# Feature Article
# Basic Ingredients of Free Energy Calculations: A Review

**CLARA D. CHRIST,[1] ALAN E. MARK,[2] WILFRED F. vAN GUNSTEREN[1]**
*[1]Laboratory of Physical Chemistry, Swiss Federal Institute of Technology,
ETH, Zürich 8093, Switzerland*
*[2]School of Molecular and Microbial Sciences, Institute for Molecular Biosciences,
University of Queensland, St Lucia, Queensland 4072, Australia*

**Abstract:** Methods to compute free energy differences between different states of a molecular system are reviewed with the aim of identifying their basic ingredients and their utility when applied in practice to biomolecular systems. A free energy calculation is comprised of three basic components: (i) a suitable model or Hamiltonian, (ii) a sampling protocol with which one can generate a representative ensemble of molecular configurations, and (iii) an estimator of the free energy difference itself. Alternative sampling protocols can be distinguished according to whether one or more states are to be sampled. In cases where only a single state is considered, six alternative techniques could be distinguished: (i) changing the dynamics, (ii) deforming the energy surface, (iii) extending the dimensionality, (iv) perturbing the forces, (v) reducing the number of degrees of freedom, and (vi) multi-copy approaches. In cases where multiple states are to be sampled, the three primary techniques are staging, importance sampling, and adiabatic decoupling. Estimators of the free energy can be classified as global methods that either count the number of times a given state is sampled or use energy differences. Or, they can be classified as local methods that either make use of the force or are based on transition probabilities. Finally, this overview of the available techniques and how they can be best used in a practical context is aimed at helping the reader choose the most appropriate combination of approaches for the biomolecular system, Hamiltonian and free energy difference of interest.

© 2009 Wiley Periodicals, Inc.     J Comput Chem 00: 000–000, 2010

**Key words:** free energy; affinity; molecular simulation; sampling

## Introduction

The estimation of free energies using molecular simulation techniques has been an active field of research for many decades. The free energy is a highly desirable quantity to compute. It is in essence the factor that determines how a process will proceed and the probability that a system will adopt a given state. The ability to calculate free energies from molecular simulations not only allows one to understand the underlying processes on an atomic level but also to probe states of a system not accessible experimentally. In particular, if precise and accurate estimates of the free energy of a system could be obtained directly from numerical simulations, the need to measure thermodynamic properties of a system, such as ligand binding constants, by experiment would be greatly reduced. This is why, for example, free energy calculations have attracted much interest in areas such as rational drug design and material science. However, to obtain a reliable estimate of the free energy of a system, a number of challenges must be met.

The free energy $F$ of a system in the canonical ensemble, i.e., at constant number of particles, volume, and temperature, is given by

$$F = -\frac{1}{\beta} \ln Q, \tag{1}$$

where $\beta$ is the inverse temperature divided by Boltzmann's constant $k_B$ and $Q$ is the partition function of the system. For simplicity, we will restrict ourselves to a classical description of the system in Cartesian coordinates. The system is also assumed to be at thermodynamic equilibrium. In this case, the partition function can be expressed as

$$Q = \frac{1}{h^{3N}N!} \iint e^{-\beta H(p,r)} dp dr, \tag{2}$$

where $N$ is the number of particles in the system and $h$ is Planck's constant. The factor $N!$ appears only for indistinguishable particles. The integral is performed over all $3N$ positions $r$ and conjugate momenta $p$, respectively. The Hamiltonian $H(p,r)$ gives the total energy of the system in a given configuration, i.e., a given set of momenta and coordinates.

The absolute free energy [eq. (1)] can only be calculated directly in a limited number of cases where an analytical expression for the partition function can be obtained. This is primarily for small simple systems governed by a very simple Hamiltonian. For larger systems with strong interactions between the particles, an analytical formulation of the partition function is generally not possible. In such cases, one is restricted to the calculation of the difference in free energy between the system of interest and a given reference state. However, if the free energy of the reference state is known, such as for an ideal gas, for a gaseous system, or an ideal crystal, for a solid phase system, the absolute free energy of the system can still be obtained. In these cases, an analytical expression for the partition function can be obtained either due to lack of interactions between the particles (ideal gas) or simplifications due to symmetry (ideal crystal). Defining a suitable reference state for a liquid phase system is more difficult although there have been ongoing attempts to do so.[1,2]

In most cases, where we wish to compare to experiment, our aim is to determine relative free energies, for example, the difference in binding of two compounds to the same receptor. Note, the relative free energy between the bound and the unbound state of a single compound is referred to by some as the "absolute" free energy of binding but this should not be confused with the absolute free energy of the system discussed earlier. Generally speaking we seek to estimate the difference in free energy between two states A and B (or possibly a series of pairs of states A and B).

$$\Delta F_{BA} = F_B - F_A = -\beta^{-1} \ln \frac{Q_B}{Q_A}, \qquad (3)$$

where A and B (i.e., $H_A$ and $H_B$) might differ in the way the particles interact with each other, i.e., when calculating a difference in free energy between two different compounds in an "alchemical" perturbation. However, A and B can also correspond to different conformations of the same molecular system where the accessible conformational space is restricted in A and B to the desired regions for, e.g., a set of restraints.

The main challenges that have to be met when attempting to determine differences in free energy using molecular simulation techniques relate to the choice of Hamiltonian and the sampling scheme on which the estimate of the relative free energy will be based. That is, a free energy calculation consists of three basic components.

- A model Hamiltonian (see "Molecular Model" Section).
- A sampling protocol, which will be used to generate a representative ensemble of configurations (see "Sampling" Section).
- A method to estimate the free energy difference (see "Estimation of the Free Energy" Section).

The various choices available for the basic components will be discussed in detail below. The aim is to enable the reader to classify the vast array of methods that have been described in the literature by identifying which choices have been made for these three basic components. Although many of these different techniques should in theory yield equivalent results, in practice given limitations in respect to the Hamiltonian and finite sampling, the results obtained using certain techniques are much more precise and accurate than others. Here, we focus primarily on those techniques that are most applicable in practical free energy calculations. We also focus on perturbations where only part of the system differs between state A and B, such as in "alchemical" free energy perturbations, where one molecule is perturbed into another while the majority of the system (e.g., the solvent) remains unperturbed. Note, about a decade ago, an expression was proposed for the relative free energy $\Delta F_{BA}$ in terms of the work done to force the system from state A to state B under non-equilibrium conditions.[3] This led to much research into the relation between equilibrium and non-equilibrium methods to compute $\Delta F_{BA}$. However, since non-equilibrium methods are yet to be shown to have practical advantages over equilibrium methods in terms of accuracy or efficiency we will focus our discussion on equilibrium approaches in "Estimation of the Free Energy" Section.

## Molecular Model

To estimate free energy differences reliably a molecular model is required that describes the thermodynamics of the system correctly. That is, the Hamiltonian used to calculate the energy and forces must be chosen such that all configurations have the correct relative probability. In practice, the choice of Hamiltonian is often a compromise between accuracy and efficiency. The evaluation of the energy and forces needs to be computationally inexpensive enough to permit sufficient sampling, yet of sufficient accuracy to estimate the free energy reliably.

The cost of the energy and force evaluation is primarily determined by two factors: the degrees of freedom that are considered and the functional form of the Hamiltonian. One can move from a quantum-mechanical description of the system where the electronic degrees of freedom are modeled explicitly, to a classical description where one atom is treated as one particle, to a coarse-grained description where groups of atoms are merged into one particle. A reduction in the number of degrees of freedom can, however, also be obtained by reducing the system size or treating parts of the system (e.g., the solvent) as a continuum. The motion along those degrees of freedom modeled explicitly will, ideally, be governed by a potential of mean force that represents the motion along the ("implicit") degrees of freedom that have been omitted. The choice of which degrees of freedom are modeled explicitly depends on the system of interest and the property one wishes to estimate. If a certain (class of) degree(s) of freedom is believed to have no effect on the property of interest, it can be omitted and the computing time gained invested in sampling the relevant degrees of freedom more extensively. How one can enhance sampling by reducing the number of degrees of freedom is further discussed in "Sampling Configurations of One State." The computational effort will also depend strongly on the functional form of the Hamiltonian. A quantum-mechanical description of the system will in general be computationally more demanding than a classical description. Furthermore, within the realms of quantum-mechanical and

classical methodology there exist a variety of approaches for which the computational demands vary markedly depending on the choice of approximation to the time-independent Schrödinger equation or the complexity of the functional form of the classical Hamiltonian, the choice again being a compromise between accuracy and speed.

