perform.

Multi-scale models represent another essential set of tools that have made a tremendous impact on our understanding of a wide range of complex chemical systems. This culminated in the 2013 Nobel Prize in Chemistry that was awarded to Martin Karplus, Michael Levitt and Arieh Warshel. Multi-scale models integrate multiple levels of theory that work together to span broad domains in terms of time, space and accuracy. Thus, it seems natural that multi-scale models should be combined with free energy simulations to achieve a desired level of precision and accuracy. However, it is often not clear how to perform such calculations in the most efficient way.

One particularly vexing issue in most applications of free energy simulations arises before the free energy simulation is even performed. When calculating a free energy difference between two states, one implicitly attempts to enumerate the partition functions of the involved end states. In principle this means that one has to sample the complete energy surface of both states – a daunting task for any system that involves more than a couple of degrees of freedom.

A different way to state the free energy difference problem is to ask how similar are the potential energy surfaces to each other. After all, the free energy difference corresponds to the ratio of the partition functions. The way to measure this similarity is calculating the potential energy difference between the two end states for a series of points in phase space. Those points in phase space are typically provided by molecular dynamics simulations. The overall measure for the similarity of the partition functions, the free energy difference, will converge quickly if the resulting distribution of the potential energy differences is well behaved. Wild fluctuations of energy differences lead to poor convergence of the free energy calculation.

The predominant technique to improve the convergence of free energy simulations between two very different energy surfaces is the introduction of artificial intermediate states. Commonly, those intermediate states are created by mixing the properties of the two end states according to a mixing ratio  $\lambda$ , where  $\lambda = 0$  corresponds to the initial state and  $\lambda = 1$  corresponds to the final state. Figuratively speaking, instead of comparing apples with oranges, one creates a series of apple-orange hybrids and compares those to each other.

Since each  $\lambda$  state requires a separate molecular dynamics simulation, it is necessary to evaluate how many intermediate states are necessary. This decision has a serious impact on the free energy calculation. On the one hand, using insufficient intermediate states can lead to biased free energy estimates which affect the accuracy of the free energy difference. On the other hand, every intermediate state increases the computational costs. Therefore, the introduction of unnecessary intermediate states is a waste of computer time. Moreover,

**CONTACT** Gerhard König  gerhard.koenig@rutgers.edu; Darrin M. York  Darrin.York@rutgers.edu

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

superfluous intermediate states lead to opportunity costs in terms of accuracy. Often, precious computer time is invested in the creation of intermediate states although it would have been more opportune to spend it on longer simulations or additional repetitions of the simulation to evaluate the reproducibility of the results.

While certain criteria are available to decide *a posteriori* whether sufficient intermediate states have been used in a free energy simulation [1–6], there is no such criterion to decide beforehand how many  $\lambda$  intermediate states are required. The underlying reason is simply that the required measure for sufficient similarity in terms of phase space overlap in a free energy simulation is the resulting free energy difference itself (or, more specifically, the entropic contribution to the free energy, as will be discussed later on in the Theory section). Thus, to a certain degree, every free energy simulation involves a chicken-and-egg dilemma.

Clearly, there is no exact solution to the problem of finding an optimal  $\lambda$  protocol for free energy simulations, and any attempt to do so would border on the absurd. However, one can argue that finding a good  $\lambda$  protocol is easier than finding a good approximation to the free energy difference. First, the accuracy and precision demands are very different for the actual free energy result and for the data that is required for estimating a good free energy protocol. Second, it is possible to compensate for mistakes in the  $\lambda$  protocol by running longer simulations. Finally, it is straightforward to detect mistakes in the  $\lambda$  protocol using the aforementioned phase space overlap criteria. Thus errors can also be compensated by introducing more intermediate steps after the fact with the benefit of hindsight. However, our main argument for trying to estimate a good free energy protocol is pragmatic in nature: any reasonable automatic procedure to estimate the required  $\lambda$  protocol is superior to the current state of the art: Guessing followed by trial and error.

In the following, we attempt to provide some basic guidelines to estimate a good  $\lambda$  protocol with the limited information that is available before performing the actual free energy simulation. For this purpose, we borrow some ideas from Marcus theory and linear response theory [7]. In particular, we develop some simple criteria for the convergence of free energy simulations based on the Zwanzig equation [8] and Bennett's acceptance ratio method [6]. We start by introducing the two methods and compare their efficiency in terms of the convergence of some very simple model systems. Then we describe how the efficiency depends on the relative computational costs of the end states. This is relevant for multi-scale free energy simulations, where one end state can be a molecular mechanics simulation and the other end state can be a quantum-mechanical simulation. We also explain in how far the two methods benefit from parallelisation on modern hardware architectures. In the next step, we show how the computational costs depend on the introduction of intermediate states. Based on free energy simulations of twelve molecules in the gas phase, we exemplify the limitations of our simplistic approach. Finally, we present results for multi-scale free energy simulations between molecular mechanics and quantum mechanics.

## 1. Theory

### 1.1. Free energy

For readers that are not necessarily familiar with free energy simulations, we use the following sections to introduce the necessary background and our notation. The free energy difference between two systems at constant volume is given by the Helmholtz free energy,  $A$ , which is defined by

$$\Delta A_{0 \rightarrow 1} = -kT \ln \frac{Z_1}{Z_0}, \quad (1)$$

where  $k$  is the Boltzmann constant,  $T$  is the temperature and  $Z$  is the partition function. The subscript identifies the corresponding end state, where we refer to the initial state as state 0, and the final state as state 1. Based on the work of Boltzmann, partition functions are defined as sums over all possible micro-states, i.e.

$$Z_a = \sum_i e^{-\frac{U_a(\vec{r}_i)}{kT}}, \quad (2)$$

where  $i$  enumerates all possible combinations of the coordinates  $\vec{r}_i$  to generate a potential energy  $U_a$  based on the potential energy function of state  $a$ . Although the enumerator  $i$  looks very innocent, it is actually one of the reasons why a computer simulation can never fully converge, because even with the finite machine representation of floating point numbers there are usually still more possible manifestations of  $i$  than what can be covered with our computers in reasonable time. Fortunately, not all states are equally likely, therefore our computer simulations can be restricted to analysing states with a high importance, which corresponds to the low-lying energy regions. In practice, this means that each state is generated according to its probability  $p$ , which corresponds to

$$p_a(\vec{r}_j) = \frac{e^{-\frac{U_a(\vec{r}_j)}{kT}}}{Z_a}, \quad (3)$$

and leads to an expression for Boltzmann-weighted ensemble averages of a property  $X$ , i.e.

$$\langle X \rangle = \sum_i X(\vec{r}_i) p(\vec{r}_i). \quad (4)$$

In this context, it is convenient to introduce a short hand notation that is employed in the remainder of this paper. First, reduced units are used to express the potential energy, which means that energies are given in units of  $kT$ . Second, we skip writing out the dependence of the potential energies,  $U$ , on the actual coordinates  $\vec{r}$  in ensemble averages, because it is implied that the potential energies have to be evaluated for all coordinates encountered in a trajectory.

For example, it is a well-known fact that the free energy difference is composed of both an enthalpic and an entropic contribution, i.e.

$$\Delta A_{0 \rightarrow 1} = \Delta H_{0 \rightarrow 1} - T \Delta S_{0 \rightarrow 1}. \quad (5)$$

Since the contribution from the kinetic energy cancels between the end points as long as mass and the number of particles are conserved, the enthalpic contribution is given by

$$\Delta H_{0 \rightarrow 1} = \langle U_1 \rangle_1 - \langle U_0 \rangle_0, \quad (6)$$

which can be calculated with relative ease from the simulations. However, the entropic contribution is given by

$$S = \sum_i \ln p(\vec{r}_i) p(\vec{r}_i), \quad (7)$$

which means that the entropy difference is

$$\Delta S_{0 \rightarrow 1} = \sum_i [\ln p_1(\vec{r}_i) p_1(\vec{r}_i) - \ln p_0(\vec{r}_i) p_0(\vec{r}_i)]. \quad (8)$$

Basically, the entropic contribution is a measure for the similarity of the two phase spaces and represents how easy it is to move from one state to the other [9]. If the probability distributions  $p_0$  and  $p_1$  are very similar, the entropy difference is close to zero. If the initial state has more phase space available than the final state, the entropy is negative and, therefore, the free energy contribution from the entropy is positive. Since the determination of the probabilities requires a good estimate of the partition functions to obtain the proper normalisation constant, long simulations have to be performed for convergence [10]. This is the very aspect that makes free energy simulations so expensive.

## 1.2. Free energy methods

Two of the main workhorses in the field of free energy simulations are the Zwanzig equation (also known as ‘Thermodynamic Perturbation’, ‘Free Energy Perturbation’ or ‘Exponential Formula’) [8], as well as Bennett’s Acceptance Ratio (BAR) [6]. Notably, the third important free energy method is Thermodynamic Integration (TI) [11], but this method has only played a minor role in multi-scale free energy simulations so far. Therefore, we restrict ourselves to pointing out that TI can be almost as efficient as BAR if it is used correctly [4,12]. Also there are some recent innovations that make TI more efficient [13,14]. Excellent introductions to this matter are also provided in several books and papers [4,12,15–17].

A key innovation since the turn of the millennium is the rise of the non-equilibrium work techniques by Jarzynski [18] and Crooks [19]. These approaches use the non-equilibrium work carried out in short switching simulations between the two end states. In the limit of infinitely fast switching simulations, the non-equilibrium work is the same as the potential energy difference between the end states. Thus, in this limit, the Jarzynski equation is the same as the Zwanzig equation and Crooks Theorem is the same as BAR, which is another reason why we restrict our analysis to the Zwanzig equation and BAR. For the advanced reader, we also want to point out that all the conclusions drawn in the following sections should be equally valid for the respective non-equilibrium methods. One just has to substitute our expressions for the mean potential energy difference by the corresponding mean non-equilibrium work. Unfortunately, we are not aware of a straightforward connection

between those two properties, since the non-equilibrium work depends on both the switching speed and the nature of the energy surface. Thus, the presented convergence criteria for Zwanzig and BAR should be considered mere bounds for the non-equilibrium expressions.

## 1.3. Zwanzig equation

The Zwanzig equation is a one-sided free energy method, in the sense that only one of the two end states is used for sampling. This makes it attractive for multi-scale free energy simulations, where potential energy evaluations on the lower level of theory are significantly cheaper than on the higher level of theory. The Helmholtz free energy difference between two states is given by

$$\Delta A_{0 \rightarrow 1}^{Zwanzig} = -\ln \left\langle e^{-(U_1 - U_0)} \right\rangle_0. \quad (9)$$

Here, the potential energy,  $U$ , is in reduced units of  $kT$ . For each frame of the trajectory of the initial state 0, the potential energies are evaluated both with the potential energy function of the initial state ( $U_0$ ) and of the final state ( $U_1$ ). For example, in a multi-scale free energy simulation of a molecule in the gas phase,  $U_0$  corresponds to a potential energy evaluation with molecular mechanics, while  $U_1$  corresponds to the use of quantum mechanics.

Of course the Zwanzig equation can also be performed in the reverse direction, going from the final state to the initial state, yielding a  $\Delta A_{1 \rightarrow 0}$  based on a trajectory of the final state. While, theoretically, the forward process  $\Delta A_{0 \rightarrow 1}$  should yield the same result as the backward process  $\Delta A_{1 \rightarrow 0}$ , this is usually not the case in practice due to limited sampling. Thus, the resulting forward and backward Zwanzig results represent bounds to the free energy difference [6,20]. However, as noted by Lu and Kofke, the Zwanzig results are more reliable if the direction goes from a high entropy to a low entropy state due to sampling [21].

Poor convergence of the free energy result is usually addressed by introducing intermediate states. In this case, the total free energy calculation is broken down into sub-steps, i.e.

$$\Delta A_{0 \rightarrow 1} = \sum_{\lambda=0}^{1-\Delta\lambda} \Delta A_{\lambda \rightarrow \lambda + \Delta\lambda}, \quad (10)$$

where  $\lambda$  represents a mixing ratio between the two end states and  $\Delta\lambda$  is the step size in  $\lambda$  space at which simulations are conducted. It is usually easier to use a fixed step size  $\Delta\lambda$ , but it sometimes turns out that some sub-steps are more challenging in terms of convergence than others. In such situations, it makes sense to introduce more  $\lambda$  states in the recalcitrant part of the simulation, which leads to unequal step sizes. One such challenge is the so-called van der Waals end point problem that arises when the van der Waals interactions of certain atoms are turned to zero [22].

A special case of the use of the Zwanzig equation is so-called Double-Wide Sampling (DWS) [23], where  $\lambda$  states are connected via half-steps, which involves free energy calculations to virtual midpoints. Thus, the total free energy difference becomes

$$\Delta A_{0 \rightarrow 1}^{DWS} = \sum_{\lambda=0}^{1-\Delta\lambda} \Delta A_{\lambda \rightarrow \lambda + \frac{\Delta\lambda}{2}} - \Delta A_{\lambda + \Delta\lambda \rightarrow \lambda + \frac{\Delta\lambda}{2}}. \quad (11)$$

Since only half the step size in  $\lambda$  space is used, the phase space overlap between the employed end points of each sub-step is increased, which improves convergence. Another noteworthy variation of the use of the Zwanzig equation is Overlap Sampling [21,24], which considers the basic asymmetry between the forward and the backward direction of the Zwanzig results and optimises the position of the virtual intermediate state that has been introduced here in the context of DWS. By finding the optimal virtual intermediate state between two end points, the use of two Zwanzig equation results becomes equivalent to employing Bennett's acceptance ratio method [21,24].

#### 1.4. Bennett's acceptance ratio

BAR [6] is a two-sided free energy estimator, which requires simulations of both end points of a free energy step. The BAR equation can be derived in a number of ways. Bennett's original derivation is based on finding a free energy estimate that minimises the variance of the free energy outcome. Shirts and Pande helped to popularise the method by showing that BAR is equivalent to a maximum likelihood estimator [25]. This makes BAR very attractive from a statistical point of view, since maximum likelihood estimators are known to be well-behaved, asymptotically unbiased and no other estimator has a lower variance [25]. As already mentioned before, BAR is also the equilibrium equivalent of Crooks non-equilibrium work theorem [19]. Moreover, BAR is connected to multi-state methods like the Weighted Histogram Analysis Method (WHAM) [26,27], the multi-state Bennett method (MBAR) [28], the non-Boltzmann Bennett method (NBB) [29–32] and the optimal multi-state free energy estimator by Maragakis, Spichy and Karplus [33].

