

Good Practices in Free-Energy Calculations

Andrew Pohorille,^{†,§} Christopher Jarzynski,[‡] and Christophe Chipot^{*,†,||}

NASA Ames Research Center, Exobiology Branch, Mail Stop 239-4, Moffett Field, California, 94035-1000, Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland, 20742-4454, and Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Engineering, University of Illinois at Urbana-Champaign, 405 North Mathews, Urbana, Illinois 61801

Received: April 2, 2010; Revised Manuscript Received: June 5, 2010

As access to computational resources continues to increase, free-energy calculations have emerged as a powerful tool that can play a predictive role in a wide range of research areas. Yet, the reliability of these calculations can often be improved significantly if a number of precepts, or good practices, are followed. Although the theory upon which these good practices rely has largely been known for many years, it is often overlooked or simply ignored. In other cases, the theoretical developments are too recent for their potential to be fully grasped and merged into popular platforms for the computation of free-energy differences. In this contribution, the current best practices for carrying out free-energy calculations using free energy perturbation and nonequilibrium work methods are discussed, demonstrating that at little to no additional cost, free-energy estimates could be markedly improved and bounded by meaningful error estimates. Monitoring the probability distributions that underlie the transformation between the states of interest, performing the calculation bidirectionally, stratifying the reaction pathway, and choosing the most appropriate paradigms and algorithms for transforming between states offer significant gains in both accuracy and precision.

Introduction

The role of molecular-level, free-energy calculations in basic theoretical research on chemical and biological systems is firmly established. This is because free energies, or chemical potentials, are the central quantities that determine the behavior of these systems at or near equilibrium. They govern, for example, ligand binding to macromolecules, partitioning of drugs and small molecules across cell membranes, conformational changes of proteins required for many cellular functions, and association of molecules forming biological or nanotechnological structures. This means that understanding and reliable predictions of how these processes take place without knowledge of the associated free-energy changes is quite difficult if not impossible. Furthermore, free-energy calculations often provide an efficient route to estimating kinetic and dynamic characteristics of chemical and biological processes, such as rate constants, permeability coefficients, or ionic conductance.

Progress in the theory of free-energy calculations, combined with advances in other, related areas of statistical mechanics, such as enhanced sampling techniques, as well as the steady increases in computational capabilities, have brought the determination of free-energy changes from numerical simulations to the level of increasingly reliable, efficient, and well-characterized research tools. As a result, these tools are being used with increasing frequency in several important areas of biotechnology, nanotechnology, and pharmacology, such as

computer-aided drug design or protein engineering. One avenue to improving this methodology is to implement in standard calculations the various theoretical advances that are already at hand. Some of these advances were proposed quite some time ago, but are still not fully appreciated or employed sufficiently broadly. Others are relatively recent. Their common feature is that they are simple to use and provide a better framework for understanding free-energy calculations while reducing the associated errors. For this reason, we call them “good practices”. The main goal of this contribution is to review these good practices and explain why they should be broadly applied. Failing to follow these practices can lead to unreliable or inefficient free-energy estimates. It should be pointed out, however, that following good practices does not guarantee satisfactory agreement between calculated free energies and physical reality. Other problems, which are beyond the scope of this paper and are due, for example, to inaccurate potential energy functions, improper or inaccurate simulation algorithms, or difficulties in adequate sampling of phase space caused by slow time evolution of relevant degrees of freedom might also contribute importantly to errors associated with the computed free energies. Thus, good practices should be viewed as a necessary, rather than a sufficient condition for obtaining free-energy estimates both reliably and efficiently.

Among several general-purpose approaches to calculating free energies,¹ we focus on free energy perturbation (FEP) and nonequilibrium work (NEW) methods. FEP is one of the oldest, well-established, and most frequently used methods. It was developed for statistical mechanics of condensed phases by Robert Zwanzig in 1954,² but its simple derivation can be found in the earlier, classical statistical mechanics textbook by Lev Landau.³ In contrast, NEW is only 13 years old,^{4,5} and its potential is still being explored. Even though these two methods have very different histories, they share theoretical underpinnings. In fact, since the inception of NEW, it has been appreciated that FEP can be

* To whom correspondence should be addressed. E-mail: chipot@ks.uiuc.edu.

[†] NASA Ames Research Center.

[‡] University of Maryland.

[§] University of Illinois at Urbana-Champaign.

^{||} Also at University of California San Francisco, San Francisco, California, 94143-2280.

[¶] On leave from Équipe de dynamique des assemblages membranaires, UMR 7565, Nancy Université, BP 239, 54506 Vandoeuvre-lès-nancy cedex, France.



Andrew Pohorille was born in Szczecin, Poland in 1949. He received his M.Sc. and Ph.D. in physics and biophysics from University of Warsaw and did his postdoctoral training at the Institut de Biologie Physico-Chimique in Paris under Prof. Bernard Pullman. Currently, he directs the NASA Center for Computational Astrobiology at NASA Ames Research Center with a joint appointment in the Department of Pharmaceutical Chemistry at University of California, San Francisco. In addition to computer simulations of chemical and biological systems, he has conducted research and published in other areas, ranging from the structure of comets to the mechanism of anesthesia and risky decision making.

considered as a limiting case of this method.⁴ For this reason, it is convenient to consider them together.

Other general-purpose methods are also available for free energy calculations. They are based either on probability distributions and histograms or on thermodynamic integration using constrained or unconstrained dynamics.¹ These methods have an equally broad range of applications and for many problems are the methods of choice. Although a number of recommendations of good practices formulated here would also hold for these methods, they are generally and in many respects different from those relevant to FEP and NEW. For this reason, they require separate treatments beyond the scope of the present contribution.

In the next section, we present the theoretical background for FEP and NEW. Then, we address the issue of both statistical and systematic errors encountered in these methods. We start with a somewhat qualitative discussion of the reasons for the frequently observed poor convergence of these calculations, followed by a more formal error analysis and the presentation of several techniques aimed at controlling and reducing errors. In particular, we discuss how introducing intermediate states in a transformation between two states of interest (stratification) and carrying out these transformations in both directions impacts errors. Each subsection ends with a summary of recommended practices relevant to a given topic. The subsequent section is devoted to good practices in what is probably the most popular application of FEP: alchemical transformations. The paper closes with a summary of good practices recommended for both FEP and NEW calculations and suggestions for future research directions that might further improve the reliability and efficiency of free-energy estimates.

Theoretical Background

Although there are various ways to formulate the problem of free energy estimation, perhaps the simplest begins with a parameter-dependent Hamiltonian

$$\mathbf{H}(\Gamma, \lambda) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{x}_1, \dots, \mathbf{x}_N; \lambda) \quad (1)$$



Christopher Jarzynski was born in Washington, DC in 1965. After receiving an A.B. in Physics from Princeton University in 1987, he spent a year at the University of Warsaw, Poland on a Fulbright Fellowship. In 1994, he obtained a Ph.D. in Physics from the University of California, Berkeley, under the supervision of Włodzimierz Świątecki and Robert Littlejohn. After a postdoctoral position at the Institute for Nuclear Theory in Seattle (1994–96), he spent 10 years at Los Alamos National Laboratory, first as a Director-Funded Postdoctoral Fellow, then as a staff member. Since 2006, he has held a joint position at the Department of Chemistry and Biochemistry and the Institute for Physical Science and Technology at the University of Maryland, College Park, where he is currently a professor.



Chris Chipot was born in Paris, France in 1967. In 1990, he received a M.Sc. in Physical Chemistry from the University of Nancy. In 1994, he obtained a Ph.D. in Theoretical Chemistry from the University of Nancy under the direction of Bernard Maigret. He subsequently joined the group of Peter Kollman at the University of California, San Francisco. He then moved to the NASA Ames Research Center, where he investigated interfacial phenomena with Andrew Pohorille. In 1996, he was recruited by the Centre National de la Recherche Scientifique, and obtained his habilitation in 2000. Since 2006, he has held the position of Research Director at the University of Nancy with a joint appointment at the University of Illinois, Urbana–Champaign. His research interests range from the structure and function of membrane proteins to developments of intermolecular potentials as well as new strategies for free-energy calculations.

Here $\Gamma = (\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{p}_1, \dots, \mathbf{p}_N)$ denotes a point in the phase space of the N -particle system of interest, and λ is an external parameter or a collection of such parameters. For convenience, we will often write $U(\Gamma; \lambda)$, although it is understood that the potential energy depends only on the coordinates and not on the momenta. For a fixed value of λ , the equilibrium state of the system at temperature T is described by the Boltzmann–Gibbs distribution,

$$p^{\text{eq}}(\Gamma, \lambda) = \frac{\exp[-\beta H(\Gamma, \lambda)]}{\int d\Gamma' \exp[-\beta H(\Gamma', \lambda)]} \equiv \frac{1}{Q_\lambda} \exp[-\beta H(\Gamma, \lambda)] \quad (2)$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, and the temperature (T) dependence of the partition function Q_λ has been suppressed. The free-energy difference between two states at the parameter values $\lambda = 0, 1$ is given in terms of the ratio of the corresponding partition functions,

$$\Delta A \equiv A_1 - A_0 = -\frac{1}{\beta} \ln \frac{Q_1}{Q_0} \quad (3)$$

It is this difference that we wish to compute.

This parametric formulation of the problem encompasses a wide variety of scenarios,^{1,6} and in practice, the calculations at the heart of many free energy methods fit into this framework, regardless of whether the physical problem of interest explicitly involves an external parameter. For example, in “computer alchemy”,^{7–10} λ is a parameter used to interpolate between two different physical Hamiltonians, which effectively represent systems with different chemical compositions. In such calculations, one molecule is mutated into another (for instance, by the modification of interatomic interactions) through a sequence of intermediate states with no physical relevance. The parameter λ can also be used to specify an artificial confining potential, restricting the system to a particular region of configuration space. This approach is taken in the *weighted histogram analysis method*¹¹ and in *steered molecular dynamics*.¹² Alternatively, when determining the potential of mean force along a simple reaction coordinate such as a dihedral angle, we can treat this coordinate itself as an external parameter rather than a dynamical variable. Finally, of course, λ might represent an actual mechanical parameter or external field; for example, a magnetic field acting on a system of interacting spins.

Methods of estimating ΔA using numerical simulations are grounded in various identities of statistical mechanics. We now briefly discuss the two classes of identities that are the focus of this paper: *free energy perturbation* identities, which are formulated in terms of equilibrium averages, and *nonequilibrium work* identities, involving averages over nonequilibrium trajectories.

The starting point for FEP is the identity²

$$\exp(-\beta\Delta A) = \langle \exp(-\beta\Delta U) \rangle_0^{\text{eq}} \quad (4)$$

where $\Delta U(\Gamma) \equiv U(\Gamma, \lambda_1) - U(\Gamma, \lambda_0)$ and the angular brackets specify an equilibrium average,

$$\langle \dots \rangle_\lambda^{\text{eq}} = \int d\Gamma p^{\text{eq}}(\Gamma, \lambda) \dots \quad (5)$$

In the most direct application of this identity, the system is simulated in the equilibrium state 0 using either molecular dynamics or Monte Carlo sampling. Then the right side is evaluated by averaging directly over N sampled microstates,

$$\Delta A \approx -\frac{1}{\beta} \ln \frac{1}{N} \sum_{i=1}^N \exp[-\beta\Delta U(\Gamma_i)] \quad (6)$$

where $\Gamma_1, \dots, \Gamma_N$ are microstates generated during the simulation.

The application of eq 4 just described is *unidirectional*, from a given sampling distribution ($\lambda = 0$) to a given target distribution ($\lambda = 1$). Of course, the roles can be reversed: if

one samples from equilibrium state 1 rather than 0, then the FEP identity is

$$\exp(\beta\Delta A) = \langle \exp(\beta\Delta U) \rangle_1^{\text{eq}} \quad (7)$$

It is also possible, and in fact beneficial, to estimate free energy using *bidirectional* methods that combine samples from both distributions. Such methods can be derived from the identity,

$$\frac{P_0(\Delta U)}{P_1(\Delta U)} = \exp[\beta(\Delta U - \Delta A)] \quad (8)$$

where $P_0(\Delta U)$ is the distribution of values of ΔU obtained when sampling from the equilibrium state at $\lambda = 0$ and $P_1(\Delta U)$ is defined similarly for $\lambda = 1$. We will discuss these methods in detail in the section on combining forward and backward calculations.

In contrast to FEP, which is based on equilibrium sampling, NEW relies on simulations of a system driven away from equilibrium. The starting point is the identity⁴

$$\exp[-\beta\Delta A] = \langle \exp[-\beta W^F] \rangle_F \approx \frac{1}{N^F} \sum_{i=1}^{N^F} \exp[-\beta W_i^F] \quad (9)$$

Here, the brackets $\langle \dots \rangle_F$ denote an average over N^F trajectories generated during simulations, in which the system’s initial conditions are sampled from equilibrium state 0, then the system evolves as the parameter λ is varied from 0 to 1 using a predetermined schedule, or *protocol*, λ_t^F ($0 \leq t \leq \tau$). The superscript F specifies “forward”, to distinguish from an average over “reverse” trajectories, discussed below. W^F is the external work performed on the system during a single simulation, and the sum on the right is now a sum over N^F simulations (i.e., trajectories) rather than over sampled microstates. In this paper, we will assume that each simulation is performed using continuous-time molecular dynamics, although analogous results hold for discrete-time Monte Carlo simulations.^{5,13} The value of work is given by a path integral along the trajectory Γ_t ,

$$W^F = W^F[\Gamma_t] = \int_0^\tau dt \dot{\lambda}_t^F \frac{\partial U}{\partial \lambda}(\Gamma_t; \lambda_t^F) \quad (10)$$

where λ_t^F specifies the external protocol used to vary the work parameter from $\lambda_0^F = 0$ to $\lambda_\tau^F = 1$.

When considering NEW, it is useful to adopt the perspective that a trajectory Γ_t represents the microscopic history of the system during a thermodynamic process. This point of view is useful even when the simulations depict physically unrealizable situations, as is the case with computer alchemy. With this in mind, the inequality

$$\langle W^F \rangle_F \geq \Delta A \quad (11)$$

which is an immediate consequence of eq 9, can be interpreted as a statement of the second law of thermodynamics, in the form of the Clausius inequality.^{14,15} Although the second law places a bound on the average work performed during a thermodynamic process, the NEW identity, eq 9, places a

constraint on the statistical fluctuations around this average. It is this constraint that we exploit in order to estimate ΔA .