Here, we will focus on classical Hamiltonians that incorporate fixed point- charges, although this is not an essential limitation of a classical model. This type of Hamiltonian, generally referred to as a "force field," is commonly used when modeling biomolecules such as proteins, nucleic acids, lipids, or carbohydrates. Although there are differences in the functional form used in the various force fields available they contain many common features. As an example, we will consider the GROMOS force field.[4,5] The classical Hamiltonian is split into a kinetic $K(\boldsymbol{p})$ and a potential energy term $V(\boldsymbol{r})$,

$$H(\boldsymbol{p}, \boldsymbol{r}) = K(\boldsymbol{p}) + V(\boldsymbol{r}) \,. \tag{4}$$

The kinetic term,

$$K(\boldsymbol{p}) = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m_i} \,, \tag{5}$$

is independent of the particle positions if no configurational constraints are applied. Here, $m_i$ is the mass of particle $i$. The potential energy term returns the interaction energy for a given configuration determined by a set of particle coordinates $\boldsymbol{r}$. It is a sum of various terms describing the bonded (bon) and nonbonded (nonb) interactions between particles as well as special, "unphysical" interactions such as restraints,

$$V(\boldsymbol{r}) = V^{\text{phys}}(\boldsymbol{r}) + V^{\text{special}}(\boldsymbol{r}) = V^{\text{bon}}(\boldsymbol{r}) + V^{\text{nonb}}(\boldsymbol{r}) + V^{\text{special}}(\boldsymbol{r}). \tag{6}$$

The bonded interactions consist of terms describing the bond-stretching, the bond-angle bending, and the dihedral-angle bending. The set of bonded interactions also contains terms that maintain a particular chirality or ensure planarity of groups. The nonbonded interactions are modeled by a Lennard-Jones 6-12 term describing dispersion and repulsion ($V^{\text{LJ}}$) and an electrostatic term ($V^{\text{CRF}}$)

$$V^{\text{nonb}}(\boldsymbol{r}) = V^{\text{LJ}}(\boldsymbol{r}) + V^{\text{CRF}}(\boldsymbol{r}) \tag{7}$$

$$= \sum_{\substack{\text{nonbonded} \\ \text{pairs}(i,j)}} \left\{ \left[ \frac{C_{12}(i,j)}{r_{ij}^6} - C_6(i,j) \right] \frac{1}{r_{ij}^6} \right.$$

$$\left. + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1} \left[ \frac{1}{r_{ij}} - \frac{\frac{1}{2}C_{\text{rf}} r_{ij}^2}{R_{\text{rf}}^3} - \frac{1 - \frac{1}{2}C_{\text{rf}}}{R_{\text{rf}}} \right] \right\} \,. \tag{8}$$

Here, $r_{ij}$ is the distance between atoms $i$ and $j$, $C_{12}(i,j)$ and $C_6(i,j)$ are the repulsive and the attractive Lennard-Jones parameters, respectively, $q_i$ and $q_j$ are the partial charges of atoms $i$ and $j$, respectively, $\epsilon_0$ is the permittivity of the vacuum, and $\epsilon_1$ is the relative dielectric permittivity of the system (generally, $\epsilon_1 = 1$). $C_{\text{rf}}$ and $R_{\text{rf}}$ are parameters of the reaction field.[6]

Various force fields differ in their functional form and the way their parameters were derived.[7,8] Parameters for the bond lengths and angles are often derived from quantum chemical calculations or crystal structures. Torsional parameters can be adjusted to fit torsional profiles obtained from quantum-chemical calculations or from experiment. This is done in conjunction with fitting of the non-bonded interaction parameters as the latter have a strong influence on the torsional barriers. In the derivation of nonbonded parameters, many differences exist between the force fields. Whereas force fields such as AMBER[9,10] and CHARMM[11,12] fit the charges to reproduce the electrostatic potential obtained from quantum-chemical calculations, OPLS[13–17] and GROMOS[4,5,18–20] fit nonbonded parameters (charges and Lennard-Jones parameters) such that they reproduce the thermodynamic properties, i.e., density and heat of vaporization, of simple liquids. For example, in the latest version of the GROMOS force-field,[5] the parameters have been optimized to reproduce the free enthalpy of hydration and apolar solvation. The accuracy of biomolecular force fields is typically assessed by comparison of the results of simulations to experimental data. Here, our interest concerns comparisons which involve free energies. Shirts et al.[21] have pointed out that the solubility of side-chain analogs is underestimated by many current force fields (AMBER, CHARMM, and OPLS-AA). As most biomolecular processes of interest take place in aqueous media, it is to be expected that a failure to reproduce the solubility of analogs of amino acid side chains is likely to result in a failure to reproduce the secondary structure ensembles correctly. The latest GROMOS force field[5] has, therefore, been parameterized not only to reproduce the properties of pure liquids but also the solvation free energies of the amino acid side-chain analogs. Interestingly, it was found that using a fixed charge model it was impossible to simultaneously reproduce the solvation free energies in cyclohexane and water, i.e., with a single set of partial charges. Recently, Mobley et al.[22] have reported hydration free energies for 504 small molecules parameterized using the AMBER Antechamber program[23] to assign GAFF[24] parameters. They were able to identify systematic errors for particular classes of compounds. With the availability and reliability of free energy calculations increasing, there are growing opportunities to use free energy calculations in force-field development.

## Sampling

As pointed out in Introduction, the present work is primarily concerned with the estimation of the difference in free energy $\Delta F_{\text{BA}}$ between two states A and B [eq. (3)] of a system. The challenge is to estimate a ratio of partition functions and in this section will be discussed how simulations can be used to estimate such quantities.

One instructive estimator (see Estimation of the Free Energy section for other choices) makes use of energy difference distributions

$$\rho_{\text{A}} \left( \Delta H; \Delta H_{\text{BA}} \right) = \langle \delta \left[ \Delta H - (H_{\text{B}} - H_{\text{A}}) \right] \rangle_{\text{A}} \tag{9}$$

$$\rho_{\text{B}} \left( \Delta H; \Delta H_{\text{BA}} \right) = \langle \delta \left[ \Delta H - (H_{\text{B}} - H_{\text{A}}) \right] \rangle_{\text{B}} \,. \tag{10}$$

Here, $\langle \cdot \rangle_X$ indicates an average over an ensemble obtained from a simulation at state $X$, $\delta$ is the delta function, and $\Delta H_{\text{BA}}(\boldsymbol{p}, \boldsymbol{r}) = H_{\text{B}}(\boldsymbol{p}, \boldsymbol{r}) - H_{\text{A}}(\boldsymbol{p}, \boldsymbol{r})$ returns the energy difference between states A

and B for a given configuration. The free energy difference can be expressed in terms of these energy difference distributions,[25–27]

$$\rho_B \left( \Delta H; \Delta H_{BA} \right) \exp \left[ -\beta \Delta F_{BA} \right] = \rho_A \left( \Delta H; \Delta H_{BA} \right) \exp \left[ -\beta \Delta H \right] . \tag{11}$$

Equation (11) states that the difference in free energy $\Delta F_{BA}$ is equal to the difference in energy $\Delta H$ at the point where the two energy difference distributions $\rho_A$ and $\rho_B$ intersect. This implies that to obtain a reliable estimate of $\Delta F_{BA}$ all "important" configurations at states A and B must be sampled in order that converged energy difference distributions can be constructed. Section "Sampling Configurations of One State" defines what are "important" configurations and why the sampling of all these configurations during a simulation may be challenging. A sufficient condition to estimate a difference in free energy via eq. (11) is that the intersection region of the energy difference distributions be sampled adequately. In many practical applications, however, the distributions obtained from simulations at the end states A and B do not show any overlap, precluding the determination of the point of intersection. It is therefore necessary to find a Hamiltonian that "connects" or "bridges" the two (or more) states of interest. Possible ways of connecting the states and sampling protocols that ensure sampling of the important configurations of the "combined" Hamiltonian are discussed in "Sampling Relative Probabilities of States" Section.

It should be stressed that the sampling, i.e., the generation of configurations from which the desired quantities can be computed, is the most time consuming part of a free energy calculation. In contrast, the estimation of the free energy from the sampled configurations, which is discussed in "Estimation of the Free Energy" Section, is comparatively fast.

### *Sampling Configurations of One State*

#### *Probability Densities and Ensemble Averages*

In classical statistical mechanics, an ensemble is a collection of an infinite number of systems, each in a particular configuration of $\boldsymbol{p}$ and $\boldsymbol{r}$, which obeys a given probability density. In the canonical ensemble, this probability density $p(\boldsymbol{p}, \boldsymbol{r})$ is given by

$$p(\boldsymbol{p}, \boldsymbol{r}) = \frac{e^{-\beta H(\boldsymbol{p}, \boldsymbol{r})}}{\iint e^{-\beta H(\boldsymbol{p}, \boldsymbol{r})} d\boldsymbol{p} d\boldsymbol{r}} . \tag{12}$$

If the Hamiltonian [eq. (4)] is separable in $\boldsymbol{p}$ and $\boldsymbol{r}$, the kinetic contribution can be integrated out and we obtain the configurational probability density

$$\rho(\boldsymbol{r}) = \frac{e^{-\beta V(\boldsymbol{r})}}{\int e^{-\beta V(\boldsymbol{r})} d\boldsymbol{r}} . \tag{13}$$

In the following, a collection of configurations which obeys this probability density will be called Boltzmann distributed.

We can define an ensemble average of an observable $A(\boldsymbol{r})$ as

$$\langle A \rangle = \int A(\boldsymbol{r}) \rho(\boldsymbol{r}) d\boldsymbol{r} . \tag{14}$$

To estimate such an ensemble average from a simulation of a given system, we need a method which can be used to generate configurations that have the desired probability density $\rho(\boldsymbol{r})$. Assuming that the integral in eq. (14) is dominated by $\rho(\boldsymbol{r})$, sampling configurations from $\rho(\boldsymbol{r})$ to obtain $\langle A \rangle$ is efficient as the configurations which have a high probability in $\rho(\boldsymbol{r})$ contribute most to the ensemble average $\langle A \rangle$. However, there are cases where the integral in eq. (14) is dominated not by $\rho(\boldsymbol{r})$ but by $A(\boldsymbol{r})$, e.g., because $A$ contains an exponential function as do estimators for the free energy (see "Estimation of the Free Energy" Section). In these cases, sampling from a Boltzmann distribution is not the optimal choice and other sampling strategies should be pursued as discussed in "Sampling Relative Probabilities of States" Section.

#### *Methods to Sample from a Boltzmann Distribution*

The two most popular methods used to generate Boltzmann distributed ensembles are the Metropolis Monte Carlo algorithm[28, 29] and molecular dynamics simulation techniques.[30, 31]

Using the Metropolis Monte Carlo algorithm, a random walk through configuration space is constructed such that the probability of visiting a certain configuration $\boldsymbol{r}$ is proportional to $\exp[-\beta V(\boldsymbol{r})]$. One possible way of constructing such a random walk is to generate a new configuration by applying a random displacement to a particle chosen randomly. This new configuration $\boldsymbol{r}'$ is then accepted with the probability $p = \min \left\{ 1, \exp[-\beta(V(\boldsymbol{r}') - V(\boldsymbol{r}))] \right\}$.