To perform a BAR calculation, the potential energy differences between the two end states have to be calculated for each frame of the two trajectories. Again we resort to the use of reduced units of  $kT$ . The free energy difference results from

$$\Delta G_{0 \rightarrow 1}^{BAR} = -\ln \left( \frac{\langle f(U_0 - U_1 + C) \rangle_1}{\langle f(U_1 - U_0 - C) \rangle_0} \right) + C, \quad (12)$$

where  $f(x)$  denotes the Fermi function ( $\frac{1}{1+\exp(x)}$ ). The unknown variable  $C$  is found using an iterative procedure that sets the ratio of the two ensemble averages to unity. A good initial guess for  $C$  can be obtained by estimating the free energy difference with Double-Wide Sampling or Overlap Sampling. Alternatively, if histograms of the  $\Delta U$  distributions for both end states are available, one can also employ the  $\Delta U$  at which the two probability distributions  $p_0(\Delta U)$  and  $p_1(\Delta U)$  intersect. In our experience, it is advisable to create the histograms of  $\Delta U$  in any case, since the inspection of the overlap integral of the two distributions provides the necessary information to judge whether the resulting free energy is converged. As a rule of thumb, the overlap should exceed 1% in order to be considered reliable. This aspect is also discussed in the subsequent sections. Commonly an equal number of data points is produced for both end states, but BAR can also be employed if the number of data points differs. In some cases, better results are possible by producing more data points for the end point with higher entropy (cf. Equations (13)–(16) in Ref. 6). Unfortunately, it is

usually not known *a priori* which end state exhibits the higher entropy, therefore it is reasonable to start with equal amounts of sampling. Finally, it is important to point out that BAR and its derivatives are currently the most efficient methods for determining free energy differences [3,4,34–37].

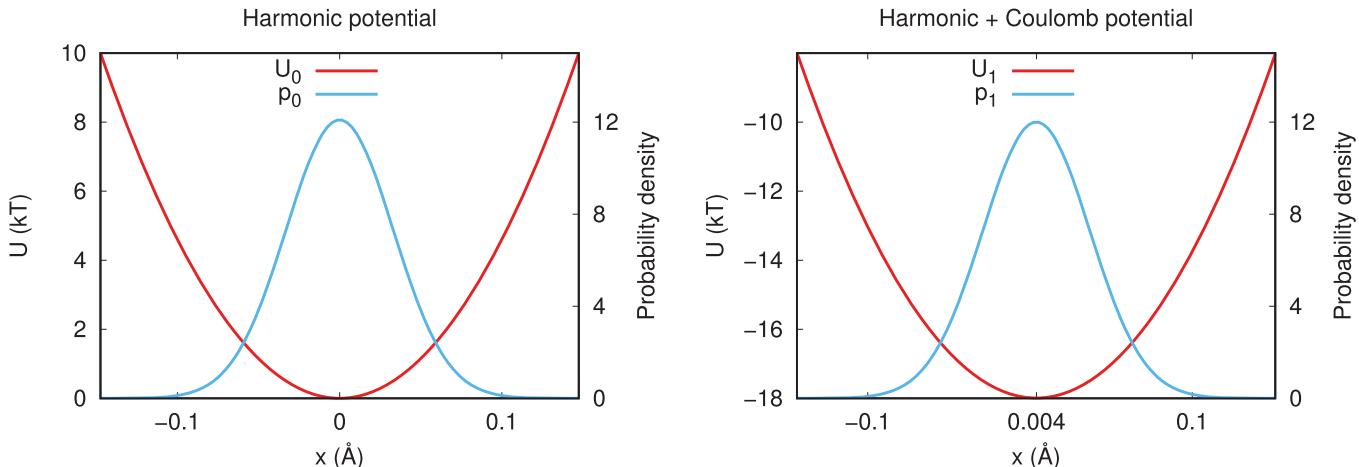
#### 1.5. Introduction of the model system

As already pointed out in the Introduction, the starting point of our endeavour to predict the convergence properties of a free energy simulation protocol is the use of linear response theory. Linear response theory is very well established in the context of Marcus theory to explain the rates of electron transfer reactions [7]. It also forms the basis of the Linear Interaction Energy (LIE) method, which is sometimes employed in drug design [38–43]. However, it is an approximate method as it assumes that the underlying energy fluctuations obey Gaussian statistics. This corresponds to a second-order approximation of the free energy difference. Hummer, Pratt and Garcia demonstrated with hydration free energies of ions that the electrostatic potential energy distribution is Gaussian [44], and Åqvist and Hansson showed that the linear response approximation holds well for monovalent ionic solutes [45]. However, it was also shown by Smith and van Gunsteren that this approach breaks down for particle insertion and deletion due to the coupling between the different free energy contributions [46]. A recent analysis by Heid, Moser and Schröder indicates that the linear response approximation can also break down for electrostatic changes under some circumstances [47]. Nevertheless, this approach has proven to be useful in practice, if one keeps aware of its limitations. We thus follow in this pragmatic tradition, but also dedicate a separate section in Results to the discussion of the limitations of this approach.

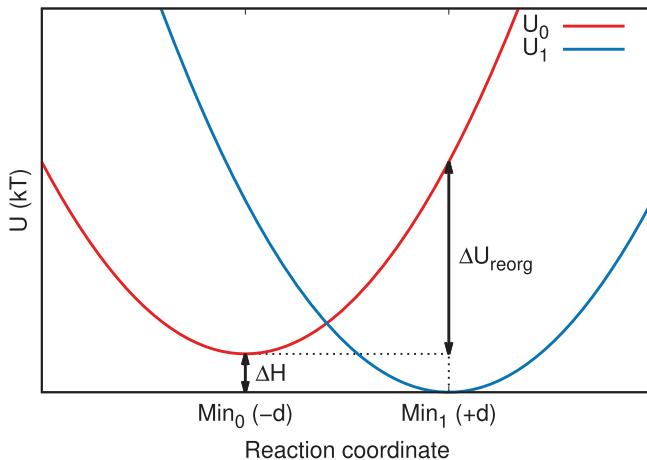
Our choice of model system is motivated by the fact that most multi-scale and absolute free energy calculations employ a single topology setup with identical numbers of atoms in the initial and final state. This setup entails no change of the general topology of the structure, and no change of the force constants, but still uses different Hamiltonians at the end points. Similar calculations are also used in transfer free energy calculations or binding affinity calculations. Thus, the free energy calculations mostly gauge the response to a change of the environment.

The strongest interactions within a molecule are usually the chemical bonds, whose potential energy surface can be approximated by harmonic potentials. The dominant change of the environment usually involves electrostatic interactions, which are described by the Coulomb potential. Combining a Coulomb potential with a significantly stronger harmonic potential leads to a change of the position and depth of the energy minimum. An example of such a change is shown in Figure 1, where the axis labels reflect the main differences. The difference between the left and the right side in Figure 1 corresponds to the change of the system as attractive electrostatic interactions are introduced. This mimics the transfer of small molecule from the gas phase into a polar solution.

As can be seen in the example of Figure 1, the electrostatic interaction energy is relatively strong with ca.  $-18 kT$ . Yet, the introduced anharmonicity and the change of the force constant are rather small. In addition, the probability distributions (blue



**Figure 1.** (Colour online) Comparison of a harmonic potential (red) at room temperature with a force constant of  $460 \text{ kT \AA}^{-2}$  (left) with the same harmonic potential under the influence of a Coulomb potential with charges of  $q_0 = +0.165e$  and  $q_1 = -e$  at a distance of  $5 \text{ \AA}$  from the equilibrium position of the harmonic potential (right). The electrostatic interactions shift the position of the energy minimum as well as the well depth. However, the resulting probability densities (blue) still remain mostly Gaussian and the phase space overlap between the two systems is high. This model mimics the effect of changing the charges of molecule or inserting an ion close to a thiol or hydroxyl hydrogen.



**Figure 2.** (Colour online) Description of two harmonic oscillators. The potential energy,  $U$ , is shown relative to an arbitrary reaction coordinate. The initial state is called state 0, while the final state is state 1. The minima reside at a distance  $d$  from the midpoint between the two minima. The different well depths define the enthalpy difference  $\Delta H$ . The energetic cost to visit the energy minimum of the other state is given by the reorganisation energy ( $\Delta U_{\text{reorg}}$ ).

lines) before and after the change are very similar (compare left and right side of Figure 1, where the main change is the shift of the centre of the Gaussian probability distribution from 0 to ca.  $0.04 \text{ \AA}$ ). From the perspective of phase space overlap, this response of the harmonic system can thus be modelled by two shifted harmonic oscillators with identical force constant. This should also approximately be the case in a multi-scale free energy calculation between MM and QM, if the MM force field has been properly parameterised, and the main difference is the impact of polarisation on the charge distribution.

In the following, we define the basic components of our model system, which consists of two harmonic oscillators. Thus, the probability distributions are Gaussian, as required by linear response theory. The underlying potential energies of the harmonic oscillators are defined as  $U_0(x) = K(x + d)^2$  and  $U_1(x) = K(x - d)^2 + \Delta H$ , where  $x$  is the reaction coordinate and

the subscript indicates the state (0=initial state, 1=final state).  $K$  is the force constant, and  $d$  is the deviation of the equilibrium position of the harmonic oscillator from the midpoint between the two energy minima. Here, the force constants  $K$  include the conventional factor of  $\frac{1}{2}$  and are in reduced units of  $kT$ .  $\Delta H$  signifies the enthalpy difference, which corresponds to the energy difference between the two energy minima (see Figure 2). The resulting probability distributions are  $p_0(x) = Z_0^{-1} e^{-U_0(x)}$  and  $p_1(x) = Z_1^{-1} e^{-U_1(x)}$ , where the partition functions are given by the Gaussian integral with  $Z_0 = \sqrt{\frac{\pi}{K}}$  and  $Z_1 = \sqrt{\frac{\pi}{K}} e^{\Delta H}$ . The potential energy difference,  $\Delta U$ , is given by  $\Delta U(x) = U_1(x) - U_0(x) = -xK4d + \Delta H$ . Following Marcus theory, we also define the reorganisation energy, which is the potential energy difference between the coordinates of the energy minima using the potential energy function of the state under consideration, i.e.

$$\Delta U_{\text{reorg}}^{0 \rightarrow 1} = U_0(d) - U_0(-d) = 4Kd^2 \quad (13)$$

and

$$\Delta U_{\text{reorg}}^{1 \rightarrow 0} = U_1(-d) - U_1(+d) = 4Kd^2 \quad (14)$$

(see also Figure 2). In the case of our model system, the reorganisation energies are identical for both states, since the force constants are the same (i.e.  $\Delta U_{\text{reorg}}^{0 \rightarrow 1} = \Delta U_{\text{reorg}}^{1 \rightarrow 0} = \Delta U_{\text{reorg}}$ ). In practice,  $\Delta U_{\text{reorg}}$  can be obtained by an energy minimisation that starts with the coordinates of the energy minimum of the other state.

### 1.6. Convergence properties of free energy estimates

Equipped with this simple model system, we now attempt to determine under which circumstances it is worthwhile to perform simulations of both end states (as required for BAR) and under which circumstances it is enough to just conduct a simulation for one of the end states (corresponding to the use of the Zwanzig equation). This is especially relevant if sampling one of the end states would incur significantly higher computational expenses, as is the case in multi-scale free energy simulations,

where one compares an MM energy surface to a QM or QM/MM energy surface.

The efficiency and bias of the Zwanzig equation and Bennett's acceptance ratio method have been described in great detail by Shirts and Pande, [48] as well as Zuckerman and Woolf [49]. Therefore, we restrict ourselves to a short summary here. We also just discuss the variance of the free energy estimate and not the bias, since the expected bias of the Zwanzig result is half of its variance [48,49].

### 1.6.1. Zwanzig equation

As described in Equations (19)–(23) of Ref. [48], the variance of a free energy result based on the Zwanzig equation,  $\sigma_{Zwanzig}^2$ , is proportional to the variance of the underlying potential energy difference distribution,  $\sigma_{\Delta U}^2$ . I.e.,

$$\sigma_{Zwanzig}^2 = \frac{e^{\sigma_{\Delta U}^2} - 1}{n}, \quad (15)$$

where  $n$  is the number of data points/potential energy difference evaluations used in the free energy estimate. For two shifted harmonic oscillators, there is a direct mapping between the probability distribution of the reaction coordinate position on the one hand, and the probability distribution of the potential energy difference on the other hand. Since the reaction coordinate can be expressed in terms of the potential energy difference, i.e.  $x = -\frac{\Delta U - \Delta H}{4Kd}$ ,  $\sigma_{\Delta U}^2$  becomes

$$\sigma_{\Delta U}^2 = 8Kd^2 = 2\Delta U_{reorg}. \quad (16)$$

Thus, the variance of the Zwanzig free energy estimate can be connected to the reorganisation energy via

$$\sigma_{Zwanzig}^2 = \frac{e^{2\Delta U_{reorg}} - 1}{n}. \quad (17)$$

A comparison of the theoretical variance with variances from actual simulations of harmonic systems is shown on top in Figure 7 of the Results section. Importantly, it is not necessary to know the exact force constant nor the nature of the underlying reaction coordinate that leads to the reorganisation energy if we assume that the force constants of the end points are identical. Thus, no normal mode analysis is required to obtain the variance estimate. This makes the analysis in practice very convenient, since only two simple energy minimisations are required to obtain the reorganisation energy.

### 1.6.2. Bennett's acceptance ratio

To the best of our knowledge, there are no amenable estimates for the variance of BAR. Equation (11) of Bennett's original paper [6] specifies that

$$\sigma_{BAR}^2 = \frac{2}{n} (\Omega^{-1} - 1) \quad (18)$$

where  $n$  lies between the number of data points for the initial state ( $n_0$ ) and the final state ( $n_1$ ). The phase space overlap  $\Omega$  is given by

$$\Omega = \int \frac{2p_0(x)p_1(x)}{p_0(x) + p_1(x)} dx, \quad (19)$$

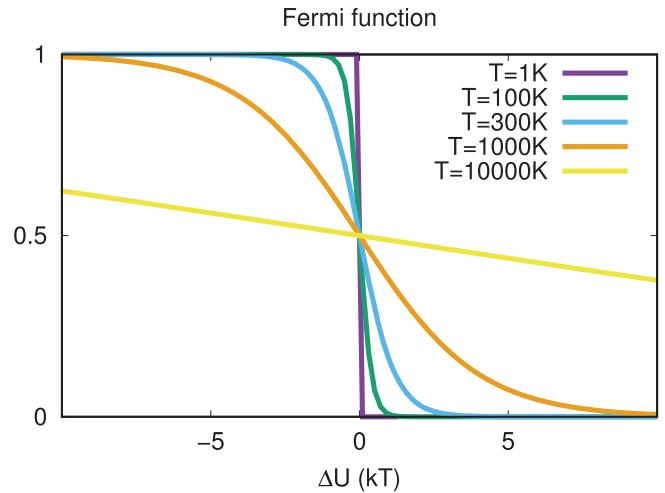


Figure 3. (Colour online) The Fermi function for different temperatures.

which depends on the details of the energy function. This corresponds to

$$\begin{aligned} \Omega &= 2 \int \frac{e^{-U_0(x)} e^{-(U_1(x) - \Delta H)}}{Z_0 (e^{-(U_1(x) - \Delta U(x))} + e^{-(U_1(x) - \Delta H)})} dx \\ &= 2 \int p_0(x) f(-\Delta U(x) + \Delta H) dx, \end{aligned} \quad (20)$$

where  $f$  again denotes the Fermi function  $f(X) = 1/(1 + e^X)$ . Notably, a constant shift of the potential energy difference in form of the enthalpy difference  $\Delta H$  has no effect on the probability distribution, because it cancels out.