We can also imagine simulations in which initial conditions are sampled from equilibrium state 1, then the system evolves under the same MD scheme as when the work parameter is switched from 1 to 0, using the reverse protocol,

$$\lambda_t^R = \lambda_{\tau-t}^F \quad (12)$$

In this case, the value of ΔA can be estimated using the identity

$$\exp(\beta\Delta A) = \langle \exp(-\beta W^R) \rangle_R \quad (13)$$

where $W^R[\Gamma_t] = \int_0^\tau dt \lambda_t^R \partial_\lambda U(\Gamma_t; \lambda_t^R)$.

Moreover, as with FEP, there is a bidirectional version of NEW (see the section on bidirectional calculations), which combines data from trajectories generated using both the forward and the reverse protocols, λ_t^F and λ_t^R . The starting point is the following fluctuation theorem obtained by Crooks,^{13,16} which is the nonequilibrium analogue of eq 8:

$$\frac{P^F(+W)}{P^R(-W)} = \exp[\beta(W - \Delta A)] \quad (14)$$

Here, $P^F(W)$ is the distribution of work values obtained from trajectories generated using the forward protocol, and $P^R(W)$ is defined correspondingly for the reverse protocol. (Note the minus sign in the denominator on the left side of the equation.)

We note that nonequilibrium simulations can also be used to determine the *potential of mean force* (PMF) along a reaction coordinate, rather than a free-energy difference between two states 0 and 1.^{17–20} This topic is, however, outside the scope of the present article.

There is an evident correspondence between the FEP identities (eqs 4, 7) and the NEW identities (eqs 9, 13). The relationship can neatly be understood by considering simulations performed in the “sudden limit” of infinitely rapid switching, $\tau \rightarrow 0$. In this situation, the system has no opportunity to evolve during the process: $\Gamma_t = \Gamma_0$ for all $0 < t < \tau$. Equation 10 then reduces to $W^F = \Delta U(\Gamma_0)$, and the average over forward trajectories becomes an average over microstates Γ_0 sampled from state 0. For the reverse process, similar comments apply, but with $W^R = -\Delta U(\Gamma_0)$. These observations allow us to interpret the FEP identities as special cases of the NEW identities, emerging in the limit of sudden switching, $\tau \rightarrow 0$.

Managing Errors in FEP and NEW Calculations

The Problem of Poor Convergence. Although the FEP and NEW identities are exact, both methods suffer acutely from poor convergence. Consider a straightforward implementation of NEW involving a complex system with many degrees of freedom. Simulations are performed, values of work are computed, and the right-hand side of eq 9 is evaluated. Suppose a running estimate of the free-energy difference, $\Delta\hat{A}$, is tallied and plotted to examine the convergence of the estimate as a function of the number of simulations, N . In all likelihood, the plot will not show steadily decreasing fluctuations as the current estimate of ΔA , $\Delta\hat{A}$, approaches an asymptotic value; rather, we will see a characteristic sawtooth pattern, punctuated by sudden drops in the value of $\Delta\hat{A}$, as illustrated in Figure 1.

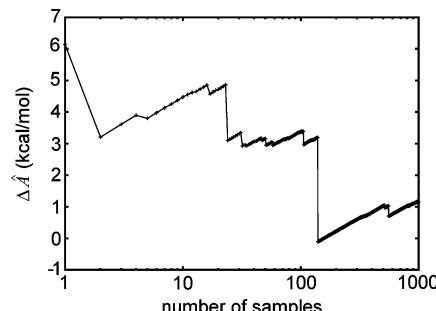


Figure 1. Schematic illustration of the sawtooth pattern that often arises when taking exponential averages. The sudden changes in $\Delta\hat{A}$ are the signature of an average dominated by rare events.

A sawtooth pattern indicates that the average is dominated by rare events: the sudden changes in $\Delta\hat{A}$ are due to trajectories with unusually low values of work. In this situation, the number of simulations required to achieve reasonable convergence might be prohibitively large. In this section, we discuss this issue in detail, focusing our attention on those “rare events” that must be sampled to achieve convergence.

In the context of FEP, convergence is closely related to the phase space overlap between the equilibrium distributions $p_0^{eq} = p^{eq}(\Gamma, 0)$ and $p_1^{eq} = p^{eq}(\Gamma, 1)$.^{21,22} Specifically, when evaluating the right side of eq 6 by simulating equilibrium state 0, the dominant contribution to the exponential average comes from those microstates that are typical of state 1. When there is little phase space overlap between the distributions p_0^{eq} and p_1^{eq} , these dominant microstates are generated very rarely, leading to slow convergence.

Since the right sides of eqs 4 and 7 involve averages over quantities that depend only on ΔU , these averages can be expressed in terms of $P_0(\Delta U)$ or $P_1(\Delta U)$ instead of p_0^{eq} or p_1^{eq} . That is,

$$\Delta A = -\frac{1}{\beta} \ln \int \exp(-\beta\Delta U) P_0(\Delta U) d\Delta U \quad (15)$$

and similarly for state 1,

$$\Delta A = \frac{1}{\beta} \ln \int \exp(\beta\Delta U) P_1(\Delta U) d\Delta U \quad (16)$$

From eqs 2 and 3, it follows that $P_0(\Delta U)$ and $P_1(\Delta U)$ are not independent, but instead, they are connected through the relation

$$\exp(-\beta\Delta U) P_0(\Delta U) = \exp(-\beta\Delta A) P_1(\Delta U) \quad (17)$$

(equivalently, eq 8). Thus, the integrand in eq 15 is proportional to the probability distribution of ΔU sampled from state 1. This reinforces the earlier observation that the integral, and therefore ΔA can be estimated reliably only if microstates representative of state 1 have been sampled from state 0.

This idea can be illustrated quantitatively if one assumes that $P_0(\Delta U)$ is Gaussian. Substituting

$$P_0(\Delta U) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\Delta U - \langle \Delta U \rangle_0)^2}{2\sigma^2}\right] \quad (18)$$

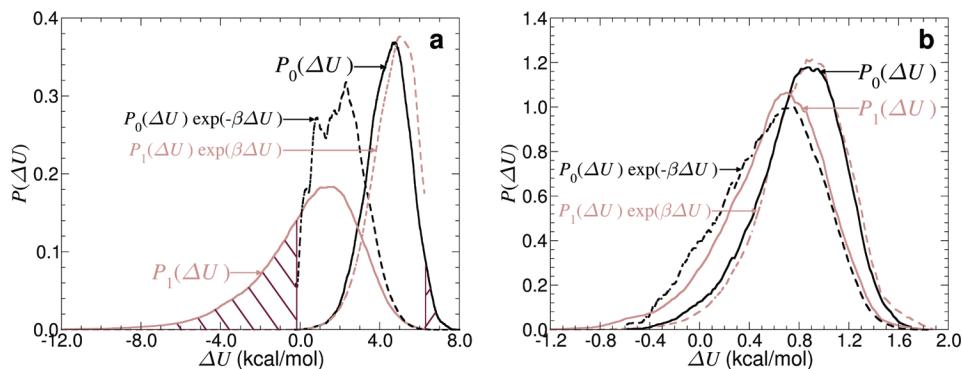


Figure 2. Histograms of probability distributions $P_0(\Delta U)$ and $P_1(\Delta U)$, and the corresponding integrands in eqs 15 and 16 obtained from FEP calculations of the hydration free energy of 4-methylimidazole (i.e., the side chain of ϵ -L-histidine) (see the Supporting Information). Forward and backward transformations correspond, respectively, to annihilation and creation of the solute. In these calculations, either an 8-window (a) or a 32-window (b) stratification strategy was employed. The probability distributions are given for a single stratum, that is, for a given value of the coupling parameter, λ . From eq 17, it follows that $P_1(\Delta U)$ should differ from $\exp(-\beta\Delta U) P_0(\Delta U)$ only by a constant factor. The same should be the case for $P_0(\Delta U)$ and $\exp(\beta\Delta U) P_0(\Delta U)$. For the 32-window strategy (b), the histograms obtained from calculations in both directions yield expected results, and therefore, the free energy estimates are expected to be reliable. For the 8-window strategy (a), the histograms are markedly broader, and the free energy estimates for annihilation are expected to be inaccurate. In this case, $\exp(-\beta\Delta U) P_0(\Delta U)$ clearly has a different shape from $P_1(\Delta U)$ because the low- ΔU region in the shaded area is almost never sampled. Stratification, creation, and annihilation are discussed in further detail in the text.

where $\sigma^2 = \langle \Delta U^2 \rangle_0 - \langle \Delta U \rangle_0^2$, we find that the integrand in eq 15 is also a Gaussian, but not normalized to unity and, compared to P_0 , shifted toward lower values of ΔU by $\beta\sigma^2$ or $\sigma/k_B T$ standard deviations. The larger this shift, the more difficult it is to sample the energies that contribute significantly to the integral. If $\sigma = k_B T$ approximately one in six sampled values of ΔU will be on the left (low ΔU) side of $\exp(-\beta\Delta U) P_0(\Delta U)$. For $\sigma = 2k_B T$, this fraction is reduced to 1/40. If σ is further increased to $5 k_B T$, the fraction drops to approximately 1/3 500 000, which means that, among statistically independent configurations sampled in simulations of typical length, we might never encounter a value of ΔU in this region. Thus, the sampling of the integrand in eqs 4 and 7 might be strongly biased, and convergence is poor.

A similar argument applies also to distributions that are not Gaussian. This is illustrated in Figure 2. If probability distributions are narrow, as in the right panel, free energy estimates are reliable. If they are broad, as in the left panel, the reliability suffers. More generally, the *relative entropy*, or *Kullback–Leibler divergence*,^{23,24}

$$D[p_0^{\text{eq}}||p_1^{\text{eq}}] = \int p_0^{\text{eq}} \ln \left(\frac{p_0^{\text{eq}}}{p_1^{\text{eq}}} \right) \geq 0 \quad (19)$$

provides an information-theoretic measure of the degree to which p_0^{eq} is distinguishable from p_1^{eq} . If the distributions are identical, then $D = 0$, whereas a value $D \gg 1$ indicates that there is little overlap between them. From eqs 1–3 and 17, it follows that

$$D[p_0^{\text{eq}}||p_1^{\text{eq}}] = D[P_0||P_1] = \beta(\langle \Delta U \rangle_0^{\text{eq}} - \Delta A) \quad (20)$$

where $D[P_0||P_1]$ is the relative entropy of $P_0(\Delta U)$ with respect to $P_1(\Delta U)$. Combining this result with eq 19 then gives us Gibbs' inequality: $\langle \Delta U \rangle_0^{\text{eq}} \geq \Delta A$.

It is useful to think of the difference,

$$W_d^0 \equiv \langle \Delta U \rangle_0^{\text{eq}} - \Delta A \quad (21)$$

as the amount of work that is dissipated during an instantaneous switching process from 0 to 1. For sampling from state 1, we similarly introduce

$$W_d^1 \equiv -\langle \Delta U \rangle_1^{\text{eq}} + \Delta A \quad (22)$$

and we have $D[P_1||P_0] = \beta W_d^1 \geq 0$, the analogue of eq 20. These quantities provide rough estimates of the number of samples needed for FEP to converge, when sampling from either state 0 or state 1,^{25–28}

$$N_0 \sim \exp(\beta W_d^1) \quad \text{and} \quad N_1 \sim \exp(\beta W_d^0) \quad (23)$$

There is a duality at play here (the convergence when sampling from 0 is dictated by the dissipation when sampling from 1, and vice versa) which has an interesting, if somewhat counterintuitive, consequence often observed in practice. Namely, if we compare the performance of eq 4 (i.e. sampling from 0) with that of eq 7 (i.e. sampling from 1), then we find that better convergence is achieved for the case of greater dissipation. A familiar example is Widom's particle insertion method.²⁹ Estimating the excess chemical potential by inserting a particle into a fluid works far better in practice than particle deletion,⁶ even though insertion is much more dissipative than deletion when the fluid is dense.

Similar considerations apply to NEW, only here, we are dealing with ensembles of trajectories (F/R) rather than canonical distributions in phase space (0/1). We distinguish between the typical trajectories obtained when simulating the forward process and the dominant trajectories that contribute the most to the average appearing in eq 9. As described more precisely elsewhere, the dominant trajectories are exactly those realizations in which time appears to be running backward.²⁷ Although this statement provides some intuition regarding the convergence of NEW, in practice, it is more useful to compare work distributions. If we construct histograms representing $P^F(W)$ and $P^R(-W)$ (the reader should note the minus sign here) then the degree of overlap between these histograms provides an indication of the likely convergence of eq 9, just as the overlap between $P_0(\Delta U)$ and $P_1(\Delta U)$ indicates the convergence of eq

4. This follows from Crooks's fluctuation theorem (eq 14). As before, the amount of overlap can be quantified using relative entropy,

$$D[P^F||P^R] = \beta(\langle W \rangle_F - \Delta A) \equiv \beta W_d^F \geq 0 \quad (24)$$

When the dissipated work, W_d^F , is much greater than $k_B T$, then there is little overlap between the work distributions, and one must resort either to stratification or to longer switching simulations to reduce dissipation. Just like with FEP, the rough number of trajectories required for convergence of NEW is governed by a duality between the forward and reverse processes

$$N^F \sim \exp(\beta W_d^R) \quad \text{and} \quad N^R \sim \exp(\beta W_d^F) \quad (25)$$

In summary, when applying FEP, it is good practice to assess the degree of phase space overlap between states 0 and 1. This is not as difficult as it might at first seem. The amount of overlap between the many-dimensional distributions of microstates in states 0 and 1 can be measured as the amount of overlap between the one-dimensional distributions $P_0(\Delta U)$ and $P_1(\Delta U)$; that is, the distributions of values of ΔU obtained when sampling from the two states. Thus, when implementing FEP, it is useful to sample from both states 0 and 1 and to construct histograms of the sampled values of ΔU . Examples of such histograms are shown in Figure 2. In many cases, one will see immediately that the overlap between the histograms is poor, indicating that the phase space overlap between the corresponding multidimensional distributions is also poor. This is illustrated in the left panel of Figure 2. In such situations, the direct application of FEP to these two states will not provide meaningful results, and stratification, which is discussed hereafter, will be most likely required.