In a molecular dynamics simulation, the equations of motion for the particles of the system are integrated forward in time. The equations of motion according to Newton can be expressed as

$$m_i \ddot{\boldsymbol{r}}_i = \boldsymbol{f}_i(\boldsymbol{r}) = -\frac{\partial V(\boldsymbol{r})}{\partial \boldsymbol{r}_i} , \tag{15}$$

where $m_i$ is the mass of particle $i$, $\boldsymbol{r}_i$ is the Cartesian position vector of particle $i$, the double dot indicates the second derivative with respect to time, and $\boldsymbol{f}_i$ is the force on atom $i$. The numerical integration can be performed using different discretization schemes.[32, 33] A molecular dynamics simulation generates (assuming perfect integration) configurations, which are distributed according to the microcanonical (constant energy) ensemble. To obtain an ensemble of configurations which is distributed according to the canonical ensemble, the system may be coupled to a thermostat to keep the temperature constant. Many different thermostats have been devised,[34] which differ in the precision with which they can keep the system in the canonical ensemble and the ease with which they can be implemented into a molecular dynamics program.

#### *The Problem of Quasi-Nonergodicity*

When estimating an ensemble average of an observable $A(\boldsymbol{r})$ from a Monte Carlo or a molecular dynamics simulation, the observable is averaged over all generated configurations. In a molecular dynamics

simulation, the ensemble average $\langle A \rangle$ is estimated by calculating the time average

$$\bar{A} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau A(\boldsymbol{r}(t))dt. \qquad (16)$$

The assumption that the time average equals the ensemble average

$$\bar{A} = \langle A \rangle, \qquad (17)$$

is called the ergodic hypothesis.[31] Most systems cannot be proven to have ergodic behavior. However, if during a simulation all "important" configurations, i.e., all configurations for which the probability $\rho(\boldsymbol{r})$ is significant are visited, then eq. (17) will hold to a first approximation. If regions of high probability in configurational space are separated by significant (free) energy barriers, it is unlikely that all important configurations will be sampled. This is known as the problem of quasi-nonergodicity. In a molecular dynamics simulation, the average kinetic energy per degree of freedom is only about $k_B T/2$, which means that the larger a barrier, the longer it will take for the barrier to be crossed. Quasi-nonergodicity can also occur in Monte Carlo simulations. Well chosen random moves can alleviate the problem, the design of such moves is, however, far from trivial.

The success of the Metropolis Monte Carlo (MC) and molecular dynamics (MD) simulation methods is due to their ability to "restrict the sampling of the configuration space to the extremely small fraction that contributes significantly to most properties of interest."[35] That is, properties for which the ensemble average is dominated by the probability distribution [see eq. (14)]. For the estimation of these properties, sampling algorithms such as MD or MC, that generate Boltzmann distributed ensembles of configurations, are in principle ideal. In practice, problems occur due to quasi-nonergodicity, i.e., the inability to visit all important configurations during the simulation. In these cases so called enhanced sampling methods, which will be discussed in the remainder of this section, may be applied.

*Methods to Alleviate the Problem of Quasi-Nonergodicity*

In the following we discuss methods that enhance the sampling of configurational space. Sampling problems can be loosely divided into two categories. Either there exists some knowledge in regard to those regions of configurational space not sampled sufficiently or there is no such knowledge. For example, a particular molecule may have two preferred conformations but only one is sampled during a simulation. In such cases, it is possible to define a variable as a function of $\boldsymbol{r}$, e.g., a torsional angle, which can be used to discriminate between the two conformations. The simulation protocol can then be changed such that configurations which correspond to values of this variable over a predefined range are sampled. Whenever such a variable, which may be an arbitrarily complicated function of the atomic positions of the molecules, can be identified we are faced with a sampling problem which resembles the one encountered in calculations of the free energy difference between two states A and B. The desired "end states" (e.g., conformations) are known. The challenge is to define a path which connects these end states, and to ensure that sampling is performed along the whole path. Because of the similarity between these sampling problems and those encountered

when estimating the difference in free energy between two states A and B, they will be discussed in "Sampling Relative Probabilities of States" Section.

If the important regions of configurational space are not known, a number of techniques can be used to enhance sampling. These include:

1. Changing the dynamics without changing the potential energy surface to emphasize motion or sampling along the slow degrees of freedom relative to the fast degrees of freedom;
2. Deforming the energy surface to increase the probability with which (free) energy barriers can be surmounted;
3. Extending the dimensionality to circumvent energy barriers;
4. Perturbing the forces;
5. Reducing the number of degrees of freedom;
6. Multi-copy approaches.

Only using technique 1 is the probability distribution of the configurations sampled not modified. Techniques 2–6 sample configurations from a modified probability distribution which focuses sampling on configurations relevant to a particular property as opposed to generating a particular distribution. Below the various approaches are illustrated using examples from the literature. Note this is not a complete list of methods that combine the various techniques. Our aim is rather to assist the reader to categorize new methods he or she encounters.

*Emphasizing Slow Degrees of Freedom Relative to Fast.* The equilibrium properties [eq. (14)] of a system described by a Hamiltonian such as eq. (4) are independent of the dynamics of the system. Hence, the dynamics of the system can be changed without altering the probability distribution of configurations [eq. (13)]. This can either be done by choosing a different functional form for the kinetic term of the Hamiltonian [eq. (5)], by changing the masses of selected particles, or by the direct modification of the momentum distribution. The goal in all of these approaches is to slow down high frequency motions and to speed up low frequency motions. Slowing down high frequency motions allows a larger time step to be used for the integration of the discretized equations of motion, speeding up low frequency motions allows configurations associated with these motions to be sampled within a shorter simulation time. Overall, this leads to configurational space being explored more efficiently. One approach to changing the functional form of the kinetic term is to replace it by a more general quadratic form $\frac{1}{2}\sum_{ij=1}^{N}\dot{r}_i M_{ij}\dot{r}_j$ where $M_{ij}$ is the so called "mass tensor."[36] An appropriate choice of $M_{ij}$ can be used to equalize the speed of the various motions in the system. However, liquid phase systems are anharmonic and devising an appropriate mass tensor is difficult. Jacucci and Rahman changed the masses in a water model to equalize the three moments of inertia in the molecule[37] allowing an increase of the integration time step. Increase of the hydrogen mass has also been shown to lead to more efficient sampling in more complex systems.[38,39] A pitfall when changing the masses is that the gain in efficiency through the ability to use a larger integration time step may be partially offset by a reduction of diffusivity, and therefore, a slower exploration of configurational space within a given length of simulation. A method that directly changes the momenta is momentum-enhanced

hybrid Monte Carlo (ME-HMC).[40] In HMC,[41] random momenta are assigned according to a Maxwell distribution. Several steps of MD are then performed using a (too) large time step. The move is then accepted or rejected according to a Metropolis-type criterion. In ME-HMC, random momenta are assigned according to a modified distribution with the aim of biasing the dynamics towards sampling of specific slow motions which are identified by time-averaging over the momenta. Here, the averaging time is a crucial but not easily determined parameter.

The methods presented above aim at slowing down the motion along the fast degrees of freedom and at speeding up the motion along the slow ones. An alternative approach is to leave the dynamics unchanged but to use different integration time steps for different parts of the system. In many of these multiple time step algorithms,[42–44] the time step used in the interaction calculation increases with the distance by which the particles involved are separated. The implementation of these methods requires great care to avoid numerical instabilities. Other challenges involve resonance effects, energy drifts, the integration of these methods with lattice sum approaches for the calculation of the electrostatic interactions, and their generalization to various thermodynamic ensembles.

*Sampling Enhancement Through Deformation of the Potential Energy Surface.* When sampling from a canonical distribution probability is proportional to $\exp[-\beta V(\boldsymbol{r})]$. This means that the probability of sampling a given configuration decreases exponentially as the energy increases. Barrier crossings are therefore rare events and important regions of configurational space may not be visited. Sampling can be improved by changing the probability distribution such that the probability of high energy configurations is increased. Sampling from a distribution other than the canonical one can be expressed in terms of canonical sampling where an additional biasing potential energy term has been added to the Hamiltonian. The configurational space is enlarged to enable barriers to be crossed. The ensemble that is sampled should, however, not deviate significantly from the original canonical distribution as otherwise time will be spent sampling configurations which do not contribute to the property of interest [eq. (14)].

One approach to increase the probability of visiting high energy states is to simulate at elevated temperature.[45–48] This corresponds to a simulation using an effective Hamiltonian where all interactions are scaled down by a constant factor. This approach can, however, quickly lead to the low energy configurations which contribute most to the ensemble average [eq. (14)] not being sampled sufficiently.