The Fermi function is depicted in Figure 3 for different temperatures. Figure 3 indicates that the Fermi function can be approximated by a step function at low temperatures. For our harmonic oscillator example,  $\Delta U - \Delta H < 0$  if  $x > 0$ , which corresponds to the tail of the probability distribution of  $p_0$ . Because of the symmetry of the Fermi function around 0, most of the error resulting from this approximation is going to cancel, except for the asymmetric contributions from  $p_0$ . Thus, the approximation by a step function is most likely underestimating the real overlap, and, consequently, overestimating the variance. With the step function approximation, the phase space overlap becomes

$$\Omega \approx 2 \int_0^\infty p_0(x) dx. \quad (21)$$

The probability distribution of a harmonic oscillator is a Gaussian,

$$\Omega \approx 2 \int_0^\infty Z^{-1} e^{-K(x+d)^2},$$

where the integral corresponds to the complementary error function,  $\text{erfc}$ , i.e.

$$\Omega \approx \frac{\sqrt{\pi}}{Z\sqrt{K}} \text{erfc}(d\sqrt{K}).$$

Using  $Z = \sqrt{\frac{\pi}{K}}$ , this becomes

$$\approx \text{erfc}(d\sqrt{K}),$$

and by employing the Chernoff–Rubin bound [50,51]  $\text{erfc}(X) \leq e^{-X^2}$

$$\Omega \leq e^{-d^2 K}, \quad (22)$$

which is equivalent to

$$\Omega \leq e^{-\frac{\Delta U_{\text{reorg}}}{4}}. \quad (23)$$

Notably, the Chernoff–Rubin bound overestimates the error function, which partially cancels the error that arises from using a step function instead of the Fermi function. We also provide an alternative way to derive this result in the Appendix 1.

Thus, the variance of the BAR free energy estimate for two shifted harmonic oscillators is

$$\sigma_{\text{BAR}}^2 \approx \frac{2}{n} \left( e^{\frac{\Delta U_{\text{reorg}}}{4}} - 1 \right). \quad (24)$$

An interesting feature of Equation 24 is that it is again absolutely independent of the underlying force constant or partition function and can be estimated based on the result of a simple energy minimisation. A comparison between theory and simulation is given at the bottom of Figure 7 in the Results section for verification. Readers that are not necessarily interested in multi-scale simulations should feel free to skip the next section and continue reading at Section 1.9.

### 1.7. Comparison of the efficiency of the Zwanzig equation and BAR for multi-scale free energy calculations

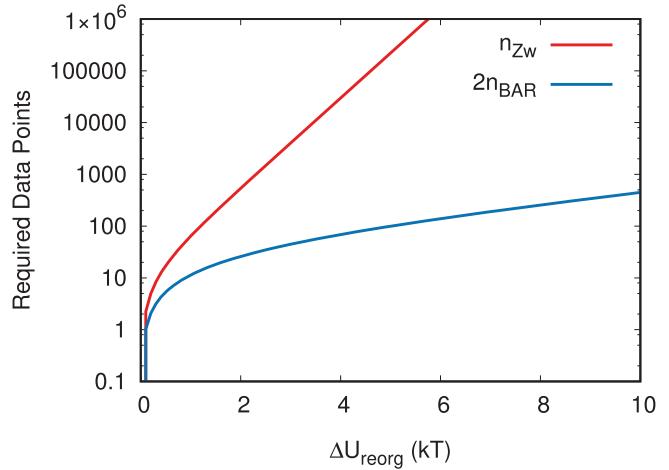
A fundamental question of multi-scale free energy calculations is whether it is more efficient to reweight trajectories from a lower level of theory (corresponding to the Zwanzig equation) or whether one should perform additional sampling on the high level of theory energy surface (corresponding to the use of BAR). The answer to this question depends on details of the method and the target system. Let us assume that the computational costs of MM potential energy evaluations are  $c_{\text{MM}}$  (for example in CPU time), while the computational costs of a QM potential energy evaluation are  $c_{\text{QM}}$ . The sampling algorithm and the roughness of the underlying energy surface also have an impact on the choice, since consecutive potential energy evaluations are not statistically independent. We therefore assume that  $\tau$  energy evaluations are required to produce statistically independent data points, where  $\tau$  represents the relevant relaxation time in terms of necessary energy evaluations.

Equation (17) provides us with an estimate of the required number of data points for the Zwanzig equation ( $n_{\text{Zw}}$ ) to obtain a certain precision ( $\sigma_p$ ) given a particular  $\Delta U_{\text{reorg}}$ , with

$$n_{\text{Zw}} = \frac{e^{2\Delta U_{\text{reorg}}} - 1}{\sigma_p^2}. \quad (25)$$

Thus, the total computational costs,  $C$ , for employing the Zwanzig equation in multi-scale free energy calculations to reach a certain level of precision are

$$C_{\text{Zwanzig}} = n_{\text{Zw}}(\tau c_{\text{MM}} + c_{\text{QM}}), \quad (26)$$



**Figure 4.** (Colour online) Required number of data points to reach a variance of 0.1 kT using the Zwanzig equation (red) and BAR (blue).  $n_{\text{BAR}}$  is multiplied with 2 since we assume that equal amounts of sampling are performed for the energy surfaces of the initial and the final state.

where  $n_{\text{Zw}}\tau c_{\text{MM}}$  represents the costs of the MM simulation and  $n_{\text{Zw}}c_{\text{QM}}$  represents the costs of the post-processing with QM.

Similarly, the required number of data points for BAR based on Equation (24) is

$$n_{\text{BAR}} = \frac{2}{\sigma_p^2} \left( e^{\frac{\Delta U_{\text{reorg}}}{4}} - 1 \right) \quad (27)$$

with the additional requirement that approximately  $n_{\text{BAR}}$  independent data points have to be sampled from both the initial state (MM) and the final state (QM).<sup>1</sup> A comparison of the number of data points to reach a certain level of precision for the Zwanzig equation and BAR is given in Figure 4. The figure highlights that BAR requires significantly fewer data points than the Zwanzig equation to reach a certain level of precision (unless the difference between the systems is very small). The total computational costs,  $C_{\text{BAR}}$ , for employing a BAR-like approach in multi-scale free energy calculations are

$$C_{\text{BAR}} = n_{\text{BAR}}(\tau c_{\text{MM}} + c_{\text{QM}} + \tau c_{\text{QM}} + c_{\text{MM}}), \quad (28)$$

since sampling ( $n_{\text{BAR}}\tau c$ ) and post-processing ( $n_{\text{BAR}}c$ ) have to be performed on both energy surfaces.

Correspondingly, the relative computational costs (Rel. Costs) of performing multiscaling by post-processing of MM trajectories vs. performing additional sampling of the QM energy surface is

$$\text{Rel. Costs} = \frac{C_{\text{BAR}}}{C_{\text{Zwanzig}}} = \frac{n_{\text{BAR}}}{n_{\text{Zw}}} \frac{\tau c_{\text{MM}} + c_{\text{QM}} + \tau c_{\text{QM}} + c_{\text{MM}}}{\tau c_{\text{MM}} + c_{\text{QM}}} \quad (29a)$$

$$= \frac{n_{\text{BAR}}}{n_{\text{Zw}}} \left( 1 + \frac{\tau c_{\text{QM}} + c_{\text{MM}}}{\tau c_{\text{MM}} + c_{\text{QM}}} \right), \quad (29b)$$

where Rel. Costs  $> 1$  indicate that post-processing is more worthwhile than performing additional simulations on the other energy surface. The equation becomes shorter when introducing the actually more relevant relative computational costs of the QM method compared to the MM method,  $c^* = \frac{c_{\text{QM}}}{c_{\text{MM}}}$ ,

$$\text{Rel: Costs} = \frac{n_{\text{BAR}}}{n_{\text{Zw}}} \left( 1 + \frac{\tau c^* + 1}{\tau + c^*} \right). \quad (30)$$

For example, if the computational costs of both energy surfaces are the same, i.e.  $c^* = 1$ , the expression in the parenthesis on the right side of the equation becomes two, which boils down to the ratio of the two curves shown in Figure 4.

### 1.8. The effect of parallelisation

Time is the most precious resource at our disposal. However, the equation above does not properly reflect this aspect, since it neglects the possible parallelisation of tasks. The ability to parallelise is also of eminent importance given modern multicore hardware architectures. The effect of parallelisation is commonly included via a speed up factor,  $\eta$ , which is related to Amdahl's law [52] via

$$\eta = (1 - f_p) + \frac{f_p}{n_p}, \quad (31)$$

where  $f_p$  is the fraction of the computational costs that can be parallelised and  $n_p$  is the number of processors that are available. The speed up factor is usually different for the sampling method ( $\eta_s$ ) and the post-processing of existing trajectories ( $\eta_p$ ), owing to the fact that the energy evaluations for the sampling have to be performed sequentially (e.g. in molecular dynamics simulations). On the other hand, the post-processing of existing trajectories is embarrassingly parallel. To simplify the equations, we define a relative speedup factor,  $\eta^* = \frac{\eta_p}{\eta_s}$ , which indicates in how far the post-processing is more parallelisable than the sampling stage. In most cases,  $\eta^*$  is proportional to the inverse of the number of processors ( $\frac{1}{n_p}$ ).

Including the speed up factors from parallelisation into Equation (30), and focusing only on the time limiting steps, leads to the relative wall clock time to perform the tasks

$$\text{Rel: Time} = \frac{n_{\text{BAR}}}{n_{\text{Zw}}} \left( 1 + \frac{\tau c^* + \eta^*}{\tau + \eta^* c^*} \right). \quad (32)$$

This equation already hints that for high levels of parallelisation the costs of the post-processing can be neglected (i.e.  $\eta \rightarrow 0$  and  $\eta^* c^* \rightarrow 0$ ) and the effect of  $\tau$  cancels out between the numerator and denominator, leaving just the bare relative costs ( $c^*$ ) to compete with the relative efficiencies of BAR and the Zwanzig equation ( $\frac{n_{\text{BAR}}}{n_{\text{Zw}}}$ ).

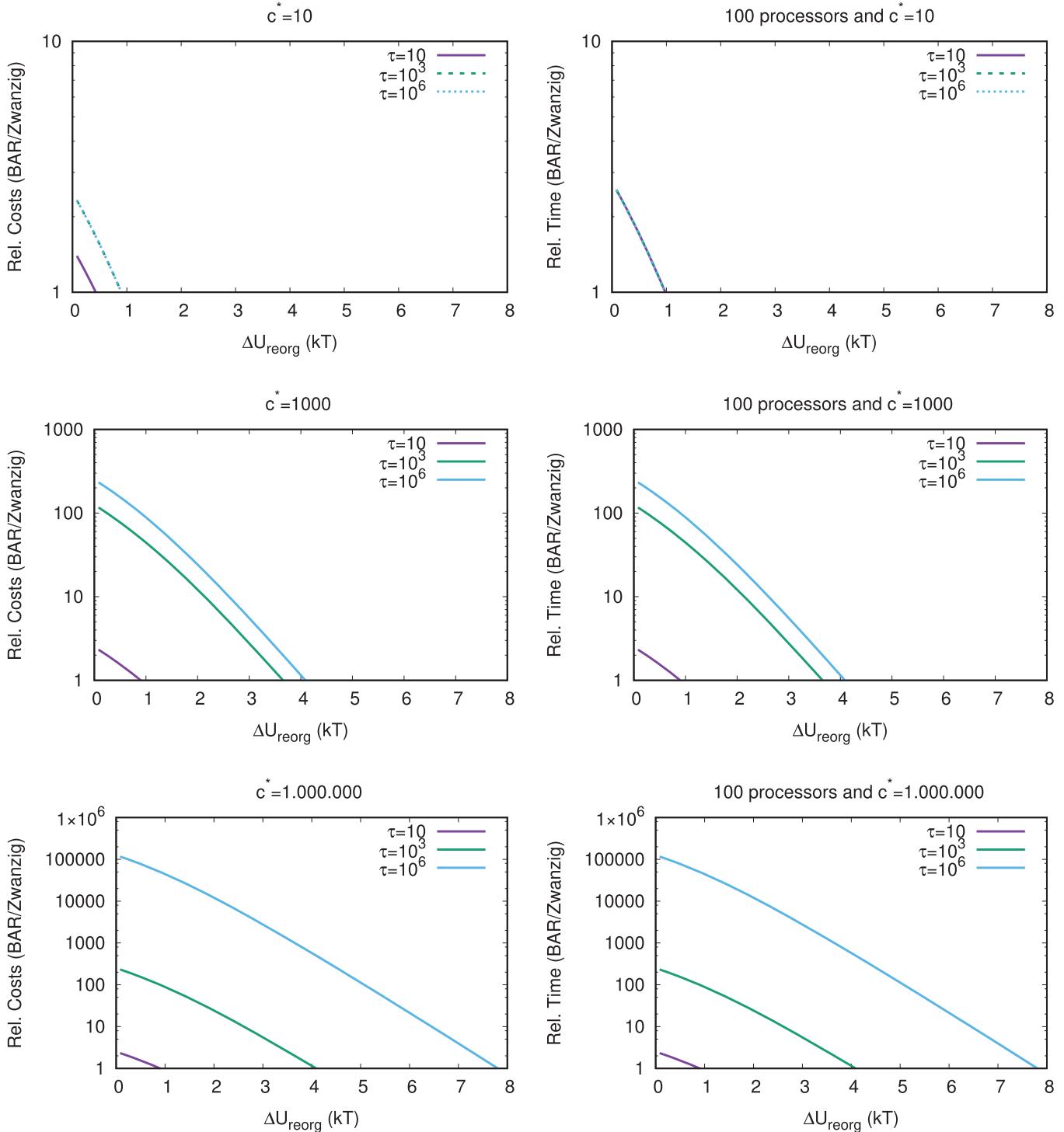
The effect of parallelisation is illustrated in Figure 5, where the left side shows the use of a single processor (where the wall-clock time corresponds to the total computational costs), while the right side shows the required wall-clock time to obtain sufficiently converged results when using 100 processors (i.e.  $\eta_p = \frac{1}{100}$ ). The three rows represent different relative costs of MM and QM ( $c^* = 1, 1000, 1,000,000$ ). Each plot shows the relative efficiency as a function of the reorganisation energy of our model system. While BAR is more efficient than the Zwanzig equation, the possibility to parallelise the post-processing and the requirement to produce independent samples favours the Zwanzig equation in some cases. For example, for  $\tau = 10$  and  $c = 10$  (the violet line on the left side of the top row), the Zwanzig equation is more efficient if  $\Delta U_{\text{reorg}}$  is very

small ( $< \frac{kT}{2}$ ). However, by exploiting parallelisation, it can out-compete BAR until a  $\Delta U_{\text{reorg}}$  of ca.  $1 \text{ kT}$ . Notably, there is a limit to the benefit of parallelisation, since the performance strongly depends on the relative computational costs between MM and QM, as illustrated by the different rows. Since parallelisation is mostly limited to the post-processing, while the costs are dominated by the time spent on sampling, it often has only a minor impact. If the relative costs are increased to  $c^* = 1000$  (middle), the Zwanzig equation performs better until a  $\Delta U_{\text{reorg}}$  of ca.  $1\text{-}4 \text{ kT}$ , depending on the autocorrelation time of the system. Because of the exponential relationship between the number of required samples and  $\Delta U_{\text{reorg}}$ , the relaxation time  $\tau$  has only a limited impact on the relative efficiency if the relative computational costs are small. This is highlighted by the similarity between the plots of  $\tau = 10^3$  and  $\tau = 10^6$  in the top row, where the lines overlap. The effect of  $\tau$  is amplified by the relative costs  $c^*$ , as highlighted by the larger gaps between the different  $\tau$  as the computational costs increase. However, it is also possible to turn this argument around and say that methods that decrease  $\tau$  become more relevant as the computational costs increase. Thus, the development of improved sampling algorithms that can decrease the autocorrelation time strongly favours the use of simulations at the higher level of theory.