Error Analysis. The previous section highlighted an important point: Due to poor overlap, free-energy calculations can be burdened with substantial errors, which ought to be controlled and estimated. Supplying an error estimate for the computed ΔA , however, has been frequently overlooked or simply ignored. Yet, without any measure of error, free-energy calculations are of limited utility; it is unclear how to interpret comparisons between calculated and experimental free-energy differences or how to improve the calculated estimates if they are not satisfactory. In recent years, considerable progress has been made in understanding the nature and sources of errors involved in calculating ΔA .^{26,30–38} These studies revealed that the problem is complex and reliable error estimates are often difficult to achieve. Even though a complete and fully satisfying solution to this problem is still not available, it is helpful to understand the reasons for this state of affairs and to be aware of the tools that can be used to estimate and manage errors. We will first consider unidirectional calculations (see eqs 4 and 7) and then discuss how to combine them to reduce error.

The first step toward calculating ΔA and the associated error is to define an estimator $\hat{\Delta A}$, a function that provides a prescription for estimating ΔA from a sample of N independent values of a random variable, ΔU , generated in computer simulations. Then the mean-squared error of $\hat{\Delta A}$, $\delta^2 \varepsilon_{\hat{\Delta A}} = \mathbb{E}[(\hat{\Delta A} - \Delta A)^2]$, where $\mathbb{E}[X]$ denotes the expected value of X , can be represented as a sum of two terms,

$$\begin{aligned} \delta^2 \varepsilon_{\hat{\Delta A}} &= \mathbb{E}[\hat{\Delta A}^2] + \Delta A^2 - 2\mathbb{E}[\hat{\Delta A}]\Delta A \\ &= (\mathbb{E}[\hat{\Delta A}^2] - \mathbb{E}[\hat{\Delta A}]^2) + \\ &\quad (\mathbb{E}[\hat{\Delta A}]^2 + \Delta A^2 - 2\mathbb{E}[\hat{\Delta A}]\Delta A) \\ &= \sigma_{\hat{\Delta A}}^2 + b^2(\hat{\Delta A}) \end{aligned} \quad (26)$$

The first term on the right-hand side of this equation, $\sigma_{\hat{\Delta A}}^2$, is the variance of the sample, which is a measure of the statistical error. The second term, $b^2(\hat{\Delta A}) = (\mathbb{E}[\hat{\Delta A}] - \Delta A)^2$, is the square of the bias of the estimator. The bias is a difference between the expected value of the estimator and the true value of the quantity being estimated. It reflects systematic errors associated with a given estimator. If an estimator is unbiased, $b(\hat{\Delta A}) = 0$, and the mean-squared error is equal to the variance. Conversely, if an estimator is biased, then $\hat{\Delta A}$ is not expected to be equal to ΔA , even when samples of size N are drawn many times. In common terminology,^{1,39} the variance measures the *precision* of the calculated free energy, the bias measures its *accuracy*, and the mean-squared error measures the overall reliability of the computation. Note that both $\delta^2 \varepsilon_{\hat{\Delta A}}$ and $\sigma_{\hat{\Delta A}}^2$ depend not only on a sample but also on an estimator. For the same sample of ΔU , different estimators will yield, in general, different values of ΔA . In fact, identifying efficient estimators, which minimize the mean square error (or the variance), will be one of the main themes in the discussion that follows.

The main problem with estimating errors in free energy calculations is that ΔA is a nonlinear function of ΔU . Then, a simple relation for linear functions, f , of a random variable X , $\mathbb{E}[f(X)] = f(\mathbb{E}[X])$, no longer holds. In nonlinear cases, estimators of $f(X)$ are often biased, and exact calculation of the variance is not a simple matter. To deal with these difficulties, the *delta method*^{40–42} is usually employed. In this method, the expectation of $f(X)$ is approximated by the expectation of a polynomial approximation to $f(X)$. This polynomial is generally a truncated Taylor expansion centered at the mean of X , although other polynomials also can be used. Most error analyses of ΔA ^{26,30,34,35,37,43} are based on this approach, even if this is not explicitly stated. The main theorem underlying the delta method⁴⁴ states in a slightly simplified form that for a function $f(X_N)$ of a sample X_N consisting of N independent values of a random variable, the expectation value of $f(X_N)$ can be written as

$$\mathbb{E}[f(X_N)] = f(\bar{X}) + \sum_{j=1}^n \frac{f^{(j)}(\bar{X})}{j!} \mathbb{E}(X_N - \bar{X})^j + \mathcal{O}(N^{-(n+1)/2}) \quad (27)$$

providing that f is bounded, and the first $n + 1$ derivatives exist and are also bounded. Here, \bar{X} is the true value of the average X and $f^{(j)}(\bar{X})$ denotes the j th derivative of f at \bar{X} .

Equation 27 is an expansion of $\mathbb{E}[f]$ in the powers of N . If the series is truncated after the first order,

$$\mathbb{E}[f(X_N)] - f(\bar{X}) = \left. \frac{\partial f}{\partial X} \right|_{X=\bar{X}} \mathbb{E}(X_N - \bar{X}) \quad (28)$$

it follows immediately that

$$\sigma_{f(X)}^2 = \left[\left. \frac{\partial f}{\partial X} \right|_{X=\bar{X}} \right]^2 \sigma^2(X) \quad (29)$$

This equation is frequently used to estimate the variance of nonlinear functions of random variables. It is often called the error propagation formula, although it should be kept in mind that this is just the first-order approximation, which is satisfactory if the estimator is asymptotically normal or if the variance is small. Otherwise, accurate estimates of the variance might require higher-order terms.

Probably the most commonly used estimator in FEP calculations is

$$\Delta\hat{A}_N^{\text{FEP}} = -\frac{1}{\beta} \ln \left(\frac{1}{N} \sum_{i=1}^N \exp[-\beta\Delta U_i] \right) \quad (30)$$

and the corresponding estimator in NEW is

$$\Delta\hat{A}_N^{\text{NEW}} = -\frac{1}{\beta} \ln \left(\frac{1}{N} \sum_{i=1}^N \exp[-\beta W_i] \right) \quad (31)$$

where ΔU_i and W_i are the energy difference and the work for sample i . These two estimators, which have been already introduced in eqs 6 and 9 are asymptotically unbiased, that is, they converge to ΔA as $N \rightarrow \infty$. For a finite N , however, they remain biased. As has been illustrated in the previous section (see Figure 2), low- ΔU or low- W tails are undersampled, which produces systematic errors. Thus, both the variance and the bias have to be accounted for in error estimation. Other estimators have also been proposed.²⁶ They also can be mapped into expansions using the delta method. At this point, it is not clear that they have any advantage over $\Delta\hat{A}_N^{\text{FEP}}$ or $\Delta\hat{A}_N^{\text{NEW}}$ in general cases.

The variance of the estimator can be calculated using the first-order delta method, eq 29. If a sample of N_0 values of ΔU has been generated from the ensemble 0, then the variance of the mean ΔA , $\sigma_{\Delta A}^2$, can be approximated as

$$\begin{aligned} \sigma_{\Delta A}^2 &= \frac{1}{N_0 \beta^2} \frac{1}{\langle \exp(-\beta\Delta U) \rangle_0^2} \times \\ &\quad [\langle \exp(-2\beta\Delta U) \rangle_0 - \langle \exp(-\beta\Delta U) \rangle_0^2] \\ &= \frac{1}{N_0 \beta^2} \frac{\langle \exp(-2\beta\Delta U) \rangle_0}{\langle \exp(-\beta\Delta U) \rangle_0^2} - \frac{1}{N_0 \beta^2} \end{aligned} \quad (32)$$

The bias of the estimators in eqs 30 and 31 can be obtained by comparing the average of M separate evaluations of ΔA , each obtained from N sampled values of ΔU or W , with ΔA estimated from a single calculation, in which the sample size is MN . Since the estimators are asymptotically unbiased, the difference between these two estimates in the limit of $M \rightarrow \infty$ is equal to the bias. Wood et al.³⁰ derived the leading-term approximation to this bias. This result was subsequently extended and generalized by Zuckerman and Woolf³³ in terms of the estimated first three moments of $\exp(-\beta\Delta U)$ or $\exp(-\beta W)$: $\hat{\sigma}$, $\hat{\mu}$, or $\hat{\mu}_3$,

$$E[\Delta\hat{A}_N] - \Delta A = \frac{\phi_1}{\beta N} + \frac{\phi_2}{\beta N^2} + \mathcal{O}\left(\frac{1}{N^3}\right) \quad (33)$$

where

$$\begin{cases} \phi_1 &= \frac{\hat{\sigma}^2}{2\hat{\mu}^2} \\ \phi_2 &= -\frac{4\hat{\mu}\hat{\mu}_3 - 9\hat{\sigma}^4}{12\hat{\mu}^4} \end{cases} \quad (34)$$

The first term can be considered as a special case of the second-order delta method, eq 27. It has been shown that this result applies to a broad class of functions f , which need not be bounded.⁴⁵ Limited numerical tests^{30,46} indicate that the first-order bias correction appears to be reliable only when the variance is small (i.e. is on the order of $k_B T$) and $P(\Delta U)$ is not very far from a Gaussian. The range of applicability of the higher order approximation that involves also ϕ_2 has not yet been well established.

There are also other sources of systematic errors that lead to differences between the calculated and the actual values of ΔA . The most common, and potentially the most serious ones, are associated with inadequate sampling of ensemble configurations due to the existence of long-lived, metastable states, or slow degrees of freedom, as well as inaccuracies in the potential functions. Other sources of systematic errors include persistent long-time correlations, integration errors in molecular dynamics, finite system size effects, and other artifacts of computer simulations.⁴⁷ Even though they are clearly important, their assessment properly belongs to other areas of theory and, therefore, fall beyond the scope of this paper.

So far, it has been assumed that a sample of N independent measurements is available. However, configurations generated in computer simulations are typically correlated. If this is the case, a number of results presented above do not hold unless the sample is uncorrelated or correlations are properly accounted for. Depending on how the sample has been generated, a number of techniques exist for this purpose.^{6,48,49} One of them is to use averages from blocks of sampled configurations sufficiently large to ensure almost complete decay of correlations between the averages in consecutive blocks. If a sample has been generated from a time sequence (e.g. from a molecular dynamics trajectory), then the average number of independent configurations in a sample of N_0 configurations generated every Δt is $N_0/(2\tau/\Delta t + 1)$, where τ is the correlation time.⁵⁰ The variance from eq 32 should then be divided by this factor instead of N_0 .

In summary, it is strongly recommended to provide error estimates, which are an integral and essential part of free-energy calculations. In FEP and NEW methods, both the variance and the bias are, in general, expected to contribute to the error. The variance can be approximated using the first-order delta method, as was done in eq 32. Estimating the bias is usually more complicated. At first sight, it might appear that for large sample sizes, systematic errors should not be important because the leading term in the bias decreases with N as $1/N$, whereas the standard deviation (i.e. the square root of the variance, which is a measure of statistical errors) decays only as $1/\sqrt{N}$. This reasoning, however, is correct only if the variance is small or the distribution is reliably Gaussian. Otherwise, the bias might be much larger than the variance and higher terms in the approximation, eq 27, are no longer negligible. Including them, however, might not be very helpful because higher Boltzmann moments are difficult to evaluate accurately and the convergence of the expansion is not guaranteed. Taken together, there are good reasons to keep the variance small, on the order of $1-2k_B T$. In this regime, it is expected that for simulations of typical lengths, the variance obtained from eq 32 should be fairly

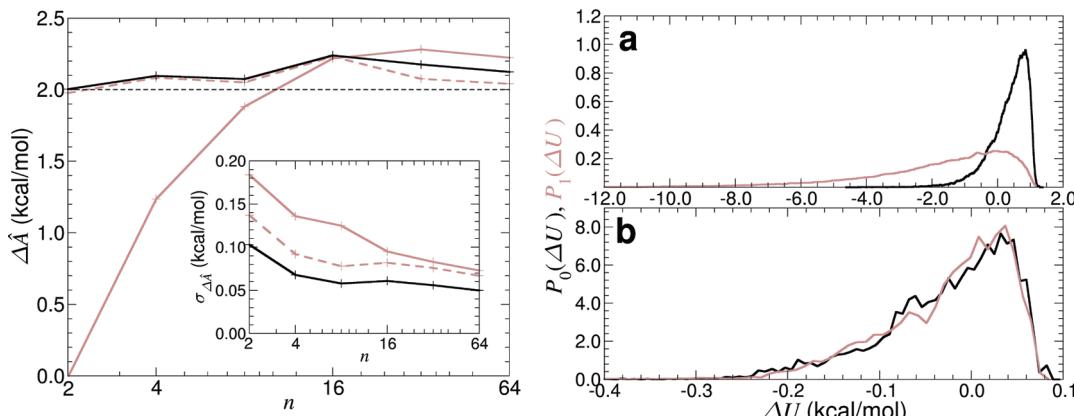


Figure 3. The influence of stratification on the free-energy estimates and the associated statistical errors is illustrated in the case of the hydrophobic hydration of argon. Shown in the figure are the results of the reversible annihilation (see the section on alchemical transformations) of a single argon atom in water using between 2 and 64 intermediate stages but maintaining the overall length of the simulation constant (the details of the simulations are provided in the Supporting Information). Left: The net free-energy differences as a function of the number of stages for the forward, annihilation and the backward, creation, transformations are shown as light, solid, and dashed lines, respectively. The BAR estimate is shown as a dark, solid line. The target, experimental hydration free energy of +2.00 kcal/mol⁵² is displayed as a dashed horizontal line. Inset: Standard deviation associated with the free-energy estimates obtained from the different stratification strategies. Line coding is identical as above. The statistical error appears to level off as the number of strata increases, but the reader is reminded that because fewer samples are considered, determination of accurate correlation lengths becomes more difficult, hence, casting doubt on the estimates of the standard deviation. Right: Underlying probability distributions for a single stratum in the forward, $P_0(\Delta U)$ (dark lines); and backward, $P_1(\Delta U)$ (light lines) transformations based on a 4- (a) and a 64-window (b) free-energy calculation. Note that as the number of strata increases from 4 to 64, the probability distributions become significantly narrower.

accurate, and the bias should be small. It has been suggested^{33,34} that the latter could be tested by plotting the estimate of ΔA , $\Delta \hat{A}_n$, for different sample sizes, n , as a function $1/n$, with $n \leq N$. It is expected that this relation will be linear for a sufficiently small variance. How to manage the variance is the subject addressed in the next section.