In the multi-canonical ensemble method[49–51] and related approaches,[52] the probability of visiting high energy states is enhanced in a more controlled fashion. Although there might be no information available on the important configurations which are not sampled, the energy can always be chosen as a variable along which sampling is enhanced. In the multi-canonical ensemble method, the sampling scheme is modified such that the probability distribution is flat over a certain energy range. This is achieved by giving weights to the different macrostates which correspond to a given energy value. These weights must be determined iteratively as they correspond to the probability, i.e., the free energy, of a given energy macrostate (see also "Estimation of the Free Energy" Section). A "flat" sampling over a given energy range is also obtained in the Wang-Landau

sampling method.[53] The acceptance criterion is chosen such that each energy level is visited with the same probability. The weights are adapted during the simulation multiplying by a factor each time a particular energy level is visited. Other methods that can be viewed as extensions of the multi-canonical ensemble method include the integrate-over-temperature approach of Gao[54] and the equi-energy sampler of Kou et al.[55]

Instead of achieving uniform sampling in the energy, sampling can be performed according to a generalized probability distribution function. One such generalized probability distribution function was suggested by Tsallis.[56] By adaptation of a parameter in the sampling distribution, sampling can be biased towards sampling of high energy regions or low energy regions. For a particular choice of parameter, the canonical ensemble distribution is recovered. Sampling according to a Tsallis distribution can be reformulated as canonical sampling of an effective Hamiltonian.[57] The approach has been generalized to allow sampling from a broad class of probability distribution functions.[58]

Other strategies to enhance sampling that rely on a direct modification of the energy surface are hyperdynamics,[59–61] accelerated molecular dynamics,[62,63] puddle jumping,[64] and related approaches.[65] In these methods, energy wells are filled up or elevated to enable more frequent barrier crossing. Care has, however, to be taken to ensure that important low energy regions of configurational space are sampled sufficiently.

The molecular model can also be modified directly to make the potential energy surface smoother. This can, for e.g., be done by using soft-core potentials.[66–69] These potentials may be used instead of the van der Waals and electrostatic non-bonded potential energy terms normally used. They do not show a singularity at zero interatomic distance but return a tunable, finite energy. This can be used to decrease the potential energy barriers between conformational states.

*Extension of Dimensionality.* An extension of the three-dimensional physical space to more dimensions can also lead to more effective barrier crossing by opening up new low energy pathways.[70,71] Mapped back to three dimensions, the potential energy surface of the system with increased dimensionality corresponds to a system in which the potential energy surface has been deformed.

*Modifying the Forces.* Enhanced sampling can also be achieved by the direct modification of the forces acting on the system. In the approach of Wu and Wang[72–74] the sampling of configurational space is accelerated by the addition of a guiding force which, in an ideal case, corresponds to the gradient of the local free-energy surface. The guiding force is estimated from the forces experienced by the system by the use of a memory function. As the method uses time-averaged information, it generates irreversible trajectories which may lead to errors in the calculated canonical averages.[40] In the PEACS method[75] the forces are modified such that a set of configurations with constant potential energy is generated. This leads to enhanced sampling.

*Reducing the Number of Degrees of Freedom.* The number of degrees of freedom in the system determines the number of energy and force evaluations that are necessary. Degrees of freedom that

are believed to have no influence on the property of interest should be omitted. This is especially true in the case of degrees of freedom involved in fast motion, as their omission can lead to a substantial increase in efficiency due to the possibility of using a larger integration time step. A very common example is the application of bond constraints using, for e.g., the SHAKE[76] algorithm. Bond vibrations belong to the fastest molecular motions which are generally modeled in a biomolecular simulation and the application of bond length constraints allows the use of a two to four times larger integration time step at a 10–20% increase in computing effort.[32]

When modeling larger biomolecular systems such as lipid bilayers models with a reduced number of degrees of freedom compared to atomistic models are used frequently,[77] In these coarse-grained models,[78] groups of atoms are merged into one particle (see also "Molecular Model" Section) resulting in a smoother potential energy surface. Simulations using coarse-grained models show enhanced diffusion and explore configurational space quicker. Furthermore, the smoother potential energy surface allows the use of a larger integration time step.[79]

A drastic reduction in number of degrees of freedom can be obtained by the use of implicit solvent models.[80, 81] Here, the solvent is represented as a continuum. As in general a large number of molecules is needed to model the solvent explicitly, implicit solvent models can yield a considerable increase in computational efficiency. However, although the reduction in the number of degrees of freedom that must be treated leads to a faster exploration of configurational space, there is no guarantee that the configurations obtained represent the canonical probability distribution of the original system. In some cases, this problem can be alleviated by combining the approaches discussed above with multi-copy methods (see below).

*Multi-Copy Approaches.* In multi-copy approaches, several replicas of the system are simulated, e.g., at various temperatures.[82–84] After a certain number of steps, an exchange of replicas is attempted and accepted or rejected based on a Metropolis Monte Carlo criterion. This swapping procedure enables the enhanced sampling at the higher temperature replica to be propagated into the low temperature replica while still maintaining an appropriate probability distribution. This approach can be generalized such that the Hamiltonian as opposed to the temperature is varied between replicas.[85–87] For example, different levels of soft-core potentials ranging from normal nonbonded interactions to very soft nonbonded interactions which enhance sampling may be used in the different replicas.[88] Hamiltonian replica exchange can also be used to connect models of varying "grain level."[89–92] These approaches can be used to exploit the sampling abilities of coarse-grained models without loosing information at an atomic level.

### Sampling Relative Probabilities of States

Canonical sampling at the end states is in general not optimal when one wishes to estimate the difference in free energy between these states. As has been pointed out at the beginning of the section using energy difference distributions $\rho(\Delta H; \Delta H_{BA})$, it is necessary to sample all important configurations of the end states. How this can be achieved and which obstacles are encountered has been discussed in "Sampling Configurations of One State" Section. The difference in free energy between two states A and B can be identified as the energy difference at the point where the two energy difference distributions $\rho_A(\Delta H; \Delta H_{BA})$ and $\rho_B(\Delta H; \Delta H_{BA})$ intersect (see "Sampling" Section). Therefore, the two (or more) states have to be connected as part of the same overall system to allow sufficient sampling in the intersection region. In the following we will discuss how such a combined Hamiltonian can be constructed and which strategies can be applied to ensure the configurational distribution function is sampled appropriately.

### Definition of a Combined Hamiltonian

Assume we want to calculate the free energy difference between two states A and B [eq. (3)]. We wish, therefore, to define a combined Hamiltonian $H_{comb}$ such that the important configurations of this combined Hamiltonian are primarily composed of configurations important to states A and B. The combined Hamiltonian will be some function of the end state Hamiltonians $H_A$ and $H_B$, or in practice $V_A$ and $V_B$. It can either

- be dependent on a coupling parameter $\lambda$ such that for a particular value of $\lambda$ (e.g., 0) $V_{comb} = V_A$ and for another (e.g., 1) $V_{comb} = V_B$, or
- not be explicitly dependent on a coupling parameter although it may still be possible to define a function $\lambda = \lambda(\boldsymbol{r})$ such that for $\lambda = 0$ we obtain $V_{comb} \approx V_A$ and for $\lambda = 1$ we obtain $V_{comb} \approx V_B$.

The first approach, the coupling-parameter approach,[93] has been widely used in free energy calculations. As the free energy is a state function, i.e., it is path independent [see also eq. (3)], the dependence of the combined Hamiltonian on $\lambda$ can be chosen freely. This is of course not true if there exists a physical path along which we would like to know the potential of mean force, i.e., the free energy as a function of $\lambda$. The $\lambda$ dependence could therefore be chosen such that sampling problems are minimized.[94] It has long been realized that a simple linear combination $V_{comb} = \lambda V_B + (1 - \lambda)V_A$ leads to numerical problems when particles are deleted during the perturbation. This is due to the singularity in the Lennard-Jones interaction term [eqs. (7) and (8)]. There are several ways to let a particle "smoothly" disappear. This can either be achieved via a non-linear scaling scheme or by making the end state Hamiltonians explicitly dependent on $\lambda$ as in the so-called "soft-core potentials."[95, 96]

The second approach of combining two (or more) Hamiltonians has been less frequently used. Possible schemes include

- the combination of $H_A$ and $H_B$ following a valence bond formulation.[97] The combined potential energy function then reads

$$V_{comb}(\boldsymbol{r}) = \Big(V_A(\boldsymbol{r}) + V_B(\boldsymbol{r}) + E_A^R + E_B^R - $$
$$\Big((V_A(\boldsymbol{r}) - V_B(\boldsymbol{r}) - E_A^R + E_B^R)^2 + \beta^{-2}\Big)^{-1/2}\Big)/2, \quad (18)$$

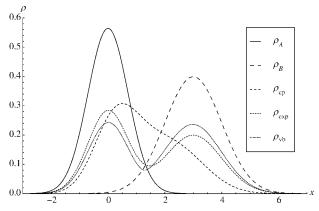where $E_A^R$ and $E_B^R$ are adjustable parameters.

**Figure 1.** Pictorial representation of possible state coupling schemes. The potential energy function of state A, $y_A = x^2$, is coupled to the potential energy function of state B, $y_B = \frac{1}{2}(x - 3)^2$ in three different ways: Following a coupling parameter approach $y_{cp} = \lambda y_B + (1 - \lambda)y_A$, via a sum of Boltzmann factors $y_{exp} = -\beta^{-1} \ln \left[ \exp(-\beta y_A) + \exp(-\beta y_B) \right]$, or using a valence bond formulation $y_{vb} = (y_A + y_B - ((y_A - y_B)^2 + \beta^{-2})^{\frac{1}{2}})/2$. The normalized probability distributions are shown [eq. (13), $\beta = 1$]. In case of the coupling parameter approach, the distribution was integrated over $\lambda$ before normalization.

- A combination of $H_A$ and $H_B$ by an exponential sum followed by taking the logarithm.[98–102] That is,

$$V_{comb}(\boldsymbol{r}) = -(\beta s)^{-1} \ln \left\{ \exp \left( -\beta s \left( V_A(\boldsymbol{r}) - E_A^R \right) \right) + \exp \left( -\beta s \left( V_B(\boldsymbol{r}) - E_B^R \right) \right) \right\}, \quad (19)$$

where $E_A^R$, $E_B^R$, and $s$ are adjustable parameters.