To summarise, Equation (32) provides a rule of thumb to determine whether multi-scaling via post-processing is worthwhile or not ( $E_{\text{MS}} > 1$  indicating that post-processing is preferable over additional sampling of the QM energy surface). The main component of the underlying Equations (25) and (27) is an energy minimisation, which determines the potential energy difference between an MM-optimised structure and a QM-optimised structure ( $\Delta U_{\text{reorg}}$ ). The same energy minimisation can also be used to estimate the relative computational costs of QM compared to MM ( $c^*$ ). In addition, one can also include an estimate of the relevant relaxation time of the system in terms of the number of potential energy evaluations to reach statistical independence ( $\tau^*$ ). Finally, the rule of thumb incorporates the parallelisability of the post-processing ( $\eta_p$ ), which is mainly determined by the number of available processors. The data indicate that some form of Amdahl's law of diminishing returns also applies to the post-processing of trajectories, leaving the relative costs of MM and QM ( $c^*$ ) and the potential energy difference  $\Delta U_{\text{reorg}}$  as the main criteria. If the relative computational costs are low, or if  $\Delta U_{\text{reorg}}$  is too high, it is more efficient to run additional simulations on the QM energy surface and employ BAR. On the other hand, post-processing MM trajectories in parallel can yield significant gains under the right conditions.

### 1.9. On the introduction of intermediate states

While simulating so-called  $\lambda$  intermediate states increases the computational costs, they can also significantly increase the phase space overlap, and, therefore, the convergence of free energy calculations [4,12,17]. However, it is often unclear how many intermediate states are necessary in order to reach convergence. Thus, we employ our linear response model system to obtain an approximate solution for the optimal number of  $\lambda$  steps in a free energy calculation ( $n_{\text{steps}}^{\text{opt}}$ ). For simplicity, we are



**Figure 5.** (Colour online) Comparison of the efficiency of post-processing vs. additional sampling according to the relative computational costs (Rel. Costs in Equation (30), shown on the left side), and the relative required wall clock time (Rel. Time) based in Equation (32) when using 100 processors, shown on the right side. Rel. Costs or Rel. Time  $> 1$  indicate that post-processing based on the Zwanzig equation/reweighting is more efficient than employing BAR. The point where the relative costs become 1 signifies that sampling on the more expensive energy surface is more efficient. The similarity between the left and the right side indicates that parallelisability only plays a significant role as long as the computational costs of the end states are comparable. The top row uses relative computational costs between QM and MM of  $c^* = \frac{c_{QM}}{c_{MM}} = 10$ , while the middle row is created with  $c^* = 1000$ , and the bottom row used a  $c^* = 1,000,000$ . As the computational costs increase, the Zwanzig equation becomes more efficient, but only as long as the reorganisation energy is on the order of a couple of  $kT$ . This highlights the need for proper balance between accuracy and speed.

assuming again that the computational costs are the same for the two end points of each sub-step.

### 1.9.1. Zwanzig equation

Starting with the Zwanzig equation, and assuming that the intermediate states are distributed evenly in  $\Delta U_{reorg}$  space, the relative computational costs of dividing the free energy calculation into  $n_{step}$  sub-steps, are

$$\frac{n_{Zw}(\Delta U_{reorg}, n_{step})}{n_{Zw}(\Delta U_{reorg})} = \frac{e^{2\frac{\Delta U_{reorg}}{n_{step}}} - 1}{e^{2\Delta U_{reorg}} - 1} n_{step}.$$

Assuming that  $e^{2\Delta U_{reorg}} \gg 1$ , this can be approximated by

$$\frac{n_{Zw}(\Delta U_{reorg}, n_{step})}{n_{Zw}(\Delta U_{reorg})} \approx \frac{e^{2\frac{\Delta U_{reorg}}{n_{step}}}}{e^{2\Delta U_{reorg}}} n_{step} \approx e^{2\Delta U_{reorg}\left(\frac{1}{n_{step}} - 1\right)} n_{step} \quad (33)$$

Setting the partial derivative of Equation (33) with respect to  $n_{steps}$  to zero, i.e.

$$\frac{\partial \frac{n_{Zw}(\Delta U_{reorg}, n_{step})}{n_{Zw}(\Delta U_{reorg})}}{\partial n_{step}} \approx e^{2\Delta U_{reorg}\left(\frac{1}{n_{step}} - 1\right)} \left(1 - \frac{2\Delta U_{reorg}}{n_{step}}\right) = 0,$$

leads to the optimal choice of the number of  $\lambda$  steps with the Zwanzig equation,

$$n_{step}^{opt ZW} \approx 2\Delta U_{reorg}. \quad (34)$$

Notably, the necessary number of  $\lambda$  points with the Zwanzig equation can be further reduced by employing so-called double-wide sampling (DWS) [4,23]. In DWS, the interval between two simulated  $\lambda$ -points,  $\lambda_i$  and  $\lambda_{i+1}$ , is divided in half, and a virtual state  $\lambda_{i+\frac{1}{2}}$  is introduced at the midpoint of the interval. Since the virtual midpoint is not simulated, the computational costs only increase by the extra post-processing for  $\lambda_{i+\frac{1}{2}}$ . Since we can also neglect the costs of post-processing due to parallelisation, this leads to

$$\frac{n_{DWS}(\Delta U_{reorg}, n_{step})}{n_{Zw}(\Delta U_{reorg})} = \frac{e^{2\frac{\Delta U_{reorg}}{2n_{step}}} - 1}{e^{2\Delta U_{reorg}} - 1} (n_{step} + 1),$$

and

$$\begin{aligned} & \frac{\partial \frac{n_{DWS}(\Delta U_{reorg}, n_{step})}{n_{Zw}(\Delta U_{reorg})}}{\partial n_{step}} \\ & \approx e^{2\Delta U_{reorg}\left(\frac{1}{n_{step}} - 1\right)} \frac{n_{step}^2 - 2\Delta U_{reorg}n_{step} - 2\Delta U_{reorg}}{n_{step}^2} \end{aligned}$$

which leads to the optimal number of steps with double-wide sampling,

$$n_{step}^{opt DWS} \approx \Delta U_{reorg} + \sqrt{(\Delta U_{reorg})^2 + 2\Delta U_{reorg}}. \quad (35)$$

This means that the optimal step size increases as the problem becomes more challenging. For example, if the reorganisation energy is so high that 10  $\lambda$  steps have to be used, the optimal

step size in terms of  $\Delta U_{reorg}$  is ca. 4.5 kT. Since the precision is supposed to be the same, and the step size of  $\Delta U_{reorg}$  increases, the only way to achieve the same precision is an increase of the number of steps. Therefore, it follows that it is sometimes more efficient to run longer simulations than to introduce more intermediate steps. This is not necessarily an intuitive result.

### 1.9.2. Bennett's acceptance ratio

Similarly, the relative computational costs of dividing a BAR free energy calculation into  $n_{step}$  steps are

$$\frac{n_{BAR}(\Delta U_{reorg}, n_{step})}{n_{BAR}(\Delta U_{reorg})} \approx \frac{e^{\frac{\Delta U_{reorg}}{4n_{step}}} - 1}{e^{\frac{\Delta U_{reorg}}{4}} - 1} \frac{n_{step} + 1}{2},$$

which, if  $e^{\frac{\Delta U_{reorg}}{4}} \gg 1$ , can be approximated by

$$\begin{aligned} \frac{n_{BAR}(\Delta U_{reorg}, n_{step})}{n_{BAR}(\Delta U_{reorg})} & \approx \frac{e^{\frac{\Delta U_{reorg}}{4n_{step}}}}{e^{\frac{\Delta U_{reorg}}{4}}} \frac{n_{step} + 1}{2} \\ & \approx e^{\frac{\Delta U_{reorg}}{4}\left(\frac{1}{n_{step}} - 1\right)} \frac{n_{step} + 1}{2}. \end{aligned} \quad (36)$$

Setting the partial derivative of Equation (36) with respect to  $n_{steps}$  to zero, i.e.

$$\begin{aligned} & \frac{\partial \frac{n_{BAR}(\Delta U_{reorg}, n_{step})}{n_{BAR}(\Delta U_{reorg})}}{\partial n_{step}} \\ & \approx e^{\frac{\Delta U_{reorg}}{4}\left(\frac{1}{n_{step}} - 1\right)} \frac{4n_{step}^2 - \Delta U_{reorg}n - \Delta U_{reorg}}{8n_{step}^2} = 0 \end{aligned}$$

leads to the optimal choice of the number of  $\lambda$  steps with BAR,

$$n_{step}^{opt BAR} \approx \frac{1}{8} \left( \Delta U_{reorg} + \sqrt{\Delta U_{reorg}} \sqrt{\Delta U_{reorg} + 16} \right). \quad (37)$$

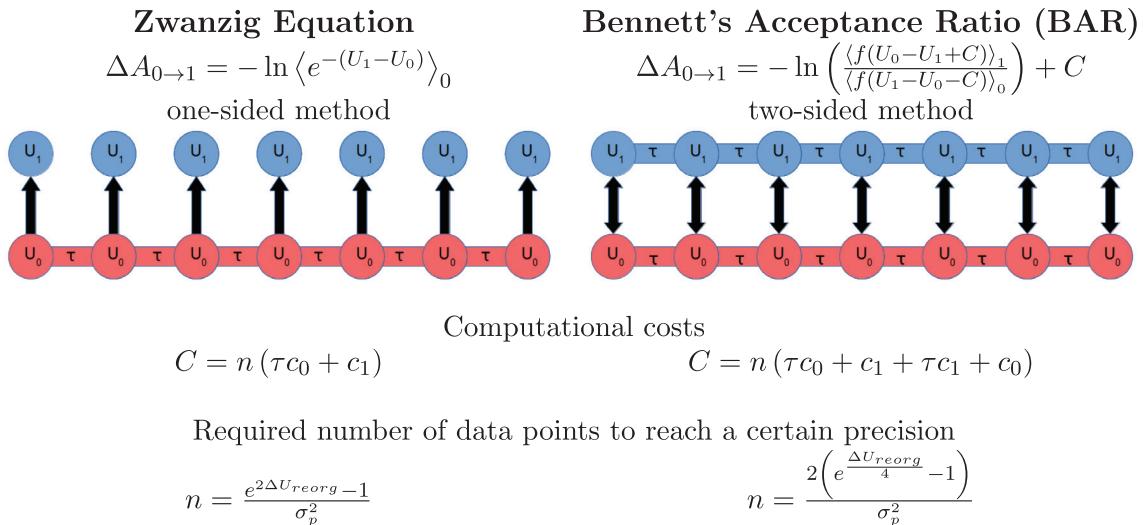
This indicates that the optimal step size in terms of  $\Delta U_{reorg}$  again increases as the reorganisation energy gap widens. Thus, it is again necessary to increase the number of steps to achieve the required precision. The main results of our comparison are shown in Figure 6.

## 2. Methods

All calculations were conducted with CHARMM [53,54], using the CHARMM General Force Field [55] unless stated otherwise. The time step was 1 fs. No cut-off radius was employed. Free energy differences were evaluated with the FREN module of CHARMM.

### 2.1. Simulations of harmonic model systems

Following the same guidelines as provided in Refs. [56–58], two linear model systems were created, one consisting of 10 atoms and one consisting of 51 atoms. The molecules were simulated in the gas phase using Langevin dynamics with a friction coefficient of 5 ps<sup>-1</sup> and a temperature of 300 K. The simulation length was 100 ns, and coordinates were saved every 1000 steps. The initial



**Figure 6.** (Colour online) Comparison of the two main free energy methods, using reduced units of  $kT$ . The initial state is shown in red and the final state is represented in blue. Potential energy evaluations are marked by arrows and trajectories are represented by crossbars.  $\tau$  refers to the saving frequency of the trajectory, which should correspond to the relevant auto-correlation time of the process, in order to generate statistically independent samples.

state featured bond lengths of 1.53 Å and bond angles of 113.6°, which mimics a CT2 alkane chain. The force constants were 222 kcal/mol/Å² and 58.350 kcal/mol/rad². Mutations were performed by changing the geometrical parameters of 1 to 50 atoms. The mutations involved bond length changes of 0.001, 0.005, 0.01, 0.05 and 0.1 Å for bonds, or 0.1, 0.5, 1, 5, 10 and 20° for angles. The variances were obtained from 10 repetitions of each simulation. In several cases, the BAR results failed to converge because of lacking overlap. For this reason, the BAR data in Figure 7 is more sparse for high  $\Delta U_{reorg}$  than the Zwanzig data.

## 2.2. Gas phase free energy simulations of twelve molecules

Based on a previous study [59], 12 molecules were used: water, methanol, ethanol, methanethiol, acetamide, tetrahydrofuran, benzene, aniline, phenol, ethane, n-hexane and cyclohexane. Each molecule was simulated in the gas phase using Langevin dynamics with a friction coefficient of 1 ps⁻¹ and a temperature of 300 K. The simulation time scale was 500 ns and coordinates were saved every 20 ps.  $\lambda$ -Hamiltonian Replica Exchange [60] was employed to enhance sampling by exchanging structures between neighbouring  $\lambda$  points every 20,000 steps based on the REPD module of CHARMM. The free energy calculation was broken down into two parts. First, the charges were scaled to zero in three steps ( $\lambda = 0.0, 0.2, 0.55$  and 1.0). Second, the van der Waals interactions were turned off using  $\lambda = 0.0, 0.15, 0.30, 0.45, 0.60, 0.75, 0.87, 0.96$  and 1.00. Soft cores, as implemented with the PSSP command in the PERT module of CHARMM, were used with the default parameters to avoid the end point problem [53,61]. Free energy differences were calculated between all possible combinations of  $\lambda$  states. Standard deviations were calculated from four repetitions of each free energy simulation.  $\Delta U_{reorg}$  values were determined based on the potential energy difference between the minimised structures of the respective end points using 10,000 steps of the adopted basis Newton–Raphson method [53].