Stratification. Probably the simplest, universally applicable strategy for reducing systematic errors is stratification, sometimes also called multistage sampling.⁵¹ In stratified calculations, one defines a sequence of intermediate Hamiltonians corresponding to values of λ between 0 and 1. The unidirectional or bidirectional strategies are used to estimate free energy differences between adjacent pairs of equilibrium states in this sequence (see Figure 3), which are then summed to give the total value ΔA . This strategy essentially breaks the problem into manageable pieces and allows for the sampling of all relevant regions of phase space, even those associated with very low-probability measures in the $\lambda = 0$ and $\lambda = 1$ states. Formally, for FEP calculations, stratification relies on the identity

$$\Delta A = -\frac{1}{\beta} \ln \left(\frac{Q_1}{Q_0} \right) = -\frac{1}{\beta} \ln \left(\frac{Q_M Q_1}{Q_0 Q_M} \right) = \Delta A_{0,M} + \Delta A_{M,1} \quad (35)$$

where M is an intermediate ensemble between 0 and 1. $\Delta A_{0,M}$ and $\Delta A_{M,1}$ are the free-energy differences between macroscopic states 0 and M , and M and 1, respectively. Clearly, eq 35 can be generalized to n strata by applying the identity serially,

$$\Delta A = -\frac{1}{\beta} \sum_{i=0}^{n-1} \ln \langle \exp(-\beta \Delta U_i) \rangle_i \quad (36)$$

Here, for convenience of notation, the final state is denoted by the index n rather than 1; intermediate states are denoted by the index i , $i = 1, \dots, n-1$; and $\Delta U_i = U(\Gamma, \lambda_{i+1}) - U(\Gamma, \lambda_i)$.

If the potential energy is represented as a linear function of a general extent, or coupling parameter, λ , where $0 \leq \lambda \leq 1$,

$$U(\lambda_i) = \lambda_i U_0 + (1 - \lambda_i) U_1 = U_0 + \lambda_i \Delta U \quad (37)$$

which is a common, although not necessarily optimal choice, then ΔA can be represented as

$$\Delta A = -\frac{1}{\beta} \sum_{i=0}^{n-1} \ln \langle \exp(-\beta \Delta \lambda_i \Delta U) \rangle_i \quad (38)$$

where $\Delta \lambda_i$ is the change of the coupling parameter between states i and $i+1$.

Stratification is almost always “good practice” because it improves overlap between the probability distributions of ΔU in two consecutive states. For example, if distributions for all n strata are Gaussian with the same variance, the Kullback–Leibler divergence (eq 20) between two consecutive states is reduced by a factor of n . This is, however, achieved at a possible cost of reducing efficiency. Striking a perfect balance between these two conflicting criteria is difficult because it would require prior knowledge of the dependence of ΔA on λ . Yet, if linear scaling of the Hamiltonian is employed, it is straightforward to manage the variance by adjusting $\Delta \lambda_i$. As can be seen from eq 38, only the total ΔU needs to be calculated in each stratum, or stage. If its sampled values have been stored, a $\Delta \lambda$ that yields the desired value of the variance can then be determined at a postprocessing stage. If ΔU is a complex function of λ , for example, as discussed later on in the section on paradigms for alchemical transformations, it might be beneficial to calculate ΔU at each step for several values of the coupling parameter, which can be done at only modest additional computational effort, and then choose $\Delta \lambda$ that yields the variance closest to the target value. In most instances, choosing equal variances in all stages is nearly optimal, even though for nonlinear functions of a sampled random variable it does not have to be strictly so. Keeping the

variance at each stage on the order of $1-2 k_B T$ usually guarantees good overlap between probability distributions in two consecutive strata and a reliable error estimate. Assuming that the values of the variance at each stage are uncorrelated, the total variance can be subsequently obtained from the Bienaymé formula for a random variable X , $\text{var}(\sum_{i=0}^{n-1} X_i) = \sum_{i=0}^{n-1} \text{var}(X_i)$. We note that since stratification reduces systematic but not statistical error, increasing the number of strata while keeping a fixed, total sample size is not useful when the former becomes smaller than the latter.

Combining Forward and Backward Calculations. If both equilibrium ensembles 0 and 1 have been sampled, it is generally advantageous to construct the corresponding free-energy difference, ΔA , by combining both data sets. To use a universal terminology for both FEP and NEW, we call calculations initiated from 0 and 1 “forward” and “backward”, respectively. The search for optimal solution to this problem can be formulated in several ways. One is to consider additional stratification, in which ΔA is obtained from the free-energy differences between each end point, 0 or 1, and an intermediate state M , so that

$$\exp(-\beta\Delta A) = \frac{Q_1}{Q_0} = \frac{Q_M}{Q_0} \left| \frac{Q_M}{Q_1} \right| = \frac{\langle \exp(-\beta(U_M - U_0)) \rangle_0}{\langle \exp(-\beta(U_M - U_1)) \rangle_1} = \frac{\langle w \exp(-\beta\Delta U/2) \rangle_0}{A_{M,0} - A_{M,1}} \quad (39)$$

Probably the simplest choice of M is such that $U_M = (U_0 + U_1)/2$. This choice has been proposed a number of times and is sometimes referred to as “half umbrella”,⁵³ or “simple overlap sampling” (SOS).³⁶ It is also closely related to the idea of “double-wide sampling”.⁷ It yields

$$\exp(-\beta\Delta A) = \frac{\langle \exp(-\beta\Delta U/2) \rangle_0}{\langle \exp(\beta\Delta U/2) \rangle_1} \quad (40)$$

The choice of M , however, is arbitrary and can be optimized subject to some criteria. To do so, one can define a function, $w(\Gamma, \lambda) = \exp\{-\beta[U_M - (U_0 + U_1)/2]\}$ and rewrite eq 39 as

$$\exp(-\beta\Delta A) = \frac{\langle w \exp(-\beta\Delta U/2) \rangle_0}{\langle w \exp(\beta\Delta U/2) \rangle_1} \quad (41)$$

Here, for simplicity, the arguments of w have been suppressed. The intermediate state M or, equivalently, the function w can then be optimized by requesting that the variance of the mean of ΔA , σ^2 , be minimized with respect to w , as described further in this section.

Another approach to optimizing bidirectional calculations is to introduce a finite function, $w(\Gamma, \lambda)$, that weights contributions from configurations sampled from ensembles 0 and 1 such that the variance of ΔA is minimized. This is the essence of the Bennett acceptance ratio (BAR) method.⁴³ In this approach, free energy can be expressed as

$$\begin{aligned} \exp(-\beta\Delta A) &= \frac{Q_1}{Q_0} \\ &= \frac{Q_1}{Q_0} \frac{\int w \exp[-\beta(U_0 + U_1)/2] d\Gamma}{\int w \exp[-\beta(U_0 + U_1)/2] d\Gamma} \\ &= \frac{Q_1}{Q_0} \frac{\int w \exp(-\beta\Delta U/2) \exp(-\beta U_0) d\Gamma}{\int w \exp(\beta\Delta U/2) \exp(-\beta U_1) d\Gamma} \\ &= \frac{\langle w \exp(-\beta\Delta U/2) \rangle_0}{\langle w \exp(\beta\Delta U/2) \rangle_1} \end{aligned} \quad (42)$$

where w is the same as in eq 41. Since eq 42 is identical to eq 41, minimizing σ^2 yields the same results in both cases.⁵⁴ In the first order approximation, eq 29, the optimal w is the hyperbolic secant function

$$w = \operatorname{sech}[\beta(\Delta U - C)/2] = \frac{2}{\exp[\beta(\Delta U - C)/2] + \exp[-\beta(\Delta U - C)/2]} \quad (43)$$

where

$$C = \Delta A + \frac{1}{\beta} \ln \frac{N_1}{N_0} \quad (44)$$

Here, N_0 and N_1 are the number of independent configurations sampled from states 0 and 1. Substitution of w from eq 43 to eq 41 or 42 yields the formula for ΔA ,

$$\exp(\beta\Delta A) = \frac{\langle f[-\beta(\Delta U - C)] \rangle_1}{\langle f[\beta(\Delta U - C)] \rangle_0} \exp(\beta C) \quad (45)$$

where

$$f(x) = \frac{1}{1 + \exp(x)} \quad (46)$$

is the Fermi function. Note that in the original formulation of BAR,⁴³ a slightly different weighting function, $w' = w e^{\beta(U_0 + U_1)/2}$, was used, but the final results remain unchanged. The quantity C that appears in these equations must be solved self-consistently: starting with an initial value for C , the estimate of ΔA obtained from eq 45 is fed into eq 44, and these steps are iterated until convergence is reached. This procedure is implemented after all the samples have been drawn and converges rapidly with insignificant computational overhead.

An alternative derivation of BAR begins with eq 8, in which the factor $\exp(-\beta\Delta A)$ can be considered as a proportionality constant connecting the distributions $P_0(\Delta U)$ and $P_1(\Delta U)$. By applying the maximum likelihood (ML) method to determine the value of ΔA that is most consistent with the sampled values of ΔU , given that $P_0(\Delta U)$ and $P_1(\Delta U)$ must satisfy eq 8, one again recovers eqs 43–46.^{28,35}

Although the methods for combining forward and backward calculations were originally developed for equilibrium sampling, they can be also applied in NEW. From Crooks's fluctuation theorem, eq 14, one can derive the result^{55,56}

$$\Delta A = \frac{1}{\beta} \ln \frac{\langle f[\beta(W^R + C)] \rangle_R}{\langle f[\beta(W^F - C)] \rangle_F} + C \quad (47)$$

As with eq 45, for N^F forward and N^R reverse trajectories, this estimate is optimized when $C = \Delta A + (1/\beta) \ln(N^R/N^F)$.

Since both optimizing the intermediate state and the ML approach are equivalent to BAR, all three approaches to combining forward and backward calculations lead to the same estimator of ΔA . Since eq 45 is an implicit equation for ΔA , that is, this quantity appears on both sides of the equation, the estimator, $\hat{\Delta A}^{\text{BAR}}$ is also implicit,

$$\left\{ \begin{array}{l} \hat{\Delta A}^{\text{BAR}} = \frac{1}{\beta} \ln \frac{\sum_{i=1}^{N_1} \frac{1}{1 + \exp[-\beta(\Delta U_i^1 - C)]}}{\sum_{j=1}^{N_0} \frac{1}{1 + \exp[-\beta(\Delta U_j^0 - C)]}} \\ \quad + C - \frac{1}{\beta} \ln \frac{N_1}{N_0} \\ C = \hat{\Delta A}^{\text{BAR}} + \frac{1}{\beta} \ln \frac{N_1}{N_0} \end{array} \right. \quad (48)$$

where ΔU_j^0 and ΔU_i^1 are the energy differences for configurations sampled from states 0 and 1, respectively. The same estimator can be applied in NEW by substituting $\Delta U_i^1 \rightarrow -W_i^R$, $\Delta U_j^0 \rightarrow W_j^F$, $N_1 \rightarrow N^R$, $N_0 \rightarrow N^F$.

If the values of ΔU or W sampled during both forward and backward simulations have been stored, these two coupled equations can be solved iteratively at a postprocessing stage with only very small, additional computational effort. The solution is always unique. Once it becomes available, it is straightforward to calculate the variance in the first-order approximation to the logarithmic function. Several equivalent formulas for the variance have been given.^{28,35,43} A simple one for numerical applications is⁴³

$$\sigma_{\Delta A}^{\text{2BAR}} = \frac{1}{N_0 \beta^2} \left[\frac{\langle f^2(x) \rangle_0}{\langle f(x) \rangle_0^2} - 1 \right] + \frac{1}{N_1 \beta^2} \left[\frac{\langle f^2(-x) \rangle_1}{\langle f(-x) \rangle_1^2} - 1 \right] \quad (49)$$

where function $f(x)$ is defined in eq 46 and $x = \beta(\Delta U - C)$. If configurations generated in states 0 and 1 are correlated, then N_0 and N_1 should be corrected for the correlation lengths in the forward and backward transformations, as discussed in the section on error estimation.

In addition to calculating variance, one can also construct somewhat conservative error bounds on the estimate of ΔA obtained with BAR.⁴³ To do so, we return to eq 48, but before optimizing C . For the optimized estimate of $\hat{\Delta A}^{\text{BAR}}$, the numerator and denominator on the right-hand side of this equation must be equal. We abbreviate the corresponding value of C as C_x . If C_1 and C_0 are defined as the values at which the numerator and the denominator are, respectively, equal to unity,

$$\sum_{i=1}^{N_1} \frac{1}{1 + \exp[-\beta(\Delta U_i^1 - C_1)]} = \sum_{j=1}^{N_0} \frac{1}{1 + \exp[\beta(\Delta U_j^0 - C_0)]} = 1 \quad (50)$$

then $\hat{\Delta A}(C_0)$ and $\hat{\Delta A}(C_1)$ provide lower and upper error bounds, respectively, on $\hat{\Delta A}^{\text{BAR}}$,

$$\hat{\Delta A}(C_0) \leq \hat{\Delta A}(C_x) \leq \hat{\Delta A}(C_1) \quad (51)$$

as discussed in greater detail by Bennett.⁴³ The same method can be applied to simulated nonequilibrium trajectories, provided that we substitute ΔU_i^0 with W_i^F and $-\Delta U_i^1$ with W_i^R in eq 51.

From the fact that $\hat{\Delta A}^{\text{BAR}}$ is a maximum likelihood estimator, it follows that it is asymptotically unbiased, that is, its bias converges to zero as the sample size increases to infinity. It is also asymptotically efficient, which means that no other asymptotically unbiased estimator has lower asymptotic mean-squared error.³⁵ It is, however, not clear what these formally satisfying properties tell us about the behavior of this estimator for finite sample sizes typical of those obtained in free-energy calculations because its convergence properties have not been well characterized.