A possible choice of state indicator function for these combined Hamiltonians would be $\lambda(\boldsymbol{r}) = \exp(-\beta V_B(\boldsymbol{r}))/[\exp(-\beta V_A(\boldsymbol{r})) + \exp(-\beta V_B(\boldsymbol{r}))]$. Note that it is not possible to make a clean separation between those approaches that explicitly depend on $\lambda$ and those that do not. In the methods mentioned above, the parameters $E_A^R$ and $E_B^R$ can be chosen such that $V_A$ or $V_B$ are recovered. For $E_A^R \gg E_B^R$ we obtain $V_{comb} \approx V_A$, i.e., we could, e.g., choose $E_B^R = 0$ and $E_A^R = n(\lambda - \frac{1}{2})$, with a sufficiently large, positive constant $n$. The resulting $\lambda$-dependent combined Hamiltonian yields $V_{comb} \approx V_A$ for $\lambda = 0$ and $V_{comb} \approx V_B$ for $\lambda = 1$. Note, however, that also for $E_A^R = E_B^R = 0$ the important configurations of the combined Hamiltonian comprise those of states A and B. This can be seen in Figure 1, which depicts different choices of combined Hamiltonians for a harmonic oscillator example.

The choice of combined Hamiltonian has a critical influence on the efficiency of the free energy estimation. Combined Hamiltonians should be designed such that the important configurations of states A and B are among the important configurations of the combined state. However, due to the choice of path connecting states A and B in the combined Hamiltonian, the combined state will necessarily contain additional configurations with significant Boltzmann weights. Ideally, these configurations are those which

correspond to the intersection region of the energy difference distributions [eq. (11)]. However, this need not be the case. A poorly chosen combined Hamiltonian will contain a high proportion of configurations which are irrelevant for the free energy estimation. Irrelevant are those configurations which neither are of importance to the end state nor correspond to configurations of the intersection region of the energy difference distributions.

### Ways to Sample the Combined Hamiltonian

Once a combined Hamiltonian has been defined, a sampling scheme must be chosen that ensures all important configurations of the combined state are sampled. For combined Hamiltonians which comprise a coupling parameter $\lambda$, a scheme must be chosen which ensures the whole $\lambda$ range is sampled. Sampling may be focused to $\lambda$ regions of primary interest (e.g., at the end states $\lambda = 0$ and $\lambda = 1$). The strategies discussed below can also be applied in simulations of only one state in cases where sampling should be enhanced along a given path variable $\lambda(\boldsymbol{r})$ (see "Methods to Alleviate the Problem of Quasi-nonergodicity" Section). The variable $\lambda$ appears in the literature with different names depending on the type of perturbation described. These names include collective variable, reaction coordinate, coupling parameter, order parameter, and others. In the following the three basic sampling enhancement strategies staging, importance sampling, and adiabatic decoupling will be presented and discussed.

*Staging.* In staging, stratification, or windowing, several independent simulations at different values of $\lambda$ are performed.[103, 104] The system can be constrained to a particular value of $\lambda$. This is common practice when $\lambda$ appears as a parameter in the combined Hamiltonian. If $\lambda$ is a function of the positions of the particles, a restraint is commonly applied, e.g., a harmonic biasing potential energy term. Constraints can, however, also be applied using, e.g., the method of Lagrangian multipliers.[105] The number of staging windows should be chosen such that the energy difference distributions at subsequent $\lambda$ values overlap. Recent extensions of this approach include replica-exchange methods[106] where configurations between the different $\lambda$ windows are exchanged after regular time intervals.

*Importance Sampling.* In importance sampling, one simulation using the combined Hamiltonian is performed. If $\lambda$ appears as a parameter in the Hamiltonian, a scheme that allows $\lambda$ to be changed has to be introduced. This can, e.g., consist of Monte Carlo moves in $\lambda$.[107] Another possibility is to treat $\lambda$ as an additional degree of freedom with an associated mass.[108] The combined Hamiltonian is now changed such that the desired probability distribution along $\lambda$ is obtained. This can be done by adding a biasing potential energy term or by the direct modification of the parameters in the combined Hamiltonian [e.g., by adjusting $E_A^R$, $E_B^R$, and $s$ in eq. (19)]. The strategy of adding a biasing potential energy term was introduced by Torrie and Valleau[104] under the name umbrella sampling. They proposed a staging approach where a different biasing potential energy term is applied in each simulation window. These local (in terms of $\lambda$ values) biasing potential energy terms are often easier to determine than a single biasing potential energy term which aims at sampling over the whole range of $\lambda$.

To build such a biasing term, the change in free energy as a function of λ has to be computed. How free energy differences can be estimated from the configurations sampled is discussed in the next section. The bias is normally built iteratively as the free energy is the quantity to be estimated and is not known at the beginning of the simulation.

Which probability distribution is desired may vary. If the whole free energy profile along λ is of interest, a uniform probability along λ is convenient. In cases where we are only interested in the relative free energy of the states described by the extreme λ values (e.g., 0 and 1), a probability distribution that gives more weight to these end states may be a more appropriate choice. The probability at intermediate λ values must, however, be large enough to allow for crossing events between the two end states.

Instead of implementing the bias into the Hamiltonian, the force can be modified. Instead of adding a biasing potential energy term to the Hamiltonian based on an estimate of the free energy, a bias based on the derivative of the free energy can be added to the force. In the adaptive biasing force method,[109–111] this derivative of the free energy with respect to λ is estimated from the mean force on λ (see "Estimation of the Free Energy" Section). This force is then removed to obtain uniform sampling along λ.

*Adiabatic Decoupling.* In adiabatic decoupling or separation, the aim is to decouple the motion along λ from all other degrees of freedom in the system.[112–115] That is, one or more collective variables (λ) are adiabatically decoupled from the rest of the system. Assignment of a higher mass relative to the other degrees of freedom ensures the adiabaticity of the decoupling. In the case of perfect adiabatic separation, the free energy along λ can be obtained directly from the probability distribution function along λ observed. Allowing the collective variables to evolve at high temperature can result in a rapid mapping of the free energy profile (or surface) along the collective variable(s). If the Hamiltonian does not depend explicitly on λ, but the collective variable λ is instead a general function of the particle positions, i.e., λ(*r*), two alternative approaches can be taken. One can perform a coordinate transformation from Cartesian coordinates *r* to generalized coordinates *q*(*r*), which explicitly contain the collective variable(s).[112,115] However, this method requires extensive modification of the molecular dynamics software. The second possibility is to treat the collective variables as extended phase space variables[114,115] similar to the approach used in metadynamics.[116] The advantage of this second approach is that no coordinate transformation is needed. As a consequence it is easier to implement into existing software. The disadvantage is the need for a harmonic coupling force constant as an additional parameter. Multiple time scale integration[43,44,115] may be used to integrate this harmonic force term which is inexpensive but oscillates rapidly.

*Embarrassment of Riches.* Which of these three strategies is optimal will depend on the system of interest. More precisely, the choice of sampling protocol will depend on the relaxation times of the degrees of freedom orthogonal to the "reaction coordinate" λ. If these relaxation times are long, then problems may occur when using importance sampling or adiabatic decoupling because the system is never in equilibrium. In such cases, a staging approach, in which simulations at different but fixed λ-values are used, may be more appropriate as the system is allowed to relax to equilibrium in each

of the simulation windows. Another problem that can occur when using importance sampling is due to slow diffusion along λ. Even if the probability distribution along λ is flat the time required for the system to diffuse along the reaction coordinate can be prohibitively long. This problem is not encountered in staging approaches or adiabatic decoupling. In the former, the system is constrained to sample over the whole range of λ. In the latter, the evolution of λ at elevated temperature ensures the diffusion along the reaction coordinate is rapid. Staging approaches, on the other hand, may suffer from broken ergodicity within a sampling window. This is likely to occur if important configurations corresponding to a particular λ value are separated by large barriers. This sampling problem might be self-inflicted as there may exist a low energy path connecting these important configurations which, however, requires variations in λ. Last but not least, all three approaches can be combined. Importance sampling and adiabatic decoupling can be applied simultaneously within a staging window.

## Estimation of the Free Energy

### *Common Free Energy Estimators*

Once an appropriate set of configurations has been sampled ("Sampling" Section), various approaches can be used to estimate the difference in free energy [eq. (3)]. Compared to the computational effort of generating configurations, the calculation of the free energy from those configurations is comparatively inexpensive. Traditionally, particular sampling protocols have been used with particular estimation schemes. However, in many cases, different estimators can be applied to the same data. The most commonly used estimation methods are:

1. Global methods:
   a. *Counting the number of visits (visited states method):* Here, one tries to estimate the probability distribution along λ. Let $\lambda_A$ and $\lambda_B$ be the λ values corresponding to state A and B, respectively. The relative probability of $\lambda_A$ and $\lambda_B$ ($p(\lambda_A)$ and $p(\lambda_B)$) can be related to the ratio of partition functions of states A and B.

$$\frac{p(\lambda_B)}{p(\lambda_A)} = \frac{\iiint \exp(-\beta H(\boldsymbol{p},\boldsymbol{r};\lambda))\,\delta(\lambda-\lambda_B)\,d\boldsymbol{p}d\boldsymbol{r}d\lambda}{\iiint \exp(-\beta H(\boldsymbol{p},\boldsymbol{r};\lambda))\,\delta(\lambda-\lambda_A)\,d\boldsymbol{p}d\boldsymbol{r}d\lambda} = \frac{Q_B}{Q_A},$$
   (20)

and, therefore, [using eq. (3)]

$$\Delta F_{BA} = -\beta^{-1} \ln \frac{p(\lambda_B)}{p(\lambda_A)}.$$
   (21)

