Published in final edited form as:
J Comput Chem. 2012 February 5; 33(4): 453–465. doi:10.1002/jcc.21989.

Convergence and error estimation in free energy calculations using the weighted histogram analysis method

Fangqiang Zhu[1] and **Gerhard Hummer**[1]
Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, 20892-0520, USA

Abstract

The weighted histogram analysis method (WHAM) has become the standard technique for the analysis of umbrella sampling simulations. In this paper, we address the challenges (1) of obtaining fast and accurate solutions of the coupled nonlinear WHAM equations, (2) of quantifying the statistical errors of the resulting free energies, (3) of diagnosing possible systematic errors, and (4) of optimal allocation of the computational resources. Traditionally, the WHAM equations are solved by a fixed-point direct iteration method, despite poor convergence and possible numerical inaccuracies in the solutions. Here we instead solve the mathematically equivalent problem of maximizing a target likelihood function, by using superlinear numerical optimization algorithms with a significantly faster convergence rate. To estimate the statistical errors in one-dimensional free energy profiles obtained from WHAM, we note that for densely spaced umbrella windows with harmonic biasing potentials, the WHAM free energy profile can be approximated by a coarse-grained free energy obtained by integrating the mean restraining forces. The statistical errors of the coarse-grained free energies can be estimated straightforwardly and then used for the WHAM results. A generalization to multidimensional WHAM is described. We also propose two simple statistical criteria to test the consistency between the histograms of adjacent umbrella windows, which help identify inadequate sampling and hysteresis in the degrees of freedom orthogonal to the reaction coordinate. Together, the estimates of the statistical errors and the diagnostics of inconsistencies in the potentials of mean force provide a basis for the efficient allocation of computational resources in free energy simulations.

INTRODUCTION

The calculation of free energies is one of the main quantitative applications of molecular dynamics or Monte Carlo simulations of molecular systems. In umbrella sampling simulations,[1] a free energy profile (or potential of mean force, PMF) $G(x)$ along a chosen physical or virtual coordinate x is obtained by performing a series of simulations with biasing potentials applied that act as local restraints on x. The weighted histogram analysis method (WHAM)[2,3] has become the standard method to combine the results from the different simulations,[4] and has accordingly been implemented in major simulation software packages.[5] Variants of WHAM can be used for the analysis of replica exchange simulations,[6,7] or in conjunction with the string method.[8] Here we present methods (1) to obtain faster and more accurate solutions of the coupled nonlinear WHAM equations, (2) to quantify the statistical errors of the estimated free energies, (3) to diagnose possible systematic errors in the free energies that result from inadequate sampling of motions

[1]Correspondence to: Fangqiang Zhu or Gerhard Hummer, Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, U.S.A.; Tel: 1-301-402-6290. Fax: 1-301-496-0825. zhuf@niddk.nih.gov. [1] hummer@helix.nih.gov .

orthogonal to x, and (4) to optimize the allocation of computational resources in free energy calculations.

Traditionally, the free energy profile in WHAM is computed by direct iteration of a set of coupled equations until self-consistency is achieved. It was long acknowledged that in practical applications the convergence of the WHAM equations may require more than 100,000 iterations.[9] More seriously, the convergence is typically judged by the variation of the free energy between iterations, which turns out to be only a poor estimator of the actual deviation from the true solution. When each WHAM iteration results in only a small change in the free energy, it may incorrectly appear as if convergence has been achieved. The traditional fixed-point iteration scheme normally has a linear convergence rate, which means that a constant number of iterations is required to gain one more significant digit in the precision of the solution.

To overcome this slow convergence, we note that the WHAM solution corresponds to a maximum likelihood estimate of the free energy parameters,[6,10] and solving the WHAM equations is thus equivalent to finding the maximum of the likelihood function. A variety of numerical techniques with superlinear convergence rate have been developed for such optimization problems,[11] including the Newton-Raphson, trust region, and nonlinear conjugate gradient methods. In these methods, the rate of convergence typically improves as the exact solution is approached. For quadratic objective functions, they can find the exact solution within a finite number of iterations.[11] In this study, we demonstrate that by adopting these superior numerical methods, both the speed and the accuracy in solving the WHAM equations can be significantly improved.

Estimating the statistical errors in the WHAM results is more challenging than solving the WHAM equations. In the original method,[2,3] the error is determined from the expected statistical uncertainty in each bin of the histogram, after properly accounting for correlation effects in the sampling. However, this method ignores the uncertainty in the normalization factor (or equivalently, the free energy) for each umbrella window,[6] and therefore fails to account for the accumulation of errors over multiple intermediate windows, which, unfortunately, is typically the major source of error in practical applications. A more reliable approach is to divide the simulation data into a certain number of blocks, to use WHAM to calculate a PMF from the data in each block, and then to determine the statistical uncertainty from the variance of the PMFs. One may further use the bootstrap strategy to generate new random data according to the estimated distribution, and then determine the uncertainty by comparing the PMFs calculated from these hypothetical trajectories or histograms.[5] In a novel method[6] based on Bayesian statistics, the underlying free energy profile is taken as the unknown quantity, and the histograms as the given observed data. The uncertainty in the free energy is then determined from the posterior likelihood of the parameters. Although the method is conceptually rigorous, the uncertainty cannot be obtained analytically, and in practice has to be obtained from statistical sampling in the parameter space under proper approximations.[6]