The functional form of w underscores again the importance of sampling in the overlap region between $P_0(\Delta U)$ and $P_1(\Delta U)$. Hyperbolic secant w is a positive definite, symmetric function that reaches the maximum, $\text{sech}(0) = 1$, when $C = \Delta U$. For $N_1 = N_0$, it corresponds to the value of ΔU for which $P_0(\Delta U) = P_1(\Delta U)$. This is the point of maximum overlap. The weight of samples with values of ΔU away from this point in either direction progressively decreases. Thus, good sampling in the overlap region is essential to obtain a reliable $\hat{\Delta A}^{\text{BAR}}$.^{28,43} This is different from the condition for a good estimate of ΔA from unidirectional calculations, which is more demanding: it is required that sampling from state 0 yields good statistics in the region of ΔU , in which $P_1(\Delta U)$ peaks. This region is located farther in the low- ΔU tail of $P_0(\Delta U)$.

The variance of ΔA depends on the choice N_0 and N_1 . It is, therefore, natural to ask: What is the best division of computer resources between forward and backward calculations? Such analysis is not simple and complicates even further if multistaged stratification is carried out, as often happens in practice. With the exception of the end points, configurations sampled at each stage are used twice: to calculate ΔU in the forward and backward directions. In multistage calculations, this leads to a chain of coupled equations that would need to be minimized simultaneously to determine the optimal number of steps for a given total precision. These equations involve the values of ΔA at each stage, which are initially unknown, but might be approximated after an initial phase of calculations.⁵⁷ In addition, correlation times for forward and backward calculations could be different. Considering these complicating factors, such an approach to determining N_0 and N_1 does not seem practical. Instead, Bennett suggested that keeping N_0 and N_1 equal should be close to optimal.⁴³ More generally, it is recommended that all variances be monitored and the number of steps at each stage be adjusted to keep them balanced. Hahn and Then²⁸ have recently proposed a specific dynamic strategy that begins with Bennett's choice ($N_0 = N_1$) and then adjusts the ratio on the fly.

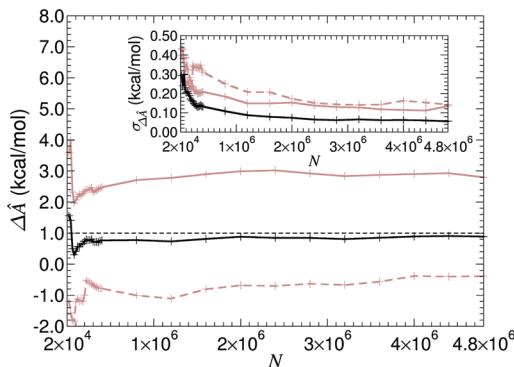


Figure 4. L-Serine (Ser) to L-alanine (Ala) point mutation in the hydrated, terminally blocked, tripeptide Ala-Ser-Ala carried out bidirectionally with a stratification strategy consisting of two windows. Here, the free-energy change corresponding to the stratum $\lambda = 0\text{--}0.5$ is reported as a function of N , the number of samples, for the forward (light, solid line) and the backward (light, dashed line) transformations and the combination thereof, using the BAR scheme (dark, solid line). The total simulation time for this calculation was equal to 12.8 ns. A separate, stratified simulation based on 16 windows gives a reproducible BAR estimate for the $\lambda = 0\text{--}0.5$ transition equal to ~ 1 kcal/mol; that is, the dashed, black line of the graph. Inset: Statistical error shown as a function of the number of samples, N , using the same line coding as described above.

If optimization of w is interpreted as the search for the best possible intermediate state, we observe that knowledge of M is not necessary during the calculations of ΔA . However, once an estimate of ΔA has been obtained, the potential energy of the system in state M can be calculated as

$$U_M = \frac{1}{\beta} \ln \{\exp[\beta(U_1 - C/2)] + \exp[\beta(U_0 + C/2)]\} - \frac{1}{\beta} \ln 2 \quad (52)$$

From this equation, it is apparent that choosing $U_M = (U_0 + U_1)/2$, as is done in SOS, or as any other linear combination of U_0 and U_1 , is not optimal in bidirectional calculations. Even though SOS, eq 40, has been reported to perform well,³⁶ there is no compelling reason to use it instead of BAR. There is even less justification to use simple or weighted averages of the free energies obtained from the forward and backward calculations,

$$\Delta A = \frac{\eta}{\beta} \ln \langle \exp(\beta \Delta U) \rangle_1 - \frac{(1-\eta)}{\beta} \ln \langle \exp(-\beta \Delta U) \rangle_0 \quad (53)$$

where $0 < \eta < 1$ might be, in general, a function of N_0 and N_1 . Since the reliabilities of the contributing free energies are usually not equal, this type of averaging might cause the free-energy estimate to deteriorate rather than improve, as compared with a properly chosen unidirectional estimate.³¹

These considerations raise a question: Is BAR more efficient than unidirectional calculations, that is, $\eta = 0$, $N_1 = 0$ for forward simulations? In other words, given a fixed amount of computer time, will BAR produce more reliable estimates of ΔA than unidirectional calculations? Numerical examples indicate that this is, indeed, the case,^{36,37} as illustrated from two different perspectives in Figure 4 for the point mutation of a single amino acid and in Figure 5 for a force-field-independent, zero-sum ethane-to-ethane alchemical transformation (details of the simulations are given in the Supporting Information). This

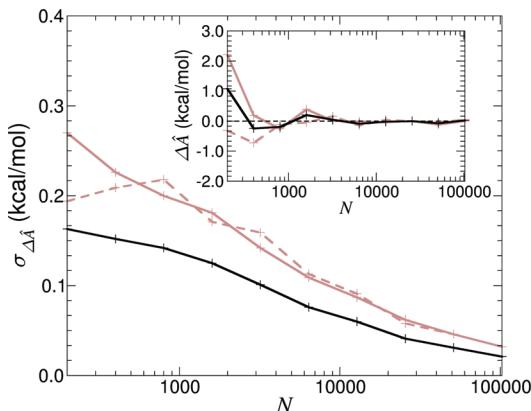


Figure 5. Zero-sum ethane-to-ethane transformation, in which the terminal methyl group of a pseudopropane molecule vanishes as the other terminal group moiety is grown.⁵⁸ In the limit of ergodic sampling, the expected free-energy change is, therefore, zero. The statistical error is shown as a function of the number of samples, n , for the forward (light, solid line) and the backward (light, dashed line) transformations, and the combination thereof, using the BAR scheme (dark, solid line). Each point corresponds to an individual free-energy calculation relying on a stratification strategy of 16 windows with an increasing sampling. Inset: Free-energy change as a function of the number of samples.

observation can be justified for most cases, in which multistaged free-energy calculations are carried out. In most instances of practical interest, only a small fraction of interactions in the system are directly affected by changing the alchemical parameter, λ .

For example, in *in silico* point mutation studies of proteins, only the part of the Hamiltonian that involves atoms in a single amino-acid side chain is modified. This implies that for sufficiently large N_0 and N_1 , most of the computational effort is expended on generating configurations from the equilibrium ensemble at each stage, and by comparison, the computer time spent on calculating ΔU in the forward or backward directions is small. Let us assume that the forward calculations produce a smaller variance of ΔA than the backward calculations and consider the intermediate-state view of BAR. Since the choice of state M is optimized to minimize the variance, it follows that any other choice of the intermediate state will produce larger σ^2 . This would, in particular, apply to a state that approaches state 1. In this case, the variance of the free-energy difference between states 0 and M approaches the variance in forward calculations, while the variance of the free-energy difference between states 1 and M approaches zero. This means that the total, optimized variance in BAR is smaller than that associated with unidirectional calculations. The computational effort incurred in forward calculations is the same in both methods for the same sample size. The additional computer time required to carry out calculations in the backward direction when using BAR is usually negligible because configurations sampled from state 1 have to be generated anyway for the next stage of multistaged calculations. Thus, BAR is expected to be more precise than unidirectional simulations for a given amount of computer time. It is also expected to be more accurate because it is equivalent to introducing an additional stratification step, which reduces systematic errors, as has been discussed previously.

In summary, BAR is not only the optimal method for combining forward and backward simulations but also, in most cases, it is expected to outperform unidirectional calculations, especially if stratification is involved. In principle, it is free of a problem commonly encountered in unidirectional calculations: forward and backward transformations are not equally reliable.

For these reasons, BAR is a recommended strategy in both FEP and NEW. Statistical error of BAR calculations can be estimated from eq 49. It should be borne in mind, however, that the optimality of BAR relies on an assumption that the first-order expansion of the logarithmic function is sufficiently accurate, which is the case in the small-variance limit. More generally, BAR is based on a specific, biased estimator, and it might be possible to define other estimators that yield a smaller mean-squared error. We will return to this issue in the last section of this paper.

Overlapping Distributions: A Simple and Valuable Consistency Check. In addition to the acceptance ratio method, Bennett⁴³ proposed an alternative method involving the interpolation of histograms. This approach is less known than the acceptance ratio method. Although it is unlikely to yield estimates of ΔA directly from the data that are superior to BAR, the identity that underlies this method (eq 55 or eq 57, below) provides a useful consistency check on the data generated by either equilibrium or nonequilibrium sampling. This consistency check is both easy to implement and sensitive to errors that might otherwise remain undetected.

Consider the functions

$$\begin{cases} g_0(\Delta U) = \ln P_0(\Delta U) - \frac{1}{2}\beta\Delta U \\ g_1(\Delta U) = \ln P_1(\Delta U) + \frac{1}{2}\beta\Delta U \end{cases} \quad (54)$$

They can be determined directly from the histograms of ΔU values obtained by sampling from states 0 and 1. From eq 8, it follows that functions g_0 and g_1 differ by a constant,

$$g_1(\Delta U) - g_0(\Delta U) = \beta\Delta A \quad (55)$$

Because of this stringent condition, it is good practice to construct explicitly functions g_1 and g_0 and to plot them together with their difference (see Figure 7.2 of ref 6 for an illustration). If there is sufficient overlap between $P_0(\Delta U)$ and $P_1(\Delta U)$, then there will be a range of values of ΔU over which the difference $g_1 - g_0$ will be constant, within statistical fluctuations. Furthermore, this difference should be consistent with the free-energy estimate obtained using, for instance, the acceptance ratio method. If this behavior is not observed, then the estimate of ΔA should not be trusted. Either the amount of sampling is not sufficient given the degree of overlap or there are underlying errors in the sampling itself. In particular, if the difference $g_1 - g_0$ appears to converge to a nonconstant function of ΔU within a region of significant overlap between P_0 and P_1 , as is illustrated in Figure 6, then this should be taken as a strong indication that states 0 and 1 are not being sampled correctly. Note that choosing the same coefficients, equal to 1/2, in front of $\beta\Delta U$ on the right-hand side of eq 54 is not required. Equation 55 remains true for any coefficients η_0 and η_1 as long as $\eta_1 = 1 - \eta_0$.

For the case of nonequilibrium sampling, the forward and reverse work distributions can be used to define the functions

$$\begin{cases} g^F(W) = \ln P^F(+W) - \frac{\beta}{2}W \\ g^R(W) = \ln P^R(-W) + \frac{\beta}{2}W \end{cases} \quad (56)$$

Using eq 14, it then follows that

$$g^R(W) - g^F(W) = \beta\Delta A \quad (57)$$

By analogy with the previous paragraph, it is good practice to construct g^F and g^R using the trajectories from the forward and reverse simulations and to plot them together with their difference. If the difference does not appear as a constant function in the overlap region or if the value of this difference is inconsistent with the free-energy estimate, then something is evidently amiss.

The Gaussian Approximation. If the probability distributions of ΔU are Gaussian estimates of statistical error, simplify markedly. Although the estimator defined in eq 30 remains biased,^{59–61} a simple, unbiased estimator, $\Delta\hat{A}_0^G$, from forward calculations,

$$\left\{ \begin{array}{l} \Delta\hat{A}_0^G = \overline{\Delta U}_0 - \frac{\beta}{2}s^2 \\ \overline{\Delta U}_0 = \frac{1}{N_0} \sum_{i=1}^{N_0} \Delta U_i^0 \\ s^2 = \frac{1}{N_0} \sum_{i=1}^{N_0} (\Delta U_i^0 - \overline{\Delta U}_0)^2 \end{array} \right. \quad (58)$$

of the exact free energy,

$$\Delta A = \langle \Delta U \rangle_0 - \frac{\beta}{2}\sigma_{\Delta U}^2 \quad (59)$$

with $\sigma_{\Delta U}^2 = \langle \Delta U^2 \rangle_0 - \langle \Delta U \rangle_0^2$, exists in this case.⁵⁹

In the absence of correlations, the exact expression for the variance of $\Delta\hat{A}_0^G$ is available.⁵⁹

$$(\sigma^G)^2 = \frac{\sigma_{\Delta U}^2}{N_0} + \frac{\beta^2 \sigma_{\Delta U}^4}{2(N_0 - 1)} \quad (60)$$

A similar expression is obtained for backward calculations,

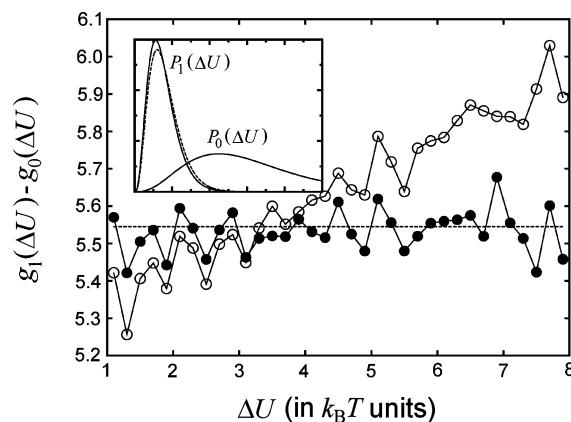


Figure 6. Overlapping distributions: Samples were taken from distributions P_0 and P_1 satisfying eq 8 (inset, solid curves), and from these data the functions g_0 and g_1 were constructed in the overlap region. Apart from statistical fluctuations, the difference $g_1 - g_0$ (solid circles) is constant, in agreement with eq 55. The exact value of $\beta\Delta A$ is shown as a horizontal dashed line. For purposes of comparison, samples were also taken from a perturbed, “erroneous” distribution P_1 (inset, dashed curve). The resulting difference $g_1 - g_0$ (open circles) clearly violates eq 55, illustrating that this test is sensitive to faulty sampling.

$$(\sigma^G)^2 = \frac{\sigma_{\Delta U}^2}{N_1} - \frac{\beta^2 \sigma_{\Delta U}^4}{2(N_1 - 1)} \quad (61)$$

from which it follows that combining forward and backward calculations reduces the variance.⁵⁹ Indeed, if $N_0 = N_1 = N/2$, then

$$(\sigma^G)^2 = \frac{\sigma_{\Delta U}^2}{N} \quad (62)$$

The same result can be obtained by combining the Gaussian estimator with SOS given in eq 40. Considering this equivalence, it is not surprising that in many instances SOS and BAR yield similar estimates of ΔA .