In cases where the configurations sampled have been obtained from a simulation in which λ is allowed to vary, the relative free energies along λ can be obtained from a histogram of λ. To this end it is irrelevant whether λ was an independent degree of freedom during the sampling or not. It can simply be counted how often a given λ value is encountered during the simulation. The bin size is determined by the desired resolution and precision. Increasing the bin size leads to a loss

in resolution and a gain in precision and vice versa. If the ensemble sampled corresponds to a biased probability distribution (as do ensembles obtained from importance sampling) the histogram obtained has to be reweighted to the desired ensemble. An ensemble average $\langle X \rangle_{\text{sampled}}$ of an observable $X$ can be reweighted to a different ensemble using

$$\langle X \rangle_{\text{desired}} = \frac{\left\langle X \exp\left[-\beta\left(V_{\text{desired}} - V_{\text{sampled}}\right)\right]\right\rangle_{\text{sampled}}}{\left\langle \exp\left[-\beta\left(V_{\text{desired}} - V_{\text{sampled}}\right)\right]\right\rangle_{\text{sampled}}} . \quad (22)$$

b. *Using energy differences:* The difference in free energy between two states A and B can be related to an exponential average over the energy difference $\Delta V_{\text{BA}}(\boldsymbol{r}) = V_{\text{B}}(\boldsymbol{r}) - V_{\text{A}}(\boldsymbol{r})$ between those states:[117]

$$\Delta F_{\text{BA}} = -\beta^{-1} \ln \langle \exp\left[-\beta \Delta V_{\text{BA}}\right]\rangle_{\text{A}} \quad (23)$$

$$= -\beta^{-1} \ln \left\{ \int \exp\left[-\beta \Delta V_{\text{BA}}\right] \rho_{\text{A}}(\boldsymbol{r}) d\boldsymbol{r} \right\} \quad (24)$$

$$= -\beta^{-1} \ln \left\{ \int \exp\left[-\beta \Delta V\right] \rho_{\text{A}}(\Delta V; \Delta V_{\text{BA}}) d\Delta V \right\}, \quad (25)$$

where $\rho_{\text{A}}(\boldsymbol{r})$ is the probability distribution of positions corresponding to state A and $\rho_{\text{A}}(\Delta V; \Delta V_{\text{BA}})$ is the probability distribution of energy differences $\Delta V_{\text{BA}}$ also corresponding to state A.[25] For simplicity, we have neglected kinetic contributions to the free energy. This method is known as free energy perturbation (FEP). The difference in free energy can be obtained from a single simulation at state A. However, canonical sampling at state A is not optimal to estimate the average eq. (23), as it is not dominated by the probability distribution but by the exponential factor $\exp\left[-\beta \Delta V_{\text{BA}}\right]$. Only configurations corresponding to the lowest values of $\Delta V_{\text{BA}}$ will contribute significantly to the average. However, these contributions normally lie in the tail of the $\rho_{\text{A}}(\Delta V; \Delta V_{\text{BA}})$ distribution, i.e., they have low probability. Therefore, sampling at state A is suboptimal as it does not concentrate the sampling effort on those configurations which contribute most to the average.

At the beginning of "Sampling" Section we have introduced the energy difference distributions [eqs. (9) and (10)] and have shown that the difference in free energy corresponds to the energy difference where those two distributions intersect [eq. (11)]. This suggests that we can obtain the difference in free energy from two simulations at states A and B. Specifically, the energy difference distributions $\rho_{\text{A}}(\Delta V; \Delta V_{\text{BA}})$ and $\rho_{\text{B}}(\Delta V; \Delta V_{\text{BA}})$ can be constructed and the intersection point determined. This implies that the distributions obtained must overlap. If this is not the case, staging must be used and the overall change split into a series of smaller perturbations. The total free energy change is then obtained by summing over the differences in free energy between subsequent windows. The simplest way of estimating the energy difference distributions is by constructing a histogram. One may, however,

also model the energy difference distributions as a series expansion[118, 119] or as an analytical function.[120] This has the advantage that the adjustable parameters of the model distribution will be determined mainly by those regions of the distributions with high probability leading to less noise in the tail regions. The disadvantage is, however, that the choice of model function is arbitrary and lacks a physical basis. Once a model for the energy difference distributions is obtained from the simulation data, the difference in free energy can be estimated from the intersection point [see eq. (11)] or by a linear regression analysis using the relation.[25–27]

$$\ln \frac{\rho_{\text{A}}(\Delta V; \Delta V_{\text{BA}})}{\rho_{\text{B}}(\Delta V; \Delta V_{\text{BA}})} = +\beta \Delta V - \beta \Delta F_{\text{BA}} . \quad (26)$$

Estimation of differences in free energy via the energy difference distribution can be performed whenever the sampling scheme allows (possibly by reweighting) the energy difference distributions to be reconstructed with sufficient accuracy.

Another estimator of the difference in free energy from two simulations has been introduced by Bennett.[121] Instead of perturbing directly from A to B two perturbations to an intermediate state R can be performed

$$\Delta F_{\text{BA}} = \Delta F_{\text{BR}} - \Delta F_{\text{AR}} = -\beta^{-1} \ln \frac{Q_{\text{B}}}{Q_{\text{A}}} \frac{Q_{\text{R}}}{Q_{\text{R}}} , \quad (27)$$

$$\Delta F_{\text{BA}} = -\beta^{-1} \ln \frac{\left\langle e^{-\beta(V_{\text{R}}-V_{\text{A}})}\right\rangle_{\text{A}}}{\left\langle e^{-\beta(V_{\text{R}}-V_{\text{B}})}\right\rangle_{\text{B}}} . \quad (28)$$

The optimal intermediate state is determined iteratively in a postprocessing step and is given by

$$e^{-\beta V_{\text{R}}(\boldsymbol{r})} = \left[\rho_{\text{A}}(\boldsymbol{r})^{-1} + \rho_{\text{B}}(\boldsymbol{r})^{-1}\right]^{-1} \text{const},$$

$$V_{\text{R}}(\boldsymbol{r}) = \beta^{-1} \ln \left[e^{\beta(V_{\text{A}}(\boldsymbol{r})-F_{\text{A}})} + e^{\beta(V_{\text{B}}(\boldsymbol{r})-F_{\text{B}})}\right] + \text{const}. \quad (29)$$

It corresponds to a state with a $\rho_{\text{R}}(\Delta V_{\text{BA}})$ distribution that shows the highest probability in the intersection region of the $\rho_{\text{A}}(\Delta V_{\text{BA}})$ and $\rho_{\text{B}}(\Delta V_{\text{BA}})$ distributions. Again, the perturbation must be split into multiple windows in cases where the energy difference distributions do not overlap.

Instead of performing two simulations one can estimate the free energy difference from a single simulation of a reference state

$$\Delta F_{\text{BA}} = -\beta^{-1} \ln \frac{\left\langle e^{-\beta(V_{\text{B}}-V_{\text{R}})}\right\rangle_{\text{R}}}{\left\langle e^{-\beta(V_{\text{A}}-V_{\text{R}})}\right\rangle_{\text{R}}} . \quad (30)$$

The reference state needs to envelop the important configurations of both states A and B. This approach can, therefore, be applied whenever we have sampled the probability distribution of the combined state within a single simulation using,

e.g., importance sampling or adiabatic decoupling (see "Ways to Sample the Combined Hamiltonian" Section). Possible choices for the Hamiltonian of the combined state are given by eqs. (18) and (19). Choosing $s = 1$, $E_A^R = F_A$, and $E_B^R = F_B$ in eq. (19) corresponds to the reference state Hamiltonian that minimizes the expected error of eq. (30).[98, 102, 121]

Equilibrium free energy differences can also be obtained using non-equilibrium methods. Replacing the energy difference $\Delta V_{BA}(\boldsymbol{r}) = V_B(\boldsymbol{r}) - V_A(\boldsymbol{r})$ in eq. (23) by the work $W_{BA}$ performed in a non-equilibrium transformation from state A to state B, we obtain Jarzynski's identity.[3] The ensemble average $\langle \cdot \rangle$ then corresponds to an averaging over initial conditions and different paths. The non-equilibrium counterpart to Bennett's optimal estimator [eqs. (28) and (29)] is Crooks' identity,[122, 123] which is again obtained by replacing $\Delta V_{BA}$ by $W_{BA}$. The non-equilibrium version of eq. (30) has to our knowledge not yet been discussed in the literature. Non-equilibrium estimators show similar strengths and weaknesses as their equilibrium counterparts. As discussed earlier, Bennett's estimator [eqs. (28) and (29)] will in general perform better than the FEP estimator [eq. (23)]. Similarly, Crooks' identity is to be preferred over Jarzynski's. It is an open question in which situations non-equilibrium approaches to free energy differences may outperform equilibrium methods.

2. Local methods:
   a. *Force estimation:* Another class of methods aims at estimating the derivative of the free energy with respect to $\lambda$ followed by integration.

$$\Delta F_{BA} = F_B - F_A \tag{31}$$

$$= F(\lambda_B) - F(\lambda_A) \tag{32}$$

$$= \int_{\lambda_A}^{\lambda_B} \frac{dF}{d\lambda} d\lambda. \tag{33}$$

These so called thermodynamic integration (TI) methods go back to Kirkwood.[124] Inserting eqs. (1) and (2) into eq. (33) we obtain

$$\frac{dF}{d\lambda} = \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda, \tag{34}$$

where $\langle \cdot \rangle_\lambda$ indicates an average over configurations corresponding to a given $\lambda$ value. The difference in free energy can thus be estimated as an integral over a generalized mean force on $\lambda$. How the partial derivative of the Hamiltonian with respect to $\lambda$ is calculated depends on the sampling scheme and the form in which $\lambda$ appears in the Hamiltonian (see "Definition of a Combined Hamiltonian" Section).

If the combined Hamiltonian shows parametric dependence on $\lambda$ (coupling parameter approach) and staged sampling was performed, the evaluation of $\left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda$ is trivial. The partial derivative can be calculated analytically and the averages from simulations at different $\lambda$ values. The free energy difference is then obtained by numerical quadrature.