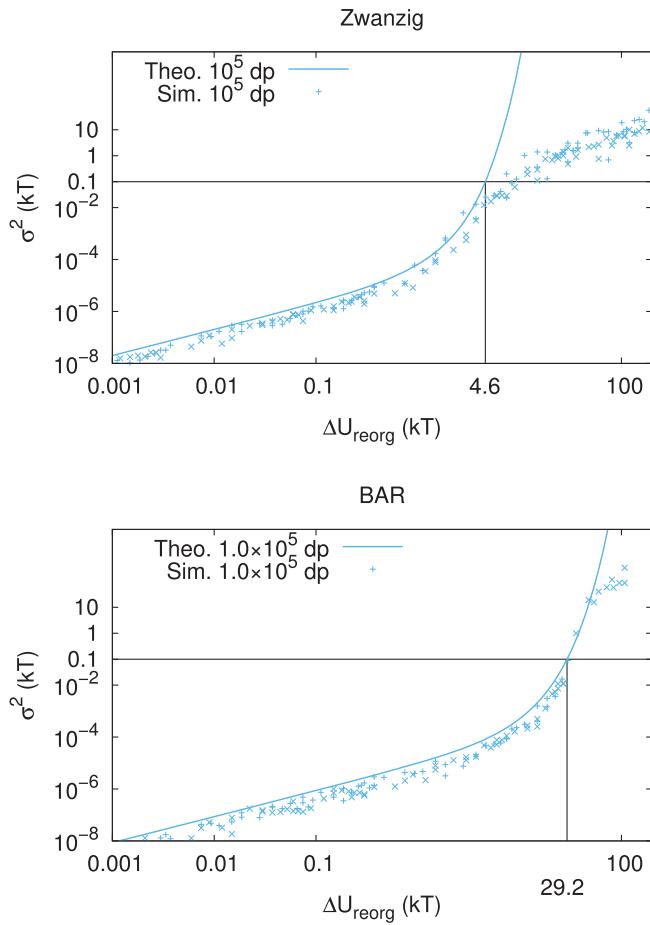
The last 18,000 frames of the physical end points in the gas phase were employed for multi-scale free energy simulations between MM and QM. QM calculations were performed with Q-Chem [62] based on the CHARMM/Q-Chem interface [63]. In particular, potential energy differences were evaluated with four different methods using the 6-31G\* basis set: [64] BLYP [65, 66], B3LYP [67], Hartree–Fock and M06-2X [68,69]. The MM-optimised structure was obtained with 1000 steps of steepest decent minimisation, followed by 1000 steps with the adopted basis Newton–Raphson method. The MM-optimised structure was then further minimised with the respective QM method, using 450 steps with the adopted basis Newton–Raphson method. The SCF convergence criterion for the gradient calculations was set to  $10^{-10}$ .  $\Delta U_{reorg}$  was calculated from the potential energy difference between the MM and QM optimised structure while using the MM Hamiltonian.

## 3. Results and Discussion

### 3.1. Verification with gas phase simulations

Because Equations (17) and (24) are of fundamental importance for our analysis, caution dictates to verify their results based on actual simulations. Although the equations are derived analytically, they can include some approximations that we are not aware of. For this reason, we resort to linear model systems that only include bonded terms, similar to the work by Boresch and Karplus in Ref. [57]. The linear response approximation should be valid for the benchmark systems because no non-bonded interactions and no force constants are changed.

The top panel of Figure 7 shows a comparison of the predicted variance of the free energy estimate based in Equation (17) (blue line) with variances of actual free energy simulations. The corresponding free energy simulations include both changes of bond lengths (+) and changes of bond angles (×). This is relevant, because the equation dictates that the variance only depends on the reorganisation energy  $\Delta U_{reorg}$ , but not on



**Figure 7.** (Colour online) Variance of free energy estimates between shifted harmonic oscillators based on the Zwanzig equation (top) and BAR (bottom).  $\Delta U_{\text{reorg}}$  indicates the potential energy difference between the energy minimum of the initial state and the energy minimum of the final state while using the Hamiltonian of the initial state. Each free energy simulation was conducted with 100,000 potential energy differences (blue). + signs indicate results for mutations of the bond length, while x signs indicate changes of bond angles. The theoretical estimates of the standard deviation based in Equation (15) are shown for 100,000 data points (dp). Interestingly, the correlation between the simulated results and the theoretical results seems to break down once the variance reaches ca. 0.1 kT. A similar behaviour was also observed by Gore et al. in Figure 3A of Ref. [70]. This can probably be attributed to sampling errors. Notably, BAR can handle significantly larger  $\Delta U_{\text{reorg}}$  than the Zwanzig equation. This is highlighted by the  $\Delta U_{\text{reorg}}$  that leads to a variance of 0.1 kT (black horizontal line), which corresponds to 4.6 kT for the Zwanzig equation and 29.2 kT for BAR.

the details of the force constants or the partition function. Thus, the equation should be equally valid for changes with a high force constants (like bond stretching) and lower force constants (like angle bending). The data in Figure 7 indicates that this is actually the case since there are no significant discrepancies between the two kinds of mutations. Also, in our mutations, we change one or more bonds at the same time, but this does not seem to affect the accuracy. Therefore, the data suggest that no normal mode analysis is required to make use of Equation (17), which renders its application very simple.

The top panel of Figure 7 makes apparent that the theoretical prediction of the variances is only valid for relatively small differences. As the variance approaches ca. 0.1 kT, which corresponds to a  $\Delta U_{\text{reorg}}$  of ca. 4.6 kT, there are clear discrepancies between the theoretical variances and the variances obtained from simulation. While the theoretical variance estimates grow exponentially, the simulated variances seem to plateau around

ca. 10 kT. It is expected that the simulated data will not perfectly agree with the theoretical result due to the finite number of repetitions that were used to estimate the uncertainty, which leads to an uncertainty of the uncertainty. However, the deviations are too large to be explained by random noise. One possible explanation is that the theoretical results assume perfect sampling, since they are directly obtained from a transformation of the Gaussian probability distribution of the micro-states. The simulated data, on the other hand, samples states only according to their Boltzmann probability and in finite time (100 ns). Thus, most of the samples reside close to the energy minimum of the initial state, while the tails of the probability distribution that are responsible for the large variance are not adequately sampled. This introduces a bias in the variance estimate, which could be addressed by employing the ‘neglected-tail’ bias model by Wu, Lu and Kofke, which was also recently discussed by Boresch and Woodcock [1,5,71]. Another possibility is the use of Umbrella Integration [72] or variational free energy profile calculations [73]. In any case, the observed sampling problems for harmonic systems in Figure 7 should serve as warning that simulated estimates of the variance are not necessarily trustworthy, since they might be affected by sampling problems.

To quantify the agreement, we have to resort to the decadic logarithm of the variances, since we have to evaluate values over several orders of magnitude. For the Zwanzig equation results, we exclude variances above 0.1 kT from this analysis, because they might be affected by sampling problems. The Pearson correlation coefficient of the remaining data with theory is very good with  $R^2 = 0.97$ . The mean signed deviation of  $\text{Log}_{10} \sigma_{\text{Theory}}^2$  and  $\text{Log}_{10} \sigma_{\text{Simulation}}^2$  is 0.36, which indicates that theory slightly overestimates the variance by a factor of ca. 1.4. The other possibility is that the simulations systematically underestimate the variance because of the limited number of repetitions (10). The root mean square deviation is 0.38, which is very similar to the mean signed deviation. This indicates that there are no major outliers in the considered range.

Turning our attention to the BAR results at the bottom of Figure 7, we do not observe discrepancies between theory and simulation, as in the case of the Zwanzig equation. However, this does not necessarily indicate that there are no sampling problems. A closer inspection of the BAR results relative to the Zwanzig results shows that the data for BAR is significantly more sparse above a  $\Delta U_{\text{reorg}}$  of ca. 29 kT. This is caused by failure of BAR to converge if the phase space overlap between the end points is too low. Without free energy results, it is not possible to calculate a variance, and, therefore, those free energy simulations are not included. While the failure of BAR to converge can be frustrating for the user, it is ultimately beneficial, because it actively forces the user to employ a proper  $\lambda$  protocol by introducing more intermediate states. In this regard, BAR can be considered far more *user friendly* than the Zwanzig equation or Thermodynamic Integration, which do not provide an equivalent implicit quality control of the free energy results. However, as can be seen from the few outliers with variances higher than 0.1 kT (which, interestingly, all arise from angle mutations with lower force constants), BAR sometimes manages to converge, even if the resulting variance is poor. For this reason, the BAR implementation in the FREN module of CHARMM

is even more restrictive and refuses to provide a free energy estimate if the overlap of the potential energy distributions is below 1%. Leaning on Equation (19), the estimated phase space overlap is based on

$$\Omega^* = \int \frac{2p_0(\Delta U)p_1(\Delta U)}{p_0(\Delta U) + p_1(\Delta U)} d\Delta U, \quad (38)$$

where, in practice, the probabilities are estimated with histograms of the potential energy differences. The cutoff of 1% is based on our past experiences with BAR [3,4,74–76], but with Equation (23) it is possible to relate this overlap to a  $\Delta U_{reorg}$  of ca. 18 kT, which corresponds to a standard deviation of 0.25 kcal/mol if merely 1000 data points are used in the free energy step. We, therefore, think that this quality criterion is not overly restrictive.

With the available BAR data, the correlation coefficient between the decadic logarithms of the variance of theory and simulation is very good with  $R^2 = 0.97$ . The mean signed deviation between theory and simulation is 0.35, and the root mean square deviation is 0.38. This, again, indicates that the theoretical estimate of the variance is higher than the one obtained from simulation, but this can be mostly explained by the uncertainty of the uncertainty. Since parts of the deviation could also arise due to the approximations introduced in Equations (21) and (22), we also provide an alternative derivation in the Appendix, which gives the same exponential behaviour, but corrected with a prefactor. While this alternative given in Equation (A2) is more accurate, we still favour Equation (17) because it simplifies the ensuing derivations.

Even a casual comparison of the BAR data with the Zwanzig data in Figure 7 reveals the fundamental superiority of BAR when it comes to bridging energy gaps between two states. Taking as an example a reasonable variance of 0.1 kT that has to be achieved with 10,000 potential energy difference samples (highlighted by the black horizontal line), the Zwanzig equation can overcome a  $\Delta U_{reorg}$  of 4.6 kT. With the same requirements, BAR can cross a  $\Delta U_{reorg}$  of almost 30 kT. In practice, this means that BAR only requires a fraction of the number of  $\lambda$  points compared to the Zwanzig equation. Taking our estimates from Equations (34) and (37), the number of  $\lambda$  points can be significantly reduced, which saves computer time.

### 3.2. Gas phase free energy simulations of 12 molecules

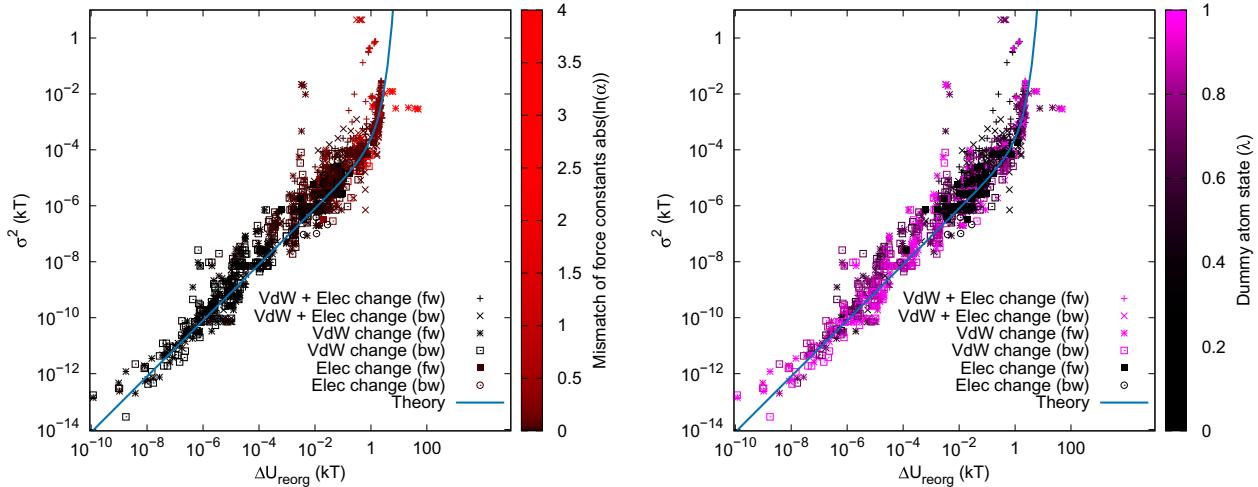
By employing our estimates of the variance of free energy simulations of twelve molecules in the gas phase, we want to highlight some of the limitations of the approach. There are several assumptions that are involved in the estimate, including: (a) The system under consideration is harmonic. (b) The structure that is used for  $\Delta U_{reorg}$  is representative for the whole energy landscape (i.e. it is the global free energy minimum and the other minima contribute only marginally to the free energy difference). (c) The force constants involved in both end states are approximately the same. Much like the model of the ‘homo oeconomicus’ in macroeconomics, those underlying assumptions are very weak and almost never fulfilled in practice. However, it can be illustrative from an academic point of view to observe

how the linear response toy system fares when compared to data from real molecules.

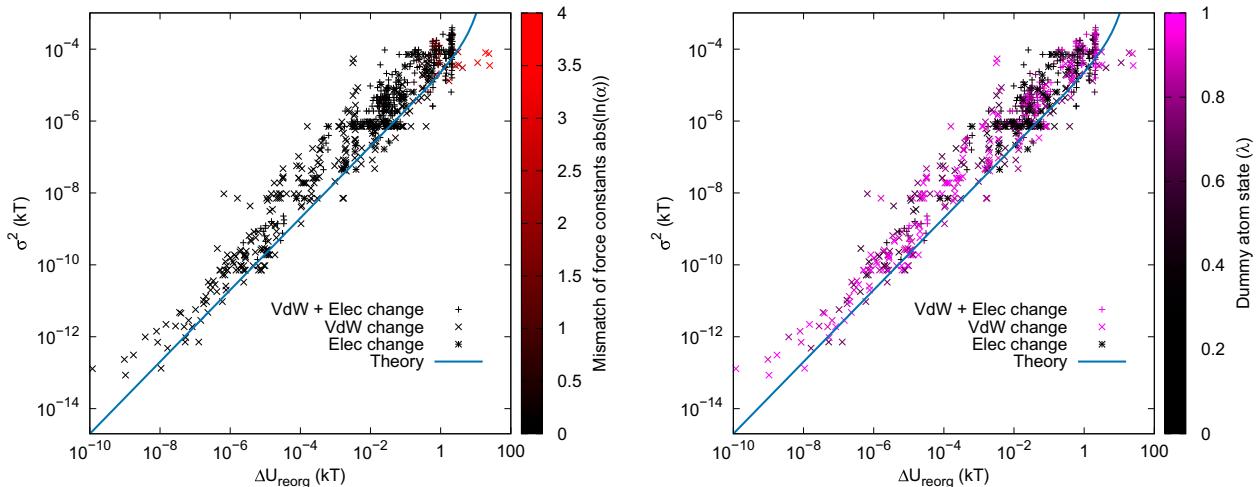
The variances of free energy differences for the twelve molecules under consideration (water, methanol, ethanol, methanethiol, acetamide, tetrahydrofuran, benzene, aniline, phenol, ethane, n-hexane and cyclohexane) are shown in Figure 8. In particular, we simulate one leg of solvation free energy calculations, where the non-bonded interactions of a molecule are sequentially turned off. In our simulations, electrostatic interactions are turned off by scaling the charges to zero (denoted by ‘Elec change’). In addition, the van der Waals interactions are scaled to zero (‘VdW change’). To reach higher values of  $\Delta U_{reorg}$ , also combinations that change both electrostatic and van der Waals interactions at the same time are considered (‘VdW + Elec change’). We also differentiate between calculations that were performed in the ‘forward’ direction (deletion) and in the ‘backward’ direction (insertion), since it is expected that the insertion direction should be more relevant [21]. *A priori*, it is expected that the linear response model works better for purely electrostatic changes.