Should the probability distributions of W be Gaussian, an analogous estimator can be constructed in NEW by substituting W for ΔU . The variance of ΔA is then given by

$$(\sigma^G)^2 = \frac{\sigma_W^2}{N} \quad (63)$$

For long simulation times, it has been shown that $(\sigma^G)^2$ scales linearly with time⁶² and, thus, behaves like linear estimators. Moreover, it has been demonstrated theoretically and confirmed in computer simulations that $\hat{\Delta A}^G$ is always more efficient than the biased estimator given in eq 30.⁴⁶ It is, therefore, recommended that this estimator, especially in its bidirectional implementation, be employed whenever the Gaussian approximation is justified. Testing for Gaussian behavior ought to be done with great care. Even though this approximation is at the heart of several classical theories^{63–65} and forms the basis of a number of approximate theories for calculating free energies,^{66–71} it turns out that in practice, probability distributions of ΔU or W are frequently non-Gaussian. In addition, although the central regions of the work distributions $P_0(\Delta U)$ and $P_1(\Delta U)$ might be well described as Gaussians (in accordance with the central limit theorem), this approximation might fail in the tails, and precisely these tails provide the dominant contributions to ΔA . Thus, it is prudent to construct both $P_0(\Delta U)$ and $P_1(\Delta U)$ (or $P^F(W)$ and $P^R(-W)$) to determine that they have the same variance, as required for Gaussian distributions. If this is not the case, the validity of the Gaussian approximation becomes questionable. Substantial deviation from Gaussian behavior causes the performance of the estimator $\hat{\Delta A}^G$ to deteriorate markedly,⁴⁶ and as a consequence, its use in such instances is not recommended.

Alchemical Transformations

From a practical perspective, most free-energy calculations relying upon perturbation theory can be categorized into three classes of transformations: (i) creations and annihilations, (ii) point mutations, and (iii) navigation along a geometrical reaction coordinate. The latter category, which encompasses potential-of-mean-force computations, is usually handled using alternative methods¹ that might be better suited for this purpose than FEP or NEW.^{46,72} The remaining three types of system modifications (creations, annihilations and point mutations) are usually called alchemical transformations, even though, strictly speaking, only the third type involves computational change of one chemical species into another one. Creations and annihilations are standard numerical schemes employed to measure solvation free energies,

which are here understood broadly, subsuming the particular case of host:guest binding free energies, where, for instance, the host is a protein and the guest is a ligand. Both terms should not be taken literally; effectively, they mean that atoms, molecular fragments, or whole molecules are coupled to or decoupled from the rest of the system.

Since computational alchemy involves transformations through nonphysical states of the system, it is prone to undesired artifacts associated with these transformations. Without reviewing this approach in detail, we briefly summarize a few good practices, aimed at controlling, reducing, or eliminating these artifacts and, by doing so, improving the convergence and efficiency of alchemical transformations. These practices should be used in combination with the methods discussed in the previous sections. The focus is entirely on FEP, even though the same considerations apply to NEW. This is because most computational alchemy has been carried out so far using the free energy perturbation approach. A number of good practices (for example proper treatment of bond and angular constraints and the use of soft potentials) also applied to other types of free energy calculations, such as thermodynamic integration.

Thermodynamic Cycles. Computational transformations of creation, annihilation, and point mutation share a common conceptual basis. They rely on constructing an appropriate thermodynamic cycle, a series of reversible transformations connecting a reference and a target state of interest. Since free energy is a state function, values of ΔA associated with each transformation in the cycle add up to zero. Thus, it is possible to replace direct calculation of the free energy difference in the transformation of interest, which might be computationally difficult, with estimating free energy differences associated with the remaining transformations in the cycle, presumably simpler to carry out.

This is illustrated in Figure 7. The cycle in the left panel of this figure provides a prescription for calculating solvation free energy of a single species at infinite dilution. It is often referred to as absolute free energy,^{73–80} although in actuality it is the difference between free energies of a solute in a solvent and in the gas phase. The cycle in the right panel illustrates a procedure for calculating relative free energy⁷ of a solute with respect to another one. The need for such calculations arises, for example, in ranking a series of ligands according to their affinity toward a given protein. Calculations of relative free energies are expected to converge much faster because the perturbation required to transmute one molecule into another one is, in general, markedly smaller than that involved in the creation or annihilation of the same molecules.^{8,9}

If feasible, it is valuable to close the thermodynamic cycle by carrying out direct calculations of $\Delta A_{\text{solvation}}$ (along the horizontal transformations in Figure 7). This can be done using, for example, a method based on probability distributions or thermodynamic integration.¹ Since each method has different sources of errors, closing the cycle to within its statistical error markedly increases our confidence that free energy calculations were carried out reliably (although the results might still be burdened with errors due to inaccurate potential functions). Conversely, if the sum of all free energy contributions along the cycle differs markedly from zero, it most likely indicates that something is wrong with free energy calculation for at least one leg of the cycle or with sampling configurations for FEP.

Paradigms for Alchemical Transformations. Practical implementation of the thermodynamic cycles outlined above requires a numerical scheme to transform between the reference and the target states of a chemical system. This is done using

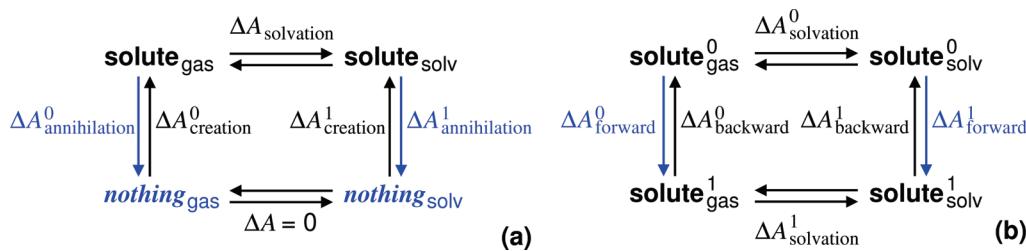


Figure 7. Examples of thermodynamic cycles used to estimate solvation free energies, $\Delta A_{\text{solvation}}$. In cycles a and b, the solvation process is described by the upper, horizontal leg, which corresponds to the transfer of the solute from the gas phase to the bulk solvent. The solvation free energy (a) can be measured by coupling the solute to its environment, that is, $\Delta A_{\text{solvation}} = \Delta A_{\text{creation}}^1 - \Delta A_{\text{creation}}^0$. This transformation is mirrored in the vertical legs of cycle a. Symmetrically, double annihilation of the solute would yield the negative of the solvation free energy. The vocabulary *nothing* utilized in the thermodynamic cycle implies that the perturbed topology is fully decoupled from the rest of the molecular system; that is, both intra- and intermolecular interactions are turned off. Relative solvation free energies of two solutes can be determined by transforming one into another in both the gas phase and in solution. This is depicted in the vertical legs of cycle b; that is, $\Delta A_{\text{solvation}}^1 - \Delta A_{\text{solvation}}^0 = \Delta A_{\text{forward}}^1 - \Delta A_{\text{forward}}^0$. To improve the free-energy estimates obtained following the thermodynamic cycles, it is recommended that the transformations be performed bidirectionally, namely, forward and backward, or creation and annihilation, and their results be combined using the relevant algorithm, for example, BAR.⁴³

either a single-topology or double-topology paradigm.⁸¹ The single-topology method was first implemented 25 years ago by William Jorgensen, who carried out the first alchemical transformation of practical interest to estimate the differential hydration free energy of ethane and methanol.⁷ Within this paradigm, a common topology for both chemical structures serves as the reference state. van der Waals parameters and charges associated with the mutation of atoms are modified from those in state 0 to those in state 1 by varying λ from 0 to 1. The missing atoms in the target structure are annihilated by progressively setting their nonbonded parameters to zero. In the mutation of ethane to methanol, in which the O–H group was represented as a united atom, the former served as the common topology. As one carbon atom was transformed to oxygen, two hydrogen atoms of the methyl group were decoupled from the rest of the system and turned into noninteracting ghost particles. An opposite approach is also possible: starting from methanol one could transform an oxygen atom to carbon and grow, or create, two hydrogen atoms. In some instances, it is advantageous to define a new common topology, M , which need not correspond to any chemical structure that exists in nature, and transform it to both 0 and 1. This will yield free-energy differences $\Delta A_{M,0}$ and $\Delta A_{M,1}$, from which the free-energy difference of interest, ΔA , can be obtained as $\Delta A_{M,1} - \Delta A_{M,0}$.

The numerical trick used in single-topology paradigm poses two obvious difficulties. When the van der Waals radii of vanishing atoms become small as λ approaches 0 or 1, these atoms can come quite close to other particles in the system. Since the vanishing atoms still carry residual charges, electrostatic interactions could increase dramatically creating large forces in the system and instabilities in integrating the equations of motion. For this reason, it might be advantageous to decouple electrostatic and nonelectrostatic transformations. Computational effort associated with this procedure is smaller than it appears because the variance of $P(\Delta U)$ for a given change of λ is reduced if each of these transformations is carried out separately, as compared with the variance associated with the coupled transformation. In addition, electrostatic interactions are usually very well behaved once decoupled from other transformations so that the corresponding $P(\Delta U)$ becomes nearly Gaussian. In fact, this observation forms one of the basis of the modern quasi-chemical theory.⁷¹ Taken together, this allows fewer intermediate λ -steps to be considered in each type of transformation.^{72,82–84}

Another difficulty is associated with the modification of chemical bonds or, more generally, bonded parameters involving the transformed atoms, as the lengths and force constants may

change during the transformation. As can be seen from Figure 7, the resulting contributions to ΔA involving mutating atoms nearly cancel out, provided that the bonds are not strongly deformed in the bound state. Hence, potentially, they could be ignored.⁸⁵ It is, however, recommended that this contribution be estimated either approximately from the averages of the appropriate energy terms or, if needed, more accurately through potential-of-mean-force-like calculations⁵⁸ with the relevant mass-metric correction.^{86–88} Special care should be taken to account properly for bonds involving appearing or disappearing atoms.⁸⁷ The same considerations apply to planar angles.

The alternative dual-topology paradigm⁸⁹ does not suffer from the above shortcomings. In this algorithm, the topologies reflecting the reference and the target states of the transformation coexist, albeit never interact, either directly, through nonbonded forces, or indirectly, through common bonded terms. In contrast to the single-topology paradigm, all interactions, with a possible exception of the bonded terms unique to each topology, are scaled as a function of λ . In the simplest case, although not optimal over the full range of λ , the potential energy at intermediate state λ can be represented as $U(\lambda) = \lambda U_1 + (1 - \lambda) U_0$. This approach, however, has its own drawbacks, which primarily stem from the possibility of the incoming topology to collide with atoms from the remainder of the chemical systems, thereby creating very large forces that might cause instabilities in the integration of the equations of motion, as well as widening the probability distribution $P(\Delta U)$. This occurs most frequently at the tail ends of the transformation. The simplest remedy consists of breaking down the reaction pathway into a number of intermediate states using $\delta\lambda$ of uneven width, which becomes increasingly smaller as λ tends toward 0 or 1. Unfortunately, even if intermediate states are spaced very closely, as close as 10^{-6} , the problem of spurious collisions involving growing or disappearing particles is not completely circumvented.

An alternative and more satisfying approach is to avoid singularities that might arise when interatomic distances, r_{ij} , approach zero during particle creation or annihilation at the end points of the reaction pathway by scaling and shifting the Lennard-Jones potential. This idea has been implemented in two popular numerical schemes. Assume that λ varies between 0 and 1 such that atom i disappears for $\lambda = 1$. One route to rewrite the conventional Lennard-Jones potential is⁹⁰

$$U^{\text{vdW}}(r_{ij};\lambda) = 4\varepsilon_{ij}(1-\lambda)\left[\left(\frac{\sigma_{ij}^2}{r_{ij}^2 + \alpha\lambda}\right)^6 - \left(\frac{\sigma_{ij}^2}{r_{ij}^2 + \alpha\lambda}\right)^3\right] \quad (64)$$

where σ_{ij} is the distance at which the van der Waals energy between atoms i and j is zero, ε_{ij} is the depth of the energy well between these atoms, and $\alpha > 0$ is a shifting parameter, which ensures that the singularity is removed, changing the conventional Lennard-Jones potential to a soft-core potential.

Another scheme is to modify the original Lennard-Jones potential as follows:⁹¹

$$U^{\text{vdW}}(r_{ij};\lambda) = 4\varepsilon_{ij}(1-\lambda)^n \left\{ \frac{1}{\left[\alpha\lambda^2 + \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right]^2} - \frac{1}{\alpha\lambda^2 + \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6} \right\} \quad (65)$$

Here again, the $\alpha\lambda^2$ term appearing in the denominator reshapes the Lennard-Jones potential into a soft-core form, devoid of singularity.

In both equations the original Lennard-Jones potential is fully recovered at $\lambda = 0$. At $\lambda = 1$, both the energy and the forces are finite and well-behaved, even if r_{ij} approaches zero. This markedly improves stability of molecular dynamics simulations. Thus, using soft-core potentials is strongly recommended. They can be applied either throughout the transformation from the initial to the final state or, after appropriate redefinition of states 0 and 1, only at both ends of this transformation. Note that once a soft-core potential is employed, ΔU is no longer linear in λ , and the strategy of choosing the stratification step in postprocessing that follows from eq 38 no longer applies. Instead, the values of ΔU in the forward and backward directions have to be calculated during the simulation rather than in postprocessing.