If the Hamiltonian shows a non-parametric dependence on $\lambda$, i.e., $\lambda = \lambda(\boldsymbol{r})$, the evaluation of the partial derivative involves more complicated expressions. Two examples shall be discussed. The first is the so-called blue-moon ensemble method (BM)[105, 125] and the second is the adaptive biasing force (ABF)[109–111] method. In BM, staged sampling is performed. Within a window, $\lambda(\boldsymbol{r})$ is fixed by applying a constraint force. The derivative of the free energy is then estimated from the average of the constraint force. In ABF, an importance sampling scheme is used (see also "Ways to Sample the Combined Hamiltonian" Section). That is, $\lambda$ is allowed to change during the simulation and uniform sampling along $\lambda$ is achieved by removing the mean force on $\lambda$. Darve et al.[109–111, 126] have shown that the average force on $\lambda(\boldsymbol{r})$ can be calculated from an expression that involves only time derivatives of $\lambda$

$$\frac{dF}{d\lambda} = -\left\langle \frac{d}{dt} \left( m_\lambda \frac{d\lambda}{dt} \right) \right\rangle_\lambda, \tag{35}$$

with $1/m_\lambda = \sum_i^N 1/m_i (\partial \lambda / \partial \boldsymbol{r}_i)^2$. That is, the derivative of the free energy can be calculated by binning the observed $\lambda$ values and evaluating eq. (35) to calculate the average force.

   b. *Estimation of transition probabilities:* Free energy differences can also be calculated by monitoring the transition probabilities between macrostates. Let $\rho_i$ be the probability of macrostate $i$. This probability corresponds to a sum over probabilities of configurations (microstates), which have the same value of a chosen observable. In our context, a macrostate $i$ could include all configurations that correspond to a given value of $\lambda$. Let $\rho(i \rightarrow j)$ denote the transition probability from macrostate $i$ to macrostate $j$. It can be shown[51] that the macrostate transition probability satisfies detailed balance

$$\frac{\rho_j}{\rho_i} = \frac{\rho(i \rightarrow j)}{\rho(j \rightarrow i)}. \tag{36}$$

The vector of macrostate probabilities can be obtained from the eigenvector of the transpose of the transition probability matrix. From the relative macrostate probabilities, the relative free energies can then be obtained [see eq. (21)]. Equation (36) was derived under the assumption that during the simulation a local equilibrium within the current macrostate may be established. This is only true if the macroscopic variables, i.e., in our case $\lambda$, move slowly compared to all other degrees of freedom. That is, if the degrees of freedom perpendicular to $\lambda$ have time to relax to equilibrium. A transition probability matrix may be estimated from any simulation at the combined state that allows changes in $\lambda$. If Monte Carlo moves in $\lambda$ are attempted, the acceptance probabilities can be accumulated to obtain an estimate for the transition matrix.

### Implications for Adaptive Methods

As the estimators aforementioned have different statistical properties,[127] it is useful to use all possible estimators on a given data

set. However, the result will depend more on whether all important configurations have been sampled than on the estimator chosen. In the approaches that involve importance sampling discussed in "Ways to Sample the Combined Hamiltonian" Section the sampling is intertwined with the estimation of the free energy. In these methods, a biasing potential energy term or force is built up iteratively to obtain the desired probability distribution along $\lambda$. Once all important configurations have been sampled, the estimates of the free energy using any of the estimators presented should not differ significantly. However, in the early iterations while determining the ideal bias, where sampling is still incomplete, particular estimators may outperform others. Smith and Bruce[51] have shown that the estimation of biasing weights based on transition probabilities outperforms estimation based on the number of visits to particular macrostates in the early iterations. Similarly, algorithms such as ABF[109–111] locally estimate the biasing force based on the configurations sampled. The biasing force can be adapted continuously. Important sampling methods that aim to construct an energy bias need to estimate the probability distribution function along $\lambda$. Unlike a force, a probability cannot be estimated from a single configuration. To estimate a probability, global information, i.e., the whole probability distribution function along $\lambda$ is needed.[128] This distribution function can only be estimated once a given number of configurations has been sampled. This implies that the bias cannot be built continuously and that in principle configurations that were sampled using a bias from previous iterations cannot be used to estimate a new bias from the current iteration step. Methods such as local elevation[129] or metadynamics[116,130] circumvent this problem by continuously building a bias that is based on the number of visits of a certain macrostate. To this end, a history-dependent bias is constructed by adding small, repulsive Gaussian functions to drive the system out of the range of $\lambda$ values that have already been visited.

### *Analysis of Staged Calculations*

When estimating the free energy profile along $\lambda$ from staged calculations (see "Ways to Sample the Combined Hamiltonian" Section), a variety of techniques can be used to estimate the offset needed to combine separate windows. Methods that optimally unbias and combine data obtained from multiple simulations at different states are the weighted histogram analysis method (WHAM),[131,132] umbrella integration,[133,134] and the recently introduced generalization of the Bennett estimator to multiple states M-BAR.[135]

## Conclusion

We have presented and discussed the three basic components of a free energy calculation: The choice of a suited model Hamiltonian, alternative sampling protocols which allow to generate a representative ensemble of configurations, and alternative estimators for the difference in free energy. The aim was to enable the reader to classify methods described in the literature by identifying which choices have been made for these three basic ingredients. Citations to the literature were chosen to exemplify particular alternatives. They do by no means present an exhaustive list of work that has been done in this field.[136]

## References

1. Ytreberg, F. M.; Zuckerman, D. M. J Chem Phys 2006, 124, 104105.
2. Tyka, M. D.; Sessions, R. B.; Clarke, A. R. J Phys Chem B 2007, 111, 9571.
3. Jarzynski, C. Phys Rev E 1997, 56, 5018.
4. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. Biomolecular Simulation: The GROMOS96 Manual and User Guide, Vdf Hochschulverlag AG an der ETH Zürich: Zürich, 1996.
5. Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. J Comput Chem 2004, 25, 1656.
6. Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. J Chem Phys 1995, 102, 5451.
7. Hünenberger, P. H.; van Gunsteren, W. F. In van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J. Eds.; Computer Simulation of Biomolecular Systems: Theoretical and Experimental Application; Vol. 3, Kluwer Academic Publishers: Dordrecht, 1997; pp. 3–82.
8. Ponder, J. W.; Case, D. A. Adv Protein Chem, 2003, 66, 27.
9. Weiner, P. K.; Kollman, P. A. J Comput Chem 1981, 2, 287.
10. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. J Comput Chem 2005, 26, 1668.
11. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
12. Foloppe, N.; MacKerell, A. D. J Comput Chem 2000, 21, 86.
13. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J Chem Phys 1983, 79, 926.
14. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. J Am Chem Soc 1996, 118, 11225.
15. Mahoney, M. W.; Jorgensen, W. L. J Chem Phys 2000, 112, 8910.
16. Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. J Phys Chem B 2004, 108, 16264.
17. Jorgensen, W. L.; Tirado-Rives, J. Proc Natl Acad Sci USA 2005, 102, 6665.
18. van Gunsteren, W. F.; Daura, X.; Mark, A. E. In Schleyer, R. Ed.; Encyclopaedia of Computational Chemistry, Vol. 2; John Wiley & Sons: New York, 1998: pp. 1211–1216.
19. Daura, X.; Mark, A. E.; van Gunsteren, W. F. J Comput Chem 1998, 19, 535.
20. Schuler, L. D.; Daura, X.; van Gunsteren, W. F. J Comput Chem 2001, 22, 1205.
21. Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. J Chem Phys 2003, 119, 5740.
22. Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. J Chem Theory Comput 2009, 5, 350.
23. Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. J Mol Graphics Modell 2006, 25, 247.
24. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. J Comput Chem 2004, 25, 1157.
25. Shing, K. S.; Gubbins, K. E. Mol Phys 1982, 46, 1109.
26. Powles, J. G.; Evans, W. A. B.; Quirke, N. Mol Phys 1982, 46, 1347.
27. Jacucci, G.; Quirke, N. Lect Notes Phys 1982, 166, 38.
28. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. J Chem Phys 1953, 21, 1087.

29. Frenkel, D.; Smit, B. Understanding Molecular Simulation (Computational Science Series), Vol. 1; Academic Press: New York, 2001; ISBN 0122673514.

30. Alder, B. J.; Wainwright, T. E. J Chem Phys 1957, 27, 1208.

31. Allen, M. P.; Tildesley, D. J. Computer Simulation of Liquids; Oxford University Press: New York, 1987.

32. van Gunsteren, W. F.; Berendsen, H. J. C. Mol Phys 1977, 34, 1311.

33. Cuendet, M. A.; van Gunsteren, W. F. J Chem Phys 2007, 127, 184102.

34. Hünenberger, P. Adv Polym Sci 2005, 173, 105.

35. Mezei, M. J Comput Phys 1987, 68, 237.

36. Bennett, C. H. J Comput Phys 1975, 19, 267.

37. Jacucci, G.; Rahman, A. Report on Workshop Methods in Molecular Dynamics: Long Timescale events; C. E. C. A. M.: Orsay, 1974; pp. 32–40.

38. Pomes, R.; McCammon, J. A. Chem Phys Lett 1990, 166, 425.

39. Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. J Comput Chem 1999, 20, 786.

40. Andricioaei, I.; Dinner, A. R.; Karplus, M. J Chem Phys 2003, 118, 1074.

41. Duane, S.; Kennedy, A. D.; Pendleton, B. J.; Roweth, D. Phys Lett B 1987, 195, 216.

42. van Gunsteren, W. F.; Berendsen, H. J. C. Angew Chem Int Ed Eng 1990, 29, 992.