While there are definitely several outliers in Figure 8, the results from the linear response approximation are not as pathetic as one would expect them to be, given the really crude underlying approximations. Like for the harmonic systems, we exclude systems with a variance higher than 0.1 kT because of the expected sampling problems. We also exclude variances below  $10^{-14}$  to avoid numerical problems. An evaluation of the decadic logarithms of the variances of the 1341 free energy results leads to an  $R^2 = 0.91$ , which is lower than the correlation coefficient for the harmonic systems, but still pretty good. The mean signed deviation is -0.32, which means that the variance is on average higher than predicted by linear response theory. The root mean square deviation is 0.82, which is twice as high as for the purely harmonic systems. However, since we are evaluating the decadic logarithm of the variance, this corresponds to a factor of ca. 6.6. If one tries to translate this ratio into an error in terms of  $\Delta U_{reorg}$  (ca. 0.94 kT), one ends up with the expected misestimate of the required number of  $\lambda$  points according to Equation (34) (1.8). Thus, if one tries to use the linear response approximation to estimate the number of  $\lambda$  points with the Zwanzig equation in the gas phase, the estimate can be expected to be wrong by about 2  $\lambda$  points. Since it is better to err on the side of caution than risking unconvolved sub-steps, one might consider adding two additional steps to the estimated  $\lambda$  protocol to account for the observed discrepancies.

Rather than indiscriminately adding more sub-steps, one can ask whether it is possible to identify particularly recalcitrant steps in the  $\lambda$  protocol. An obvious and highly recommended strategy is to conduct short simulations to see whether the predicted variances agree with the actual results. This accounts for both anharmonicity and for conformational entropy, which are not considered in the approximation. In addition, short equilibration simulations are required anyway before the actual production is started, and therefore this check does not really incur significant additional computational efforts. However, since we are aspiring to make the best use of the information that is available before trajectories are generated, we also propose two simple alternatives.



**Figure 8.** (Colour online) Variance of free energy estimates based on the Zwanzig equation of gas phase simulations of twelve molecules. The mutations consider electrostatic changes ('Elec change'), van der Waals changes ('VdW change') and combinations of both ('VdW + Elec change'). Free energy simulations are conducted both in the deletion direction ('fw') and the insertion direction ('bw').



**Figure 9.** (Colour online) Variance of free energy estimates based on BAR of gas phase simulations of twelve molecules. The mutations involve electrostatic changes ('Elec change'), van der Waals changes ('VdW change') and combinations of both ('VdW + Elec change').

Apart from our assumptions of harmonicity and of the starting structure being representative for the ensuing trajectory, the linear response approximation implies that both end points are governed by the same force constants. One possible way to verify this is to use normal mode analysis [53,77] or quasi-harmonic analysis [78]. However, some information about the difference of the force constants can also be extracted from the reorganisation energies. Since the distance between the energy minima of the two end points in terms of our reaction coordinate is the same, the ratio of the resulting reorganisation energies is the same as the ratio of the force constants, i.e.

$$\alpha = \frac{K_1}{K_0} = \frac{\Delta U_{\min}^{1 \rightarrow 0}}{\Delta U_{\min}^{0 \rightarrow 1}}, \quad (39)$$

where  $\Delta U_{\min}^{1 \rightarrow 0}$  means that the reorganisation energy is determined with the Hamiltonian of the final state, and  $\Delta U_{\min}^{0 \rightarrow 1}$  uses the potential energy function of the initial state. The resulting

ratio  $\alpha$  actually represents a measure for the vibrational entropy [57,79,80] since

$$\Delta S_{vib} \propto \ln \frac{K_1}{K_0} = \ln \alpha. \quad (40)$$

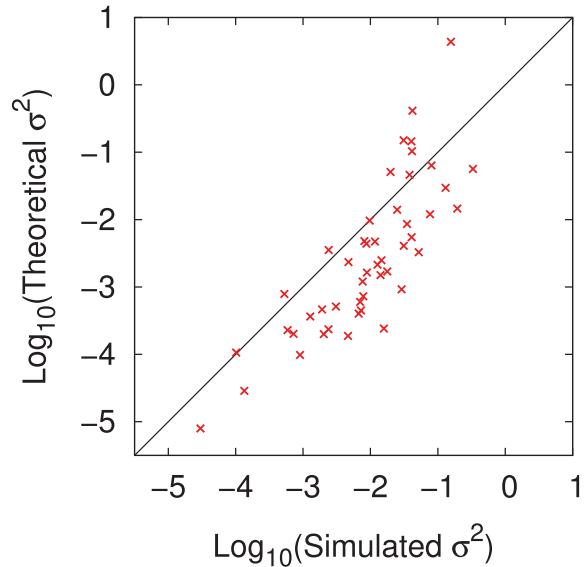
By identifying large changes of the entropy by colour coding the results in Figure 8 based on  $\ln \alpha$  (left side, where black represents small changes and red represents drastic changes), one can see that several major outliers can be explained by violations of the assumption that the end states have similar force constants. In some cases  $\ln \alpha$  is 4, which corresponds to a force constants ratio of ca. 55. Thus, the end states have very different phase spaces available. How is it possible to resolve this issue?

To some degree, it is possible to mitigate this problem by following the advice of Kofke and coworkers that free energy simulations are supposed to be performed in the insertion direction, as one should generally proceed from the higher entropy

state to the lower entropy state [1,2,21,24,81]. Focusing on a comparison of the data points marked in bright red in Figure 8, one can distinguish between free energy calculations that are performed in the forward direction ('fw', marked with stars), which corresponds to a deletion step here, and calculations that are performed in the backward direction ('bw', marked by squares), which correspond to an insertion. The outliers in the forward direction are fully visible and mark clear overestimations of the variance by the theoretical estimate (blue line). On the other hand, the backward processes (red squares marking the insertion) are more difficult to see, because they partially overlap with other data points and are significantly closer to the theoretical estimate. This indicates that the insertion process leads to a more reliable estimate. However, a closer inspection of some of the other outliers at the top of figure also shows that sometimes the backward estimates severely underestimate the variance. A possible compromise to reconcile the conflicting data is to consider the forward and backward processes as bounds to the proper estimate, and either use the mean of the two, or, as a kind of worst-case scenario, use the larger  $\Delta U_{reorg}$ .

Another factor that is not considered by our approximation is the use of so-called dummy atoms. Dummy atoms are placeholder atoms without any non-bonded interactions [57, 82]. Since no van der Waals interactions are calculated for dummy atoms, they do not cause steric clashes. However, when evaluating the potential energy difference to some other state, the dummy placeholder atoms are turned into normal atoms, which sometimes causes steric clashes, leading to the so-called end point catastrophe [53,61,83]. The effect of the end point catastrophe depends on the probability of steric clashes, when cross-evaluating potential energies with dummy states. Since the dummy-likeness depends on the  $\lambda$  value, the effect of dummy states can be observed by colour coding the results in Figure 8 based on  $\lambda$ , as is done on the right side with purple colour. In most cases, the use of dummy atoms leads to no noticeable effects, but this is only the case because our test set mostly consists of relatively small and stiff molecules. An interesting column of outliers can be observed at ca.  $10^{-2}$  kT where the variance is higher than predicted by linear response theory. All data points of this group correspond to hexane, the longest chain in the data set. In this case, the dummy state of hexane can assume any conformation that is possible in terms of the bonded terms, while the phase space of the non-dummy end state is restricted by excluded volume effects. The excluded volume is of course not considered by linear response theory, but probably could be treated with a form of Flory theory [84]. However, it would be difficult to generalise the resulting theory, therefore suffice it to say that the linear response approach may break down for dummy atoms, especially in the condensed phase.

Figure 9 shows the corresponding results for BAR. As already discussed in the context of forward and backward transformations with the Zwanzig equation, there is some ambiguity in terms of which  $\Delta U_{reorg}$  is to be used. This is especially acute in BAR, since BAR uses the potential energy distributions of both end states for its free energy estimate. As a simple *ad hoc* solution, we just employ the mean of the two  $\Delta U_{reorg}$  here. Again, the correlation coefficient is pretty good with  $R^2 = 0.92$ , but the mean signed deviation is  $-0.57$  and the root mean



**Figure 10.** (Colour online) Variances of multi-scale free energy simulations in the gas phase between the CGenFF force field and the quantum-mechanical methods BLYP, B3LYP, Hartree–Fock and M06-2X.

square deviation is 0.83. Thus, the systematic error in terms of underestimation of the resulting variances is higher than for the Zwanzig equation. Based on the root mean square deviation, this also leads to an underestimate of the required number of  $\lambda$  points by about two.

As for the Zwanzig equation, the left side of Figure 9 marks mismatching force constants between end points in red. While the BAR free energy estimate itself finds the optimal combination of the forward and reverse potential energy distributions, this is not case for our linear response approximation for estimating the variance based on  $\Delta U_{reorg}$ . This aspect will, therefore, require more attention in the future. The same can be said about the treatment of dummy atoms. Therefore, it is advisable to restrict the use of our linear response approximation to systems that mostly involve electrostatic changes and no introduction of dummy atoms. This indicates that the theory can be applied to some extent to multi-scale free energy simulations between an MM and a QM end state.

### 3.3. Multi-scale free energy simulations from MM to QM

Our final benchmark consists of multi-scale free energy simulations between MM and QM energy surfaces. In particular, we employ MM trajectories based on CGenFF and attempt to obtain free energy differences with the Zwanzig equation to QM energy surfaces based on BLYP, B3LYP, Hartree–Fock and M06-2X with 6-31G\* basis set. We employ the same twelve molecules as in the previous case, which leads to 48 free energy differences. Since performing quantum-mechanical simulations on the nanosecond time scale would be very expensive, we do not consider the use of BAR here. Instead, we attempt to check the basic transferability of our results to multi-scale free energy simulations. The results are shown in Figure 10.

Although the range of the variance data is significantly smaller compared to the more extensive test in Figure 8, the overall metrics are comparable. The  $R^2$  is significantly lower with 0.66 (compared to 0.91), but this reflects both the limited range and

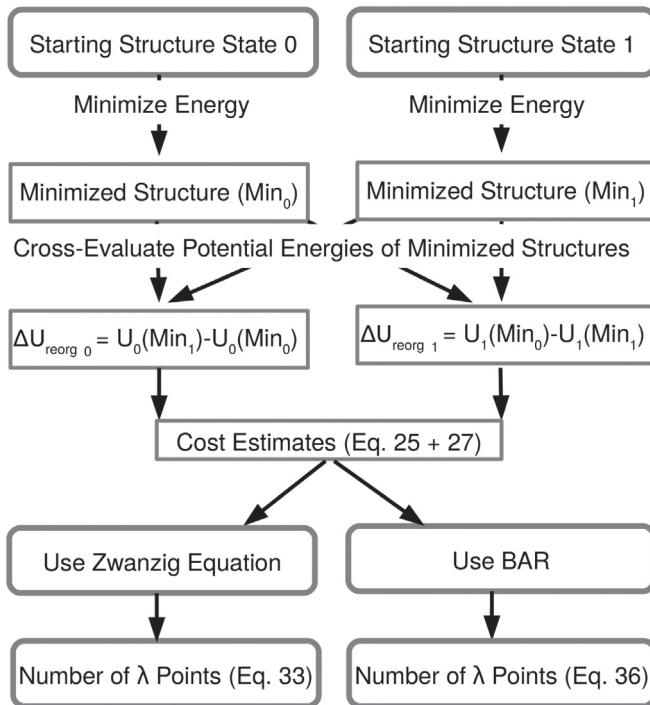


Figure 11. Overview of the proposed recursive workflow.

the fact that this benchmark is significantly more challenging due to the inherent differences of the underlying physics of both end states. The mean signed deviation is  $-0.51 \text{ kT}$  (compared to  $0.32 \text{ kT}$ ) and the root mean square deviation is  $0.83 \text{ kT}$  (compared to  $0.82 \text{ kT}$ ). Thus, the root mean square deviation is very similar to the previous results, but the systematic error is higher and tends to underestimate the variance. While the introduction of  $\lambda$  intermediate states is possible in multi-scale free energy simulations [31,85], it is less common than in molecular mechanics. Thus the linear response approximation is probably more valuable to gauge the required number of data points, which is also a way to increase precision. Alternatively, one can also use  $\Delta U_{\text{reorg}}$  to evaluate whether a different force field is more likely to yield better precision.

#### 4. Conclusions

In this paper, we have attempted to devise a pragmatic solution to one of the dilemmas involved in free energy simulations: How to devise a free energy simulation protocol when the required information for the protocol is the entropic contribution of the resulting free energy difference itself. For this purpose, we borrow some ideas from Marcus theory and linear response theory. An overview of the basic strategy is provided in Figure 11. Starting from some representative initial structures (e.g. a crystal structure), the reorganisation energies ( $\Delta U_{\text{reorg}}$ ) between the two states are determined. Based on  $\Delta U_{\text{reorg}}$ , the expected computational costs can be derived, which serve as a basis to decide which free energy method is most likely going to be more efficient, and the expected number of  $\lambda$  steps is determined. Ideally, the procedure is supposed to be used recursively on

each of the sub-steps to minimise potential errors arising from mismatching force constants.

The proposed approach was tested using free energy simulations in the gas phase of model systems and of a set of twelve molecules, reaching root mean square deviations of the decadic logarithm of the variances between 0.38 and 0.83. This indicates that the  $\lambda$  protocol can be chosen with an accuracy of about  $\pm 2$  steps. We also provide some metrics based on phase space overlap, vibrational entropy and the presence of dummy atoms to predict when the linear response approach is likely to break down. As a final verification, we employed multi-scale free energy simulations between molecular mechanics and quantum mechanics to demonstrate the applicability of our simplistic approach.