Under certain circumstances, concomitant scaling of Coulomb and van der Waals interactions may lead to a situation in which the energy and forces are dominated by the former interactions before the repulsive component of the latter is large enough to avoid spurious collisions of the constituent atoms. As has been already mentioned in the context of the single-topology paradigm, this shortcoming can be circumvented by decoupling electrostatic and van der Waals interactions. Without resorting to this somewhat extreme procedure, it is possible to scale Coulomb interactions with a sufficient delay to avoid scenarios, in which Lennard-Jones repulsion is overwhelmed by electrostatic forces. See the Supporting Information for an example of delaying strategies. Alternatively, the Coulomb potential can be modified in the spirit of eqs 64 and 65, replacing the interatomic distance, r_{ij} with a related soft-core term of the form $(\alpha\lambda^2 + r_{ij}^2)^{1/2}$.⁹¹ Either route is recommended whenever the problem arises to ensure numerical stability of the trajectories. The reader is referred to recent investigations appraising the relative merits of soft-core strategies.⁹²

Standard Binding Free Energies. Perturbation theory is a method of choice for calculating standard binding free energies, often improperly termed *absolute* binding free energies.^{1,77,79,80} This quantity, which can be measured using a variety of experimental techniques, including isothermal microcalorimetry,⁹³ electrospray ionization mass spectrometry,⁹⁴ or BIACore,⁹⁵ corresponds to the free-energy difference between two molecules of interest (e.g. a host and a guest) in their bound and free states. Recently, Deng and Roux proposed a formal framework to reconcile the computational, microscopic estimate of the binding

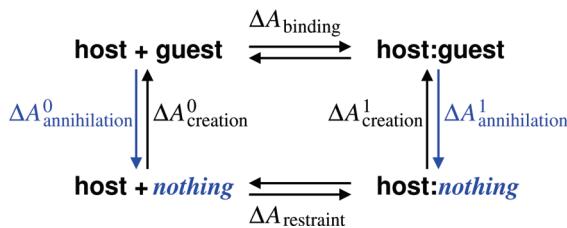


Figure 8. Thermodynamic cycle used to estimate binding free energy, $\Delta A_{\text{binding}}$. In many instances, direct, computational determination of the latter, following the upper horizontal leg of the cycle, might be difficult and computationally very intensive. Assuming that the guest binds tightly to the host, $\Delta A_{\text{binding}}$ can be inferred from a double annihilation of the former in its bound and unbound state, $\Delta A_{\text{annihilation}}^0 - \Delta A_{\text{annihilation}}^1$. As has been already mentioned, annihilation of the guest should be interpreted as its decoupling from the rest of the system. Imposition of harmonic restraints during creation transformations to prevent the growing guest from escaping the binding site of the host is tantamount to a loss of configurational entropy, which ought to be accounted for in the corresponding free-energy contribution $\Delta A_{\text{restraint}}$.

free energy with its experimental, macroscopic counterpart.⁹⁶ The main thrust of this section is to outline a series of recommendations based on the current best practices for estimating binding affinities computationally.

The underlying idea of carrying out binding free-energy calculations is to mimic the reversible association of a guest to a host, according to the thermodynamic equilibrium $\text{host} + \text{guest} \rightleftharpoons \text{host:guest}$, governed by the equilibrium constant

$$K_a = \frac{[\text{host:guest}]}{[\text{host}][\text{guest}]} \quad (66)$$

where $[\text{host:guest}]$, $[\text{host}]$ and $[\text{guest}]$ are the concentrations of the complex, the free host and the free guest, respectively. The standard binding free energy can be readily inferred from the association constant, $\Delta A_{\text{binding}}^0 = -1/\beta \ln(c^0 K_a)$, where c^0 denotes the standard concentration.

The binding free energy may be estimated computationally using the thermodynamic cycle of Figure 8, which involves double annihilation,⁹⁷ or double creation of the guest in its free and bound states. Microreversibility imposes that the forward and backward transformations yield an equal binding free-energy estimate, albeit of opposite sign. Achieving such reversibility and, thus, a correct estimate of binding free energy often requires additional computational steps.

Should the binding free energy be estimated from a double annihilation of the guest, the right leg of the thermodynamic cycle represents decoupling of the guest from the solvated host:guest complex, whereas the left leg simplifies to decoupling of the guest in a quasi-infinitely diluted solution from the rest of the system. The association constant then reduces to $K_a = V \exp[-\beta(\Delta A_{\text{annihilation}}^0 - \Delta A_{\text{annihilation}}^1)]$, under the stringent assumptions that the ligand holds steadily its initial position in the bound state and diffuses freely in the accessible volume, V , of the simulation cell in the free state. However, as the ligand is annihilated in the bound state, its affinity to the host becomes progressively weaker, so that it might move away from the binding site. The same concern is valid in the reverse, creation process. At early stages of the transformation, as electrostatic and nonelectrostatic are progressively turned on, the guest is only weakly coupled to the environment and, therefore, is generally anticipated to move away from its initial position instead of remaining steadily in the binding site of the host.

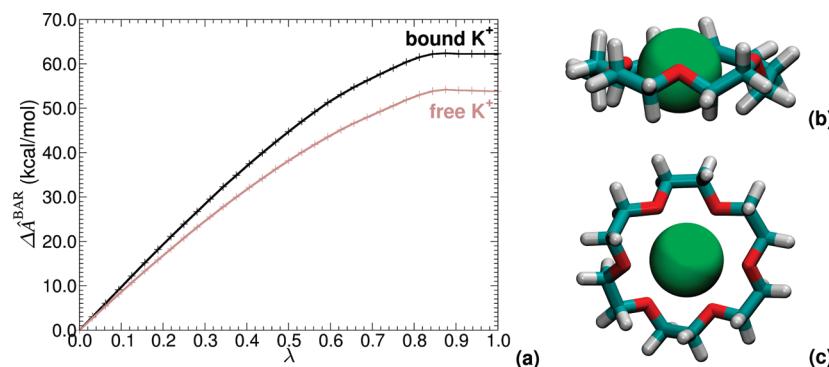


Figure 9. Binding of a potassium ion to 18-crown-6. (a) Free-energy change for the annihilation of the guest in the free state (light, solid line) and in the bound state (dark, solid line). Details of the simulations are provided in the Supporting Information. A positional restraint is enforced to prevent the alkaline cation to move astray from the center of mass of the crown ether when the two species are weakly coupled (i.e. at the end of the annihilation transformation) or at the beginning of the creation one. The free-energy contribution due to this positional restraint, $-1/\beta \ln(c^0 \Delta V)$, is equal to +4.80 kcal/mol. Reversible annihilation of the potassium ion in the free and the bound states yields a free-energy change of +53.82 and +62.25 kcal/mol, respectively, on the basis of BAR estimates. Following the thermodynamic cycle of Figure 8, the overall binding free energy is equal to -3.64 ± 0.14 kcal/mol, in good agreement with the experimentally measured value of -2.91^{100} as well as the pioneering simulation of Liem Dang and Peter Kollman.¹⁰¹ Top (b) and (c) side views of typical host:guest arrangements in the course of the simulation.

A convenient approach to circumvent this difficulty consists of introducing a series of harmonic restraints aimed at maintaining the guest in the appropriate orientation in its binding site.⁹⁸ This route is sometimes referred to as double decoupling.⁹⁷ Imposing positional restraints in the free-energy calculation corresponds to a loss of translational, rotational, and possibly conformational entropy. This yields unfavorable, positive contributions to the standard binding free energy that have to be accounted for and subsequently subtracted from the calculated standard free energy of binding. They can be determined analytically and involve two terms of distinct nature.⁹⁹ The loss of translational motion imposed by the restraining potential can be inferred from the Sackur–Tetrode equation and amounts to $-1/\beta \ln(c^0 \Delta V)$, where ΔV is a volume element of the binding site effectively explored by the center of mass of the guest. On the other hand, the loss of rotational motion amounts to $-1/\beta \ln(\Delta\Theta/8\pi^2)$, where $\Delta\Theta$ is the orientational fluctuation of the guest. In contrast to the translational term, this free-energy contribution due to a loss of rotational freedom does not depend on the concentration but, primarily, on the molecular mass of the guest. The significance of accounting for positional restraints in the determination of the binding affinity is illustrated in Figure 9 in the case of a potassium ion bound to the 18-crown-6 crown ether in an aqueous environment (details of the simulations are given in the Supporting Information).

As an alternative to positional restraints, whereby the center of mass of the guest is confined in a volume element of the binding site, the guest may be anchored to the host by means of a series of harmonic potentials aimed at preserving the contacts formed between the two molecular compounds as their interaction is perturbed from the reference to the target state. One common approach to estimate the contribution due to these restraints is perturbation theory,⁷⁵ in which the force constant of the confinement potential is progressively modified from its nominal value to zero. A similar route can be followed for a flexible guest by freezing its conformation adopted in the native host:guest complex, thereby preserving the binding mode throughout the transformation. This can be achieved by restraining the relevant dihedral angles to their nominal value. A perturbative scheme, through which, for instance, the force constant of the underlying restraints is progressively zeroed out, may subsequently be employed to determine the free-energy contribution due to these additional restraints. For consistency with the remainder of the thermodynamic cycle, the free-energy

differences which have been computed on the basis of simulations performed bidirectionally, it is advisable that the contribution arising from restraints be also measured from the combination of forward and backward runs.

Although necessary to prevent the guest from escaping or isomerizing into a conformation distinct from that of the native binding mode and, by doing so, ensuring microreversibility of the transformation, positional restraints ought to be introduced in the free-energy calculations with great care, because they may constitute important contributions to standard binding affinities and the associated errors. If a large number of restraints have been applied, measuring such contributions with an appreciable accuracy might be a challenging task. Furthermore, restraints might limit access of the system to relevant parts of the configurational space, introducing unwanted quasi nonergodicity. Thus, restraints introduce an additional source of error, which should be carefully quantified.

Summary and Outlook

When conducting free-energy calculations, it is essential to follow good practices, which can be considered as the computational equivalent of proper experimental protocols. At a modest, additional computational cost, and sometimes even saving computer time, free-energy estimates can be markedly improved. Equally importantly, it is possible to assign errors to these estimates. Without this step, comparisons between calculated and measured free-energy differences are difficult to interpret, and reliance on the computed ΔA in such areas as computer-aided drug design or in silico genetic engineering becomes questionable.

A number of specific “good-practice” steps are particularly recommended. They are discussed in detail in the preceding sections and recapitulated at the end of each subsection, so here we only provide a brief summary.

(a) All free-energy calculations should be accompanied by an error estimate. Because of the nonlinear dependence of ΔA on ΔU , obtaining these estimates is not simple. Reliable estimates of both statistical and systematic errors can be made only if the variance of the probability distributions, $P_0(\Delta U)$ and $P_1(\Delta U)$ is small, typically on the order of $1-2 k_B T$. These considerations, in combination with improved overlap, dictate that it is prudent to keep the variance in this range.

(b) Since the reliability of free-energy estimates depends critically on the overlap between $P_0(\Delta U)$ and $P_1(\Delta U)$, it is highly

advisable to plot the histograms of these distributions and the integrands in eqs 15 and 16 to assess the degree of overlap. If the overlap is poor, the resulting free-energy estimate is likely to be unreliable.

(c) Stratification provides an effective, general method for reducing the variance and improving overlap at each stage. Choosing appropriate increments of λ in each stratification step allows the variance to be adjusted such that it remains close to the desired value. However, applying stratification is not a sufficient justification to reduce sample size at each stage, since all degrees of freedom, including the slowest ones, should be adequately sampled at every stage.

(d) Combining forward and backward simulations using the BAR estimator, eq 48, is strongly recommended. This approach is simple to implement as a postprocessing algorithm and will almost always improve the reliability and the efficiency of free-energy calculations, as compared with the unidirectional method.

(e) A graphical test of consistency, as described in the section on overlapping distributions, is very valuable because it offers a possibility to detect problems associated, for example, with inadequate sampling of slow degrees of freedom that are otherwise difficult to diagnose. Again, this test is quite easy to implement in postprocessing.

(f) If the probability distributions are Gaussian, then the Gaussian estimator, eq 58, is both simpler and more accurate than the estimator given in eqs 30 and 31. Combining forward and backward calculations, as described herein, is always beneficial. It is, however, crucial to establish that the distributions are Gaussian, including in the region of overlap between the two distributions. A comparison of the variances provides a simple consistency check: if the variances are not equal, then the Gaussian approximation should not be applied.

(g) For alchemical transformations, both single- and double-topology paradigms can be used. They suffer from somewhat different shortcomings, but it does not appear that, in general, one of them should clearly be preferred over the other. Considering their respective advantages and drawbacks, the application will dictate the choice of the best suited paradigm. Decoupling electrostatic and van der Waals nonbonded interactions allows for avoiding instabilities in integrating the equations of motions whenever charged atoms are created or annihilated in the system. In addition, soft-core nonbonded potentials ought to be used under those circumstances, at least when the coupling parameter, λ , approaches 0 or 1. Special care should be taken to account correctly for the change in free energy due to modifications of parameters that describe chemical structures (e.g. chemical bonds, planar angles, dihedral angles) of the molecules in the system. Similar considerations apply to artificial potentials employed to restrain ligands near the host in determining free energies of host:guest binding. Whenever feasible, it is recommended to carry out additional free-energy calculations to close the appropriate thermodynamic cycle to confirm that its total free energy is, indeed, close to zero.

Many good practices can be applied at a postprocessing stage of calculations. A well-designed analysis package integrated with the main simulation software can readily handle even a large number of ligands. This does not mean, however, that free energy calculations can or should be fully automated. Careful inspection of results is always good practice.

Our increasing ability to calculate ΔA does not imply that further theoretical work aimed at improving the efficiency and the reliability of FEP and NEW calculations is not worth pursuing. Indeed, several research directions along these lines hold potential promise.

One is the targeted free-energy perturbation,¹⁰² or its non-equilibrium counterpart.¹⁰³ The underlying idea of this approach is to map one of the end states, for example, state 0, into another state, M , that overlaps with state 1 better than 0. The free-energy calculation between M and 1 should then converge faster than calculating ΔA directly.