43. Tuckerman, M.; Berne, B. J.; Martyna, G. J. J Chem Phys 1992, 97, 1990.

44. Schlick, T. Structure 2001, 9, R45.

45. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Science 1983, 220, 671.

46. Kaptein, R.; Zuiderweg, E. R. P.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. F. J Mol Biol 1985, 182, 179.

47. Bruccoleri, R. E.; Karplus, M. Biopolymers 1990, 29, 1847.

48. Auffinger, P.; Wipff, G. J Comput Chem 1990, 11, 19.

49. Berg, B. A.; Neuhaus, T. Phys Lett B 1991, 267, 249.

50. Berg, B. A.; Neuhaus, T. Phys Rev Lett 1992, 68, 9.

51. Smith, G. R.; Bruce, A. D. J Phys A: Math Gen 1995, 28, 6623.

52. Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. J Chem Phys 1992, 96, 1776.

53. Wang, F. G.; Landau, D. P. Phys Rev Lett 2001, 86, 2050.

54. Gao, Y. Q. J Chem Phys 2008, 128, 064105.

55. Kou, S. C.; Zhou, Q.; Wong, W. H. Ann Statist 2006, 34, 1581.

56. Tsallis, C. J Stat Phys 1988, 52, 479.

57. Plastino, A. R.; Anteneodo, C. Ann Phys 1997, 255, 250.

58. Barth, E. J.; Laird, B. B.; Leimkuhler, B. J. J Chem Phys 2003, 118, 5759.

59. Voter, A. F. Phys Rev Lett 1997, 78, 3908.

60. Voter, A. F. J Chem Phys 1997, 106, 4665.

61. Steiner, M. M.; Genilloud, P. A.; Wilkins, J. W. Phys Rev B 1998, 57, 10236.

62. Pal, S.; Fichthorn, K. A. Chem Eng J 1999, 74, 77.

63. Hamelberg, D.; Mongan, J.; McCammon, J. A. J Chem Phys 2004, 120, 11919.

64. Rahman, J. A.; Tully, J. C. Chem Phys 2002, 285, 277.

65. Gao, Y. Q.; Yang, L. J. J Chem Phys 2006, 125, 114103.

66. Huber, T.; Torda, A. E.; van Gunsteren, W. F. J Phys Chem A 1997, 101, 5926.

67. Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Chem Phys Lett 1994, 222, 529.

68. Pillardy, J.; Piela, L. J Phys Chem 1995, 99, 11805.

69. Shao, C. S.; Byrd, R.; Eskow, E.; Schnabel, R. B. J Global Optim 2000, 16, 167.

70. van Schaik, R. C.; Berendsen, H. J. C.; Torda, A. E.; van Gunsteren, W. F. J Mol Biol 1993, 234, 751.

71. Beutler, T. C.; van Gunsteren, W. F. J Chem Phys 1994, 101, 1417.

72. Wu, X. W.; Wang, S. M. J Phys Chem B 1998, 102, 7238.

73. Wu, X. W.; Wang, S. M. J Chem Phys 1999, 110, 9401.

74. Shinoda, W.; Mikami, M. Chem Phys Lett 2001, 335, 265.

75. van Schaik, R. C.; van Gunsteren, W. F.; Berendsen, H. J. C. J Comput-Aided Mol Des 1992, 6, 97.

76. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. J Comput Phys 1977, 23, 327.

77. Marrink, S. J.; de Vries, A. H.; Mark, A. E. J Phys Chem B 2004, 108, 750.

78. Voth, G. A. Ed. Coarse-Graining of Condensed Phase and Biomolecular Systems; CRC Press/Taylor and Francis Group, Boca Raton, FL, 2009; ISBN 1420059556.

79. Winger, M.; Trzesniak, D.; Baron, R.; van Gunsteren, W. F. Phys Chem Chem Phys 2009, 11, 1934.

80. van Gunsteren, W. F.; Luque, F. J.; Timms, D.; Torda, A. E. Annu Rev Biophys Biomol Struct 1994, 23, 847.

81. Roux, B.; Simonson, T. Biophys Chem 1999, 78, 1.

82. Swendsen, R. H.; Wang, J. S. Phys Rev Lett 1986, 57, 2607.

83. Hukushima, K.; Nemoto, K. J Phys Soc Jpn 1996, 65, 1604.

84. Sugita, Y.; Okamoto, Y. Chem Phys Lett 1999, 314, 141.

85. Sugita, Y.; Kitao, A.; Okamoto, Y. J Chem Phys 2000, 113, 6042.

86. Fukunishi, H.; Watanabe, O.; Takada, S. J Chem Phys 2002, 116, 9058.

87. Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. J Chem Theory Comput 2006, 2, 217.

88. Hritz, J.; Oostenbrink, C. J Chem Phys 2008, 128, 144121.

89. Lwin, T. Z.; Luo, R. J Chem Phys 2005, 123, 194904.

90. Christen, M.; van Gunsteren, W. F. J Chem Phys 2006, 124, 154106.

91. Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. Phys Rev Lett 2006, 96, 028105.

92. Lyman, E.; Zuckerman, D. M. J Chem Theory Comput 2006, 2, 656.

93. Squire, D. R.; Hoover, W. G. J Chem Phys 1969, 50, 701.

94. Mark, A. E.; van Gunsteren, W. F.; Berendsen, H. J. C. J Chem Phys 1991, 94, 3808.

95. Pitera, J. W.; Van Gunsteren, W. F. Mol Simul 2002, 28, 45.

96. Steinbrecher, T.; Mobley, D. L.; Case, D. A. J Chem Phys 2007, 127, 214108.

97. Best, R. B.; Chen, Y. G.; Hummer, G. Structure 2005, 13, 1755.

98. Han, K. K. Phys Lett A 1992, 165, 28.

99. Han, K. K. Phys Rev E 1996, 54, 6906.

100. Chen, Y. G.; Hummer, G. J Am Chem Soc 2007, 129, 2414.

101. Christ, C. D.; van Gunsteren, W. F. J Chem Phys 2007, 126, 184110.

102. Christ, C. D.; van Gunsteren, W. F. J Chem Phys 2008, 128, 174112.

103. Valleau, J. P.; Card, D. N. J Chem Phys 1972, 57, 5457.

104. Torrie, G. M.; Valleau, J. P. J Comput Phys 1977, 23, 187.

105. Sprik, M.; Ciccotti, G. J Chem Phys 1998, 109, 7737.

106. Woods, C. J.; Essex, J. W.; King, M. A. J Phys Chem B 2003, 107, 13703.

107. Pitera, J.; Kollman, P. J Am Chem Soc 1998, 120, 7557.

108. Kong, X. J.; Brooks, C. L. J Chem Phys 1996, 105, 2414.

109. Darve, E.; Pohorille, A. J Chem Phys 2001, 115, 9169.

110. Darve, E.; Wilson, M. A.; Pohorille, A. Mol Simul 2002, 28, 113.

111. Rodriguez-Gomez, D.; Darve, E.; Pohorille, A. J Chem Phys 2004, 120, 3563.

112. Rosso, L.; Minary, P.; Zhu, Z. W.; Tuckerman, M. E. J Chem Phys 2002, 116, 4389.

113. Zhang, Z. Y.; Shi, Y. Y.; Liu, H. Y. Biophys J 2003, 84, 3583.

114. Maragliano, L.; Vanden-Eijnden, E. Chem Phys Lett 2006, 426, 168.

115. Abrams, J. B.; Tuckerman, M. E. J Phys Chem B 2008, 112, 15742.

116. Laio, A.; Parrinello, M. Proc Natl Acad Sci USA 2002, 99, 12562.

117. Zwanzig, R. W. J Chem Phys 1954, 22, 1420.

118. Hummer, G.; Pratt, L. R.; Garcia, A. E. J Am Chem Soc 1997, 119, 8523.

119. Pohorille, A.; Darve, E. AIP Conf. Proc. Vol. 872, November 29, 2006; pp. 23–30. doi: 10.1063/1.2423257 Permalink: http://link.aip.org/link/?APCPCS/872/23/1

120. Nanda, H.; Lu, N. D.; Woolf, T. B. J Chem Phys 2005, 122, 134110.
121. Bennett, C. H. J Comput Phys 1976, 22, 245.
122. Crooks, G. E. Phys Rev E 1999, 60, 2721.
123. Crooks, G. E. Phys Rev E 2000, 61, 2361.
124. Kirkwood, J. G. J Chem Phys 1935, 3, 300.
125. Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. Chem Phys Lett 1989, 156, 472.
126. Darve, E.; Rodriguez-Gomez, D.; Pohorille, A. J Chem Phys 2008, 128, 144120.
127. Shirts, M. R.; Pande, V. S. J Chem Phys 2005, 122, 144107.
128. Darve, E. In Chipot, C.; Pohorille, A. Eds.; Free Energy Calculations: Theory and Applications in Chemistry and Biology; Springer: Berlin, 2007; Chapter 8, pp. 119–170.
129. Huber, T.; Torda, A. E.; van Gunsteren, W. F. J Comput Aided Mol Des 1994, 8, 695.
130. Laio, A.; Gervasio, F. L. Rep Progr Phys 2008, 71, 126601.
131. Ferrenberg, A. M.; Swendsen, R. H. Phys Rev Lett 1989, 63, 1195.
132. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. J Comput Chem 1992, 13, 1011.
133. Billeter, S. R.; van Gunsteren, W. F. J Phys Chem A 2000, 104, 3276.
134. Kastner, J.; Thiel, W. J Chem Phys 2005, 123, 144104.
135. Shirts, M. R.; Chodera, J. D. J Chem Phys 2008, 129, 124105.
136. Hansen, H.; Hünenberger, P. J Comput Chem doi: 10.1002/jcc. 21253.