While it would be rather bold to suggest that the provided theory can be used in all fields of free energy simulations, it is still valuable for educational purposes. The underlying theory allows us to address practical questions such as the relative value of speed and accuracy in molecular simulation. This is highly relevant to appreciate the value of recent developments such as polarisable force fields [86–94], semi-empirical methods [95,96], or applications of QM/MM [31,32,32,80,97–139]. As illustrated in Section 1.7, the required number of data points in a free energy calculation increases exponentially with the error of the potential energy in terms of the reorganisation energy. Therefore, a slow method can actually out-compete a much faster method if it manages to provide modest gains in accuracy. However, the optimal balance of speed and accuracy depends on the details of the application, since other factors such as parallelisability or the autocorrelation time of sampling play a role. Those factors are quantified here.

To some degree, the development of advanced sampling techniques [58,60,140–152] may also tip the balance in favour of more expensive Hamiltonians, since they decrease the auto-correlation between consecutive dynamics steps. The analysis in Section 1.8 shows that the impact of parallelisation of the post-processing steps is rather limited. However, parallelisation of sampling algorithms, also in the form of GPU computing [13,90], favours the slower, but more accurate potential energy function.

To summarise, our results show from a purely statistical mechanics point of view the connections of the efficiency of free energy calculations with underlying concepts like accuracy, speed, parallelisability, sampling and the choice of free energy method. For simple systems, we demonstrate how to make the optimal choice in terms of the free energy protocol. From a theoretical point of view, the use of more accurate polarisable force fields and, ultimately, quantum-mechanical methods for sampling is inevitable for applications of free energy simulations. While fast, approximate methods provide the psychological advantage of more immediate gratification, they are often not competitive if the speed is bought with a notable loss of accuracy. However, since the optimal performance of free energy simulations depends on several factors, it is still up to the user to weight the impact of the different choices. Being able to come up with the best possible strategy for a challenging system is therefore still one of the hallmarks of a good computational chemist.

## Note

- While different numbers of data points could in principle be used for the initial state and the final state, it is pointed out in the original BAR paper [6] that the optimal ratio depends on the variance of the Fermi functions (c.f., Equations (13)–(16) and Figure 3 in Ref. [6]) – which is equal in our model. However, spending about equal computer time on both states could also be a viable alternative..

## Acknowledgements

The authors thank Wei Yang, Lee Woodcock, Kenneth Kostenbader and George Giambaşu for very helpful discussions. The authors are much indebted to Yihan Shao and Benjamin T. Miller for their assistance during the generation of the simulation data.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the National Institutes of Health [number GM107485 to D.M.Y.] and the Max Planck Society. It utilised 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://hpc.nih.gov>), and resources provided by the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey, the National Institutes of Health [grant number S10OD012346] and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation [grant number ACI-1548562] [grant number OCI-1053575].

## ORCID

Gerhard König  <http://orcid.org/0000-0003-4898-2958>  
Walter Thiel  <http://orcid.org/0000-0001-6780-0350>

## References

- Wu D, Kofke D. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. *J Chem Phys.* **2005**;123(5):054103.
- Wu D, Kofke D. Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. *J Chem Phys.* **2005**;123(5):084109.
- König G, Bruckner S, Boresch S. Unorthodox uses of Bennett's acceptance ratio method. *J Comput Chem.* **2009**;30(11):1712–1718.
- Bruckner S, Boresch S. Efficiency of alchemical free energy simulations I: practical comparison of the exponential formula, thermodynamic integration and Bennett's acceptance ratio method. *J Comput Chem.* **2011**;32:1303–1319.
- Boresch S, Woodcock HL. Convergence of single-step free energy perturbation. *Mol Phys.* **2017**;115(9–12, SI):1200–1213.
- Bennett CH. Efficient estimation of free energy differences from Monte Carlo data. *J Comp Phys.* **1976**;22:245–268.
- Marcus R. Electron-transfer reactions in chemistry – theory and experiment (Nobel Lecture). *Ang Chem Int Ed.* **1993**;32(8):1111–1121.
- Zwanzig RW. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J Chem Phys.* **1954**;22:1420.
- Kofke D. On the acceptance probability of replica-exchange Monte Carlo trials. *J Chem Phys.* **2002**;117(15):6911–6914.
- Peter C, Oostenbrink C, van Dorp A, et al. Estimating entropies from molecular dynamics simulations. *J Chem Phys.* **2004**;120(6):2652–2661.
- Kirkwood JG. Statistical mechanics of fluid mixtures. *J Chem Phys.* **1935**;3:300–313.
- Bruckner S, Boresch S. Efficiency of alchemical free energy simulations II: Improvements for thermodynamic integration. *J Comput Chem.* **2011**;32:1320–1333.
- Lee TS, Hu Y, Sherborne B, et al. Toward fast and accurate binding affinity prediction with pmemdGTI: an efficient implementation of GPU-accelerated thermodynamic integration. *J Chem Theory Comput.* **2017**;13(7):3077–3084.
- Giese TJ, York DM. A GPU-accelerated parameter interpolation thermodynamic integration free energy method. *J Chem Theory Comput.* **2018**;14(3):1564–1582.
- Chipot C, Pohorille A. Free energy calculations: theory and applications in chemistry and biology. Berlin: Springer; **2007**.
- Leach AR. Molecular modelling – principles and applications. 2nd ed. Essex, UK: Pearson Education; **2001**.
- Pohorille A, Jarzynski C, Chipot C. Good practices in free-energy calculations. *J Phys Chem B.* **2010**;114:10235–10253.
- Jarzynski C. Nonequilibrium equality for free energy differences. *Phys Rev Lett.* **1997**;78:2690–2693.
- Crooks GE. Path-ensemble averages in systems driven far from equilibrium. *Phys Rev E.* **2000**;61:2361–2366.
- Luzhkov VB. On relation between the free-energy perturbation and Bennett's acceptance ratio methods: tracing the influence of the energy gap. *J Chem Phys.* **2010**;132(19):194104.
- Lu N, Singh JK, Kofke DA. Appropriate methods to combine forward and reverse free-energy perturbation averages. *J Chem Phys.* **2003**;118:2977–2984.
- Simonson T. Free energy of particle insertion: an exact analysis of the origin singularity for simple liquids. *Mol Phys.* **1993**;80:441–447.
- Jorgensen W, Ravimohan C. Monte Carlo simulation of differences in free energies of hydration. *J Chem Phys.* **1985**;83(6):3050–3054.
- Lu N, Kofke DA, Woolf TB. Improving the efficiency and reliability of free energy perturbation calculations using overlap sampling methods. *J Comput Chem.* **2004**;25:28–40.
- Shirts MR, Bair E, Hooker G, et al. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys Rev Lett.* **2003**;91:140601.
- Ferrenberg AM, Swendsen RH. Optimized monte carlo data analysis. *Phys Rev Lett.* **1989**;63(12):1195–1198.
- Kumar S, Bouzida D, Swendsen RH, et al. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem.* **1992**;13:1011–1021.
- Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys.* **2008**;129(12):124105.
- König G, Boresch S. Non-Boltzmann sampling and Bennett's acceptance ratio method: how to profit from bending the rules. *J Comput Chem.* **2011**;32(6):1082–1090.
- Wereszczynski J, McCammon JA. Using selectively applied accelerated molecular dynamics to enhance free energy calculations. *J Chem Theory Comput.* **2010**;6(11, 3285):3285–3292.
- König G, Hudson PS, Boresch S, 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. *J Chem Theory Comput.* **2014**;10(4):1406–1419.
- König G, Pickard FC, Mei Y, et al. Predicting hydration free energies with a hybrid QM/MM approach: an evaluation of implicit and explicit solvation models in SAMPL4. *J Comput-Aid Mol Design.* **2014**;28(3, SI):245–257.
- Maragakis P, Spichty M, Karplus M. Optimal estimates of free energies from multistate nonequilibrium work data. *Phys Rev Lett.* **2006**;96(10):100602.
- Paliwal H, Shirts MR. A benchmark test set for alchemical free energy transformations and its use to quantify error in common free energy methods. *J Chem Theory Comput.* **2011**;7(12):4115–4134.
- de Ruiter A, Boresch S, Oostenbrink C. Comparison of thermodynamic integration and Bennett's acceptance ratio for calculating relative protein-ligand binding free energies. *J Comp Chem.* **2013**;34(12):1024–1034.
- Shirts MR, Pande VS. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett



- acceptance ratio, and thermodynamic integration. *J Chem Phys.* **2005**;122:144107-1–144107-16.
- [37] König G, Brooks BR. Predicting binding affinities of host-guest systems in the SAMPL3 blind challenge: the performance of relative free energy calculations. *J Comput-Aided Mol Des.* **2012**;26(5):543–550.
- [38] Aqvist J, Medina C, Samuelsson JE. New method for predicting binding-affinity in computer-aided drug design. *Prot Eng.* **1994**;7:385–391.
- [39] Hansson T, Marelus J, Aqvist J. Ligand binding affinity prediction by linear interaction energy methods. *J Comp Aid Mol Des.* **1998**;12:27–35.
- [40] Almlöf M, Carlsson J, Åqvist J. Improving the accuracy of the linear interaction energy method for solvation free energies. *J Chem Theory Comput.* **2007**;3(6):2162–2175.
- [41] Stjernshantz E, Marelus J, Medina C, et al. Are automated molecular dynamics simulations and binding free energy calculations realistic tools in lead optimization? An evaluation of the linear interaction energy (LIE) method. *J Chem Inf Model.* **2006**;46(5):1972–1983.
- [42] Kolb P, Huang D, Dey F, et al. Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J Med Chem.* **2008**;51(5):1179–1188.
- [43] Gutiérrez-de Terán H, Åqvist J. Linear interaction energy: method and applications in drug design. Vol. 819, Computational drug discovery and design. Methods in molecular biology. **2012**. p. 305–323.
- [44] Hummer G, Pratt LR, García AE. Free energy of ionic hydration. *J Phys Chem.* **1996**;100:1206–1215.
- [45] Åqvist J, Hansson T. On the validity of electrostatic linear response in polar solvents. *J Phys Chem.* **1996**;100(22):9512–9521.
- [46] Smith PE, van Gunsteren W. When are free energy components meaningful? *J Phys Chem.* **1994**;98:13735–13740.
- [47] Heid E, Moser W, Schröder C. On the validity of linear response approximations regarding the solvation dynamics of polyatomic solutes. *Phys Chem Chem Phys.* **2017**;19(17):10940–10950.
- [48] Shirts M, Pande V. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J Chem Phys.* **2005**;122(14):144107.
- [49] Zuckerman D, Woolf T. Theory of a systematic computational error in free energy differences. *Phys Rev Lett.* **2002**;89(18):180602.
- [50] Chiani M, Dardari D. Improved exponential bounds and approximation for the Q-function with application to average error probability computation. **2002**. p. 1399–1402. IEEE Global Telecommunications Conference (Globecom). IEEE Global Telecommunications Conference (GLOBECOM 02), Taipei, Taiwan, Nov 17–21, 2002.
- [51] Chernoff H. A note on an inequality involving the normal-distribution. *Ann Prob.* **1981**;9(3):533–535.
- [52] Amdahl GM. Computer Architecture and Amdahl's Law. *Computer.* **2013**;46(12):38–46.
- [53] Brooks B, Brooks C III, Mackerell A Jr, et al. CHARMM: the biomolecular simulation program. *J Comput Chem.* **2009**;30(10, Sp. Iss. SI):1545–1614.
- [54] Brooks BR, Brucolieri RE, Olafson BD, et al. CHARMM – a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem.* **1983**;4:187–217.
- [55] Vanommeslaeghe K, Hatcher E, Acharya C, et al. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comp Chem.* **2010**;31(4):671–690.
- [56] Boresch S, Karplus M. The Jacobian factor in free energy simulations. *J Chem Phys.* **1996**;105:5145–5154.
- [57] Boresch S, Karplus M. The role of bonded terms in free energy simulations: 1. theoretical analysis. *J Phys Chem A.* **1999**;103:103–118.
- [58] König G, Miller BT, Boresch S, et al. Enhanced sampling in free energy calculations: combining SGML with the Bennett's acceptance ratio and enveloping distribution sampling methods. *J Chem Theory Comput.* **2012**;8(10):3650–3662.
- [59] König G, Mei Y, Pickard FC, et al. Computation of hydration free energies using the multiple environment single system quantum mechanical/molecular mechanical method. *J Chem Theory Comput.* **2016**;12(1):332–344.
- [60] Sugita Y, Kitao A, Okamoto Y. Multidimensional replica-exchange method for free-energy calculations. *J Chem Phys.* **2000**;113:6042–6050.
- [61] Zacharias M, Straatsma TP, McCammon JA. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J Chem Phys.* **1994**;100:9025–9031.
- [62] Shao Y, Molnar LF, Jung Y, et al. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys Chem Chem Phys.* **2006**;8(27):3172–3191.
- [63] Woodcock HL III, Hodošček M, Gilbert ATB, et al. Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations. *J Comp Chem.* **2007**;28(9):1485–1502.
- [64] Hariharan PC, Pople JA. Accuracy of AH<sub>n</sub> equilibrium geometries by single determinant molecular orbital theory. *Mol Phys.* **1974**;27(1):209–214.
- [65] Becke A. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A.* **1988**;38(6):3098–3100.
- [66] Lee C, Yang W, Parr R. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B.* **1988**;37(2):785–789.
- [67] Becke AD. Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys.* **1993**;98(7):5648–5652.
- [68] Zhao Y, Truhlar DG. 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.* **2007**;120:215–241.
- [69] Zhao Y, Truhlar DG. Density functionals with broad applicability in chemistry. *Accounts Chem Res.* **2008**;41:157–167.
- [70] Gore J, Ritort F, Bustamante C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc Natl Acad Sci USA.* **2003**;100(22):12564–12569.
- [71] Lu N, Kofke D. Accuracy of free-energy perturbation calculations in molecular simulation. I. Modeling. *J Chem Phys.* **2001**;114(17):7303–7311.
- [72] Kästner J, Thiel W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "umbrella integration". *J Chem Phys.* **2005**;123(14):144104.
- [73] Lee TS, Radak BK, Huang M, et al. Roadmaps through Free Energy Landscapes Calculated Using the Multidimensional vFEP Approach. *J Chem Theory Comput.* **2014**;10(1):24–34.
- [74] König G, Boresch S. Hydration free energies of amino acids: why side chain analog data are not enough. *J Phys Chem B.* **2009**;113(26):8967–8974.
- [75] König G. Methodological studies concerning free energy simulations [dissertation]. Vienna, Austria: University of Vienna; **2010**.
- [76] König G, Bruckner S, Boresch S. Absolute hydration free energies of blocked amino acids: implications for protein solvation and stability. *Biophys J.* **2013**;104(2):453–462.
- [77] Brooks BR, Janežič D, Karplus M. Harmonic analysis of large systems. I. Methodology. *J Comput Chem.* **1995**;16:1522–1542.
- [78] Andricioaei I, Karplus M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J Chem Phys.* **2001**;115(14):6289–6292.
- [79] Karplus M, Kushick JN. Method for estimating the configurational entropy of macromolecules. *Macromolecules.* **1981**;14:325–332.
- [80] König G, Brooks BR. Correcting for the free energy costs of bond or angle constraints in molecular dynamics simulations. *Biochim Biophys Acta – Gen Subj.* **2015**;1850(5, SI):932–943.
- [81] Lu N, Wu D, Woolf TB, et al. Using overlap and funnel sampling to obtain accurate free energies from nonequilibrium work measurements. *Phys Rev E.* **2004**;69:05772.
- [82] Boresch S, Karplus M. The role of bonded terms in free energy simulations. 2. calculation of their influence on free energy differences of solvation. *J Phys Chem A.* **1999**;103:119–136.