An interesting line of research is to identify estimators of ΔA that are more efficient than that given in eqs 6 and 9. An example is the recent extension of the BAR to include data from multiple stages or simulations.^{38,104,105}

Another approach to improving estimators of ΔA is to model the probability distributions $P_0(\Delta U)$ and $P_1(\Delta U)$ as analytical functions, or power-series expansions that depend on a small number of adjustable parameters. Thus, what is being directly estimated from the data are these parameters rather than $\exp(-\beta\Delta A)$. The underlying assumption is that they can be determined from the whole distributions with sufficient accuracy so that the model represents correctly the relevant tails. As a result, these probability distributions should be correct, even in the regions that are sampled only rarely. This would, in turn, allow for larger stratification steps without sacrificing accuracy. For unidirectional calculations, this approach can be considered as an extrapolation technique, whereas for bidirectional methods, it becomes an interpolation strategy, which is usually more accurate. The simplest example of model probability distribution is a Gaussian,¹⁰⁶ but, as has been already discussed, this approximation is often of limited accuracy. Other functional forms, more suitable in general cases, have been tried and appear to yield improved estimates of ΔA .^{107–109} The advantages and disadvantages of this potentially very fruitful approach still remain to be explored.

Another promising research direction of considerable practical interest, especially for rapid screening of ligands that might effectively bind to protein targets or for in silico genetic engineering, is to generate sample configurations from a judiciously chosen reference state that is common to a number of different ligands or amino acid side chains.^{72,82–84} Such a reference state need not be physically meaningful; it is sufficient that it contains configurations representative of each of the end point structures. Then the relative free energies of several systems can be determined from a single set of configurations at a potentially considerable computational gain.

Since the ultimate goal of molecular-level computer simulations is to describe physical, chemical, or biological processes and phenomena, methods for calculating free energies cannot be considered separate from other topics that are beyond the scope of this paper, but might influence considerably the quality of the results obtained from FEP or NEW. One is efficient sampling of representative microstates from the underlying ensembles. This could be impeded, among others, by long-lived metastable states, or the existence of slowly equilibrating degrees of freedom in the system. This is frequently the case, especially in biological systems, and leads to a biased sample, which, in turn, yields a biased ΔA . A number of enhanced sampling techniques,¹ of which parallel tempering, or replica exchange,^{110–112} is probably the most popular, have been developed to deal with this problem. Another issue that impacts both FEP and NEW calculations is the choice of the coupling parameter, λ . Although ΔA will not depend on the choice of λ , if the latter is not a dynamical variable, the efficiency of the calculations might vary to a considerable extent. Finally, the reliability of free-energy calculations is limited by the accuracy of the current force fields. In the framework of classical mechanics, perhaps the most promising extension is to add explicit contributions arising from

polarization effects to the usual pairwise-additive model of interatomic interactions and, by doing so, to improve the description of induction effects,^{113–117} especially in anisotropic environments. Further progress in applying FEP and NEW to problems in chemistry and biology depends on successfully combining advancements in the theory of free-energy calculations and in these related areas.

Acknowledgment. The authors thank Lawrence Pratt, Stefan Boresch, and Tony Lelièvre for reading and commenting on the manuscript. A.P. thanks the NASA Exobiology Program for supporting this work. C.J. acknowledges financial support from the National Science Foundation (USA), under CHE-0841557. C.C. acknowledges Johan Strumpfer and Aline Kurtzmann for stimulating discussions.

Supporting Information Available: Methodological detail of the free-energy calculations and analyses thereof. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) *Free energy calculations. Theory and applications in chemistry and biology*; Chipot, C., Pohorille, A., Eds.; Springer Verlag: Berlin, New York, 2007.
- (2) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (3) Landau, L. D. *Statistical physics*; The Clarendon Press: Oxford, 1938.
- (4) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
- (5) Jarzynski, C. *Phys. Rev. E: Stat. Nonlin. Soft Matter Phys.* **1997**, *56*, 5018–5035.
- (6) Frenkel, D.; Smit, B. *Understanding molecular simulations: From algorithms to applications*; Academic Press: San Diego, 2002.
- (7) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050–3054.
- (8) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* **1987**, *235*, 574–576.
- (9) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564–568.
- (10) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631–639.
- (11) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (12) Isralewitz, B.; Gao, M.; Schulten, K. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224–230.
- (13) Crooks, G. *J. Stat. Phys.* **1998**, *90*, 1481–1487.
- (14) Clausius, R. *The mechanical theory of heat*; MacMillan: London, 1879.
- (15) Reinhardt, W. P.; Hunter, J. E., III. *J. Chem. Phys.* **1992**, *97*, 1599–1601.
- (16) Crooks, G. E. *Phys. Rev. E* **1999**, *60*, 2721–2726.
- (17) Hummer, G.; Szabo, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3658–3661.
- (18) Park, S.; Schulten, K. *J. Chem. Phys.* **2004**, *120*, 5946–5961.
- (19) Kosztin, I.; Barz, B.; Janosi, L. *J. Chem. Phys.* **2006**, *124*, 064106.
- (20) Minh, D. D. L.; Adib, A. B. *Phys. Rev. Lett.* **2008**, *100*, 180602.
- (21) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 054103.
- (22) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 084109.
- (23) Kullback, S.; Leiber, R. *Ann. Math. Stat.* **1951**, *22*, 79–86.
- (24) Cover, T. M.; Thomas, J. A. *Elements of information theory*; Wiley-Interscience: New York, 1991.
- (25) Lu, N. D.; Kofke, D. A. *J. Chem. Phys.* **1999**, *111*, 4414–4423.
- (26) Gore, J.; Ritort, F.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12564–12569.
- (27) Jarzynski, C. *Phys. Rev. E: Stat. Nonlin. Soft Matter Phys.* **2006**, *73*, 046105.
- (28) Hahn, A. M.; Then, H. *Phys. Rev. E: Stat. Nonlin. Soft Matter Phys.* **2009**, *80*, 031111.
- (29) Widom, B. *J. Chem. Phys.* **1963**, *39*, 2808–2812.
- (30) Wood, R. H.; Muhlbauer, W. C. F.; Thompson, P. T. *J. Phys. Chem.* **1991**, *95*, 6670.
- (31) Lu, N.; Kofke, D. A. *J. Chem. Phys.* **2001**, *114*, 7303–7312.
- (32) Lu, N.; Kofke, D. A. *J. Chem. Phys.* **2001**, *115*, 6866–6875.
- (33) Zuckerman, D. M.; Woolf, T. B. *Phys. Rev. Lett.* **2002**, *89*, 180602.
- (34) Zuckerman, D. M.; Woolf, T. B. *Chem. Phys. Lett.* **2002**, *351*, 445–453.
- (35) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. *Phys. Rev. Lett.* **2003**, *91*, 140601.
- (36) Lu, N.; Kofke, D. A.; Woolf, T. B. *J. Comput. Chem.* **2004**, *25*, 28–39.
- (37) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 144107.
- (38) Minh, D. D. L.; Chodera, J. D. *J. Chem. Phys.* **2009**, *131*, 134110.
- (39) Kofke, D.; Cummings, P. *Fluid Phase Equilib.* **1998**, *150*, 41–49.
- (40) Lehmann, E.; Casella, G. *Theory of point estimation*; Springer, New York, 2003.
- (41) Rice, J. A. *Mathematical statistics and data analysis*; Duxbury Press: Belmont, CA, 2006.
- (42) Oehlert, G. W. *Am. Stat.* **1992**, *46*, 27–29.
- (43) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (44) Hurt, J. *Aplikace Matematiky* **1976**, *21*, 444–456.
- (45) Lehmann, E. *Theory of point estimation*; Wiley: New York, 1983.
- (46) Rodriguez-Gomez, D.; Darve, E.; Pohorille, A. *J. Chem. Phys.* **2004**, *120*, 3563–3578.
- (47) Lyman, E.; Zuckerman, D. M. *J. Phys. Chem. B* **2007**, *111*, 12876–12882.
- (48) Straatsma, T. P.; Berendsen, H. J. C.; Postma, J. P. M. *J. Chem. Phys.* **1986**, *85*, 6720.
- (49) Flyvbjerg, H.; Petersen, H. G. *J. Chem. Phys.* **1989**, *91*, 461–466.
- (50) Müller-Krumbhaar, H.; Binder, K. *J. Stat. Phys.* **1973**, *8*, 1–24.
- (51) Valleau, J. P.; Card, D. N. *J. Chem. Phys.* **1972**, *57*, 5457–5462.
- (52) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016–2027.
- (53) Lee, C. Y.; Scott, H. L. *J. Chem. Phys.* **1980**, *73*, 4591–4596.
- (54) Radmer, R. J.; Kollman, P. A. *J. Comput. Chem.* **1997**, *18*, 902–919.
- (55) Crooks, G. E. *Phys. Rev. E: Stat. Nonlin. Soft Matter Phys.* **2000**, *61*, 2361–2366.
- (56) Shirts, M.; Pitera, J.; Swope, W.; Pande, V. *J. Chem. Phys.* **2003**, *119*, 5740–5761.
- (57) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1989**, *90*, 2460–2470.
- (58) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1991**, *94*, 4532–4545.
- (59) Hummer, G. *J. Chem. Phys.* **2001**, *114*, 7330–7337.
- (60) Hendrix, D. A.; Jarzynski, C. *J. Chem. Phys.* **2001**, *114*, 5974–5981.
- (61) Hummer, G. *Mol. Simul.* **2002**, *28*, 81–90.
- (62) Wood, R. H. *J. Phys. Chem.* **1991**, *95*, 4838–4842.
- (63) Born, M. *Z. Phys.* **1920**, *1*, 45–48.
- (64) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.
- (65) Marcus, R. A. *Ann. Rev. Phys. Chem.* **1964**, *15*, 131–154.
- (66) Warshel, A. *J. Phys. Chem.* **1982**, *86*, 2218–2224.
- (67) Åqvist, J.; Medina, C.; Samuelsson, J. E. *Protein Eng.* **1994**, *7*, 385–391.
- (68) Hummer, G.; Garde, S.; García, A.; Pohorille, A.; Pratt, L. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8951–8955.
- (69) Sham, Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 4458–4472.
- (70) Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. *Acc. Chem. Res.* **2002**, *35*, 358–365.
- (71) Beck, T. L.; Paulaitis, M. E.; Pratt, L. R. *The potential distribution theorem and models of molecular solutions*; Cambridge University Press: New York, 2006.
- (72) Oostenbrink, C.; van Gunsteren, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6750–6754.
- (73) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742–3746.
- (74) Åqvist, J. *J. Comput. Chem.* **1996**, *17*, 1587–1597.
- (75) Dixit, S. B.; Chipot, C. *J. Phys. Chem. A* **2001**, *105*, 9795–9799.
- (76) Asthagiri, D.; Pratt, L. R.; Ashbaugh, H. S. *J. Chem. Phys.* **2003**, *119*, 2702–2708.
- (77) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.
- (78) Huang, D.; Caflisch, A. *J. Med. Chem.* **2004**, *47*, 5791–5797.
- (79) Ytreberg, F. M.; Zuckerman, D. M. *J. Chem. Phys.* **2006**, *124*, 104105.
- (80) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371*, 1118–1134.
- (81) Pearlman, D. A. *J. Phys. Chem.* **1994**, *98*, 1487–1493.
- (82) Liu, S. Y.; Mark, A. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1996**, *94*85–9494, 1749.
- (83) Oostenbrink, C.; van Gunsteren, W. F. *J. Comput. Chem.* **2003**, *24*, 1730–1739.
- (84) Christ, C. D.; van Gunsteren, W. F. *J. Chem. Phys.* **2007**, *126*, 184110.
- (85) Wang, L.; Hermans, J. *J. Chem. Phys.* **1994**, *100*, 9129–9139.
- (86) Boresch, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 5145–5154.
- (87) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 103–118.
- (88) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 119–136.

- (89) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. *Science* **1989**, *244*, 1069–1072.
- (90) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025–9031.
- (91) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- (92) Steinbrecher, T.; Mobley, D. L.; Case, D. A. *J. Chem. Phys.* **2007**, *127*, 214108.
- (93) Wadsö, I. *Thermochim. Acta* **1995**, *267*, 45–59.
- (94) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64–71.
- (95) Karlsson, R.; Larsson, A. *Methods Mol. Biol.* **2004**, *248*, 389–415.
- (96) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.
- (97) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.
- (98) Wang, J.; Deng, Y.; Roux, B. *Biophys. J.* **2006**, *91*, 2798–2814.
- (99) Yu, Y. B.; Privalov, P. L.; Hodges, R. S. *Biophys. J.* **2001**, *81*, 1632–1642.
- (100) Michaux, G.; Reisse, J. *J. Am. Chem. Soc.* **1982**, *104*, 6895–6899.
- (101) Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 5716–5720.
- (102) Jarzynski, C. *Phys. Rev. E: Stat. Nonlin. Soft Matter Phys.* **2002**, *65*, 046122.
- (103) Vaikuntanathan, S.; Jarzynski, C. *Phys. Rev. Lett.* **2008**, *100*, 190601.
- (104) Maragakis, P.; Spichty, M.; Karplus, M. *Phys. Rev. Lett.* **2006**, *96*, 100602.
- (105) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.
- (106) Goette, M.; Grubmüller, H. *J. Comput. Chem.* **2009**, *30*, 447–456.
- (107) Hummer, G.; Pratt, L.; Garcia, A. E. *J. Am. Chem. Soc.* **1997**, *119*, 8523–8527.
- (108) Nanda, H.; Lu, N.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 134110:1–8.
- (109) Pohorille, A.; Darve, E. A Bayesian approach to calculating free energies in chemical and biological systems. *Bayesian inference and maximum entropy methods in science and engineering. AIP Conference Proceedings*, Melville, N.Y., 2006; Mohammad-Djafari, A., Ed.; 872, pp 23–30.
- (110) Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.
- (111) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451–458.
- (112) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (113) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D.; Roux, B. *J. Chem. Theor. Comput.* **2006**, *2*, 1587–1597.
- (114) Patel, S.; Brooks, C. L., III. *Mol. Simul.* **2006**, *32*, 231–249.
- (115) Friesner, R. A. *Adv. Protein Chem.* **2006**, *72*, 79–104.
- (116) Archambault, F.; Chipot, C.; Gutiérrez, I. S.; Luque, F. J.; Schulten, K.; Dehez, F. *J. Chem. Theor. Comput.* **2009**, *5*, 3022–3031.
- (117) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

JP102971X