- [83] Beutler TC, Mark AE, van Schaik RC, et al. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem Phys Lett.* **1994**;222:529–539.
- [84] Flory P. Die konformation linearer makromoleküle. *Angew Chem.* **1975**;87(22):787–822.
- [85] Dybeck EC, König G, Brooks BR, et al. A comparison of methods to reweight from classical molecular simulations to QM/MM potentials. *J Chem Theory Comput.* **2016**;12(4):1466–1480.
- [86] Anisimov V, Lamoureux G, Vorobyov I, et al. Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. *J Chem Theory Comput.* **2005**;1(1):153–168.
- [87] Lopes PEM, Roux B, MacKerell AD Jr. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theor Chem Acc.* **2009**;124(1–2):11–28.
- [88] Lopes PEM, Huang J, Shim J, et al. Polarizable force field for peptides and proteins based on the classical drude oscillator. *J Chem Theory Comput.* **2013**;9(12):5430–5449.
- [89] Ponder JW, Wu C, Ren P, et al. Current status of the AMOEBA polarizable force field. *J Phys Chem B.* **2010**;114:2549–2564.
- [90] Lagardere L, Jolly LH, Lipparini F, et al. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem Sci.* **2018**;9(4):956–972.
- [91] Harger M, Li D, Wang Z, et al. Tinker-OpenMM: absolute and relative alchemical free energies using AMOEBA on GPUs. *J Comput Chem.* **2017**;38(23):2047–2055.
- [92] Huang J, Simmonett AC, Pickard FC, et al. Mapping the Drude polarizable force field onto a multipole and induced dipole model. *J Chem Phys.* **2017**;147(16):161702.
- [93] Simmonett AC, Pickard FC, Schaefer HF III, et al. An efficient algorithm for multipole energies and derivatives based on spherical harmonics and extensions to particle mesh Ewald. *J Chem Phys.* **2014**;140(18):184101.
- [94] Simmonett AC, Pickard FC, Shao Y, et al. Efficient treatment of induced dipoles. *J Chem Phys.* **2015**;143(7):074115.
- [95] Dral PO, Wu X, Spörkel L, et al. Semiempirical quantum-chemical orthogonalization-corrected methods: benchmarks for ground-state properties. *J Chem Theory Comput.* **2016**;12(3):1097–1120.
- [96] Dral PO, Wu X, Spörkel L, et al. Semiempirical quantum-chemical orthogonalization-corrected methods: theory, implementation, and parameters. *J Chem Theory Comput.* **2016**;12(3):1082–1096.
- [97] Ryde U, Söderhjelm P. Ligand-binding affinity estimates supported by quantum-mechanical methods. *Chem Rev.* **2016**;116(9, SI):5520–5566.
- [98] Stanton RV, Hartsough DS, Merz KM. Calculation of solvation free energies using a density functional/molecular dynamics coupled potential. *J Phys Chem.* **1993**;97:11868–11870.
- [99] Reddy MR, Singh UC, Erion MD. Development of a quantum mechanics-based free-energy perturbation method: use in the calculation of relative solvation free energies. *J Am Chem Soc.* **2004**;126:6224–6225.
- [100] Reddy MR, Singh UC, Erion MD. Ab initio quantum mechanics-based free energy perturbation method for calculating relative solvation free energies. *J Comput Chem.* **2007**;28:491–494.
- [101] Riccardi D, Schaefer P, Yang Y, et al. Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J Phys Chem B.* **2006**;110:6458–6469.
- [102] Yang W, Cui Q, Min D, et al. Chapter 4 – QM/MM alchemical free energy simulations: challenges and recent developments. *Annu Rep Comput Chem.* **2010**;6:51–62.
- [103] Yang W, Bitetti-Putzer R, Karplus M. Chaperoned alchemical free energy simulations: a general method for QM, MM, and QM/MM potentials. *J Chem Phys.* **2004**;120:9450–9453.
- [104] Zhang Y, Liu H, Yang W. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J Chem Phys.* **2000**;112(8):3483–3492.
- [105] Min D, Chen M, Zheng L, et al. Enhancing QM/MM molecular dynamics sampling in explicit environments via an orthogonal space-random-walk-based strategy. *J Phys Chem B.* **2011**;115:3924–3935.
- [106] Min D, Zheng L, Harris W, et al. Practically efficient QM/MM alchemical free energy simulations: the orthogonal space random walk strategy. *J Chem Theory Comput.* **2010**;6:2253–2266.
- [107] Kästner J, Senn H, Thiel S, et al. QM/MM free-energy perturbation compared to thermodynamic integration and umbrella sampling: application to an enzymatic reaction. *J Chem Theory Comput.* **2006**;2(2):452–461.
- [108] Polyak I, Benighaus T, Boulanger E, et al. Quantum mechanics/molecular mechanics dual Hamiltonian free energy perturbation. *J Chem Phys.* **2013**;139:064105–064116.
- [109] Senn HM, Thiel W. QM/MM methods for biomolecular systems. *Ang Chem Int Ed.* **2009**;48:1198–229.
- [110] Nam K, Gao J, York DM. An efficient linear-scaling ewald method for long-range electrostatic interactions in combined QM/MM calculations. *J Chem Theory Comput.* **2005**;1:2–13.
- [111] Štrajbl M, Hong G, Warshel A. Ab Initio QM/MM Simulation with proper sampling: "First Principle" calculations of the free energy of the autodissociation of water in aqueous solution. *J Phys Chem B.* **2002**;106(51):13333–13343.
- [112] Plotnikov NV, Kamerlin SCL, Warshel A. Paradynamics: an effective and reliable model for ab initio QM/MM free-energy calculations and related tasks. *J Phys Chem B.* **2011**;115(24):7950–7962.
- [113] Rod TH, Ryde U. Quantum mechanical free energy barrier for an enzymatic reaction. *Phys Rev Lett.* **2005**;94(13):138302.
- [114] Rod TH, Ryde U. Accurate QM/MM free energy calculations of enzyme reactions: methylation by catechol O-methyltransferase. *J Chem Theory Comput.* **2005**;1(6):1240–1251.
- [115] Beierlein FR, Michel J, Essex JW. A simple QM/MM approach for capturing polarization effects in protein-ligand binding free energy calculations. *J Phys Chem B.* **2011**;115(17):4911–4926.
- [116] Fox SJ, Pittcock C, Tautermann CS, et al. Free energies of binding from large-scale first-principles quantum mechanical calculations: application to ligand hydration energies. *J Phys Chem B.* **2013**;117(32):9478–9485.
- [117] Mikulskis P, Ciološec D, Andrejić M, et al. Free-energy perturbation and quantum mechanical study of SAMPL4 octa-acid host-guest binding energies. *J Comp-Aid Mol Des.* **2014**;28(4, SI):375–400.
- [118] Genheden S, Ryde U, Söderhjelm P. Binding affinities by alchemical perturbation using QM/MM with a large QM system and polarizable MM model. *J Comput Chem.* **2015**;36(28):2114–2124.
- [119] Sampson C, Fox T, Tautermann CS, et al. A "Stepping Stone" approach for obtaining quantum free energies of hydration. *J Phys Chem B.* **2015**;119(23):7030–7040.
- [120] Cave-Aylard C, Skylaris CK, Essex JW. Direct validation of the single step classical to quantum free energy perturbation. *J Phys Chem B.* **2015**;119(3, SI):1017–1025.
- [121] Ollsson MA, Söderhjelm P, Ryde U. Converging ligand-binding free energies obtained with free-energy perturbations at the quantum mechanical level. *J Comput Chem.* **2016**;37(17):1589–1600.
- [122] Gao J, Xia X. A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science.* **1992**;258(5082):631–635.
- [123] Gao J, Luque FJ, Orozco M. Induced dipole moment and atomic charges based on average electrostatic potentials in aqueous solution. *J Chem Phys.* **1993**;98(4):2975.
- [124] Luzhkov V, Warshel A. Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies. *J Comput Chem.* **1992**;13(2):199–213.
- [125] Wesolowski T, Warshel A. Ab initio free energy perturbation calculations of solvation free energy using the frozen density functional approach. *J Phys Chem.* **1994**;98(20):5183–5187.
- [126] Gao J, Freindorf M. Hybrid ab initio QM/MM Simulation of N-methylacetamide in aqueous solution. *J Phys Chem A.* **1997**;101(17):3182–3188.
- [127] Kollman P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem Rev.* **1993**;93(7):2395–2417.

- [128] Li H, Yang W. 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.* **2007**;126(11):114104.
- [129] Woods CJ, Manby FR, Mulholland AJ. An efficient method for the calculation of quantum mechanics/molecular mechanics free energies. *J Chem Phys.* **2008**;128(1):014109.
- [130] Heimdal J, Ryde U. Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations. *Phys Chem Chem Phys.* **2012**;14:12592–12604.
- [131] Hu H, Lu Z, Yang W. QM/MM minimum free energy path: methodology and application to triosephosphate isomerase. *J Chem Theory Comput.* **2007**;3(2):390–406.
- [132] Zeng X, Hu H, Hu X, et al. Ab initio quantum mechanical/molecular mechanical simulation of electron transfer process: fractional electron approach. *J Chem Phys.* **2008**;128(12):124510.
- [133] Hu H, Lu Z, Parks JM, et al. Quantum mechanics/molecular mechanics minimum free-energy path for accurate reaction energetics in solution and enzymes: sequential sampling and optimization on the potential of mean force surface. *J Chem Phys.* **2008**;128(3):034105.
- [134] Hu H, Yang W. Elucidating solvent contributions to solution reactions with ab initio QM/MM methods. *J Phys Chem B.* **2010**;114(8):2755–2759.
- [135] Hudson PS, Woodcock HL, Boresch S. Use of nonequilibrium work methods to compute free energy differences between molecular mechanical and quantum mechanical representations of molecular systems. *J Phys Chem Lett.* **2015**;6(23):4850–4856.
- [136] Pickard FC, König G, Tofoleanu F, et al. Blind prediction of distribution in the SAMPL5 challenge with QM based protomer and pK<sub>a</sub> corrections. *J Comput-Aided Mol Des.* **2016**;30(11, 1, SI):1087–1100.
- [137] König G, Pickard FC, Huang J, et al. Calculating distribution coefficients based on multi-scale free energy simulations: an evaluation of MM and QM/MM explicit solvent simulations of water-cyclohexane transfer in the SAMPL5 challenge. *J Comput-Aided Mol Des.* **2016**;30(11, 1, SI):989–1006.
- [138] Ryde U. How many conformations need to be sampled to obtain converged QM/MM energies? The curse of exponential averaging. *J Chem Theory Comput.* **2017**;13(11):5745–5752.
- [139] Olsson MA, Ryde U. Comparison of QM/MM methods to obtain ligand-binding free energies. *J Chem Theory Comput.* **2017**;13(5):2245–2253.
- [140] Yang W. Advanced sampling for molecular simulation is coming of age. *J Comput Chem.* **2016**;37(6, SI):549–549.
- [141] Min D, Zheng L, Harris W, et al. Practically efficient QM/MM alchemical free energy simulations: the orthogonal space random walk strategy. *J Chem Theory Comput.* **2010**;6(8):2253–2266.
- [142] Li H, Yang W. 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.* **2007**;126(11):114104.
- [143] Abella JR, Cheng SY, Wang Q, et al. Hydration free energy from orthogonal space random walk and polarizable force field. *J Chem Theory Comput.* **2014**;10(7):2792–2801.
- [144] Grubmüller H. Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys Rev E.* **1995**;52:2893–2906.
- [145] Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys.* **2004**;120:11919–11929.
- [146] Berne BJ, Straub JE. Novel methods of sampling phase space in the simulation of biological systems. *Curr Opin Struct Biol.* **1997**;7:181–189.
- [147] Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett.* **1999**;314:141–151.
- [148] Hansmann UHE. Parallel tempering algorithm for conformational studies of biological molecules. *Chem Phys Lett.* **1997**;281:140–150.
- [149] Woods CJ, Essex JW, King MA. The development of replica-exchange-based free-energy methods. *J Phys Chem B.* **2003**;107:13703–13710.
- [150] Woods CJ, Essex JW, King MA. Enhanced configurational sampling in binding free-energy calculations. *J Phys Chem B.* **2003**;107:13711–13718.
- [151] Wu X, Brooks BR. Self-guided Langevin dynamics simulation method. *Chem Phys Lett.* **2003**;381:512–518.
- [152] Wu X, Brooks BR. Force-momentum-based self-guided Langevin dynamics: A rapid sampling method that approaches the canonical ensemble. *J Chem Phys.* **2011**;135(20):204101.
- [153] Maragakis P, Ritort F, Bustamante C, et al. Bayesian estimates of free energies from nonequilibrium work data in the presence of instrument noise. *J Chem Phys.* **2008**;129(2):024102.

## Appendix 1. Alternative expression for the BAR variance

In the main text, we have approximated the overlap expression  $\Omega$  using a step function instead of the Fermi function. Here, we approximate the product of the Fermi function with the probability distribution of the harmonic oscillator, which leads to a logistic-normal distribution. Since there is no simple analytical expression of the integral of the logistic-normal distribution, we replace it by a normal distribution with matched moments. This corresponds to replacing the Fermi function by a Gaussian distribution, which leads to stronger decay of the tail of the resulting distribution.

In particular, this means replacing the Fermi function by the Gaussian approximation, i.e.

$$\frac{1}{1 + e^{-(\Delta U - \Delta H)}} \approx e^{-\frac{(Kd - \ln 2)^2}{\ln 2}}, \quad (\text{A1})$$

which leads on an overlap integral of

$$\Omega <= \frac{\sqrt{\ln 2}}{\sqrt{Kd^2 + \ln 2}} e^{-Kd^2} = \frac{\sqrt{\ln 2}}{\sqrt{\frac{\Delta U_{reorg}}{4} + \ln 2}} e^{-\frac{\Delta U_{reorg}}{4}} \quad (\text{A2})$$

Notably, another possible approximation to the logistic-normal integral has been provided by Maragakis, Ritort, Karplus, Bustamante and Crooks.

[153] We readily admit that their approximation is probably superior, but our guiding criterion was simplicity, since we were searching for a simple expression that can be readily used in the ensuing analysis.