In this study we propose a simple method for the error estimation in umbrella sampling simulations, based on the statistical error of the free energy gradient, or the mean force, in each individual umbrella window. For a PMF $G(x)$ along a reaction coordinate x sampled with a series of harmonic biasing potentials $K(x - r_i)^2 / 2$ spaced evenly at $r_i = r_0 + i\Delta r$, and a reference value of $G(r_0) = 0$ in the leftmost umbrella window by definition, we show that the variance in the free energy estimator, given by the square of the cumulative statistical error, is approximately

$$\mathrm{var}\,[G(x)] \approx (K\Delta r)^2 \cdot \sum_{i=1}^{(x-r_0)/\Delta r} \mathrm{var}\left(\bar{x}_i\right). \tag{1}$$

Here $\mathrm{var}\left(\bar{x}_i\right)$ is the squared error in the estimate of the mean position of x in window i which can be obtained straightforwardly from block averages[12] (see Eq. 36 below). Eq. (1) allows us to use simple statistics to estimate the error of a PMF. Furthermore, it clearly reveals the error propagation through multiple windows, and identifies the contribution of each umbrella window to the overall statistical error, thus providing a basis for systematic improvement of the accuracy with minimal computational effort.

We also introduce consistency tests between histograms in adjacent umbrella windows, or between observed and consensus histograms. In particular, we provide a diagnostic that uses information entropy as a measure of deviation between the actual observed histogram $p_i^{\mathrm{obs}}(x)$ in window i from the consensus histogram $p_i^{\mathrm{WHAM}}(x)$ expected from the WHAM free energy:

$$\eta_i = \int p_i^{\mathrm{obs}}(x) \ln \frac{p_i^{\mathrm{obs}}(x)}{p_i^{\mathrm{WHAM}}(x)} dx \tag{2}$$

Large values of this Kullback-Leibler divergence or relative entropy indicate that the free energy surfaces sampled in different umbrella windows are inconsistent. Eq. (2) and an additional pair-wise test for adjacent histograms may help identify potential problems of insufficient equilibration of the degrees of freedom orthogonal to the chosen reaction coordinate.

We thus suggest the following procedure to analyze umbrella sampling simulations: (1) efficiently compute an accurate solution of the WHAM equations and obtain the PMF using a superlinear optimization algorithm; (2) compute the variance of the average reaction coordinate in each umbrella window by block averaging, and then estimate the statistical errors in the PMF using Eq. (1) or its more precise forms discussed in this article; (3) calculate the inconsistency coefficient in Eq. (2) or other similar measures discussed in this study, and identify windows with potential problems of insufficient equilibration. We note that all the analyses above are computationally inexpensive, and can be straightforwardly implemented. Furthermore, on the basis of Eqs. (1) and (2), if one intends to extend the sampling, the computational resources can be invested efficiently by concentrating on regions that contribute most to statistical uncertainties and inconsistencies.

In this study we focus on one-dimensional PMFs in umbrella sampling simulations, which have been widely adopted in a large body of studies on biological and synthetic channels as well as many other important systems. We note that WHAM can also be used on multidimensional histograms involving different state variables (such as temperature) or collective coordinates. The numerical methods for solving the WHAM equations and the consistency tests introduced in this study can be directly applied in such multidimensional cases. A generalization of our estimate of the statistical error to higher dimensions is outlined in Appendix A.

METHODS

In this section, we first introduce the WHAM equations and propose new numerical algorithms to solve the equations. Then we propose a new scheme to estimate the statistical errors in the free energy obtained from WHAM, and consistency tests of the histograms to help identify the potential problem of poor equilibration of the degrees of freedom orthogonal to the chosen reaction coordinate. We conclude this section with a discussion of strategies for the optimal allocation of computational resources in umbrella sampling simulations.

WHAM Equations and Numerical Algorithms

Consider a set of S independent simulations at temperature T, each corresponding to an "umbrella window" with a harmonic biasing potential

$$w_i(x) = \frac{K}{2}(x - r_i)^2, \quad i = 1, \ldots, S. \tag{3}$$

For each simulation i, the time series of x is binned into histograms, with $\{n_{il}\}$ representing the counts in bin l centered at $\{x_l\}$ ($l = 1, \ldots, M$) and $N_i = \sum_{l=1}^{M} n_{il}$ the number of samples in simulation i. If the data from a simulation are correlated, one can scale the counts by an inefficiency factor $(1 + 2\tau_i)^{-1}$ determined from the correlation time τ_i of x in window i, as measured in units of steps in its time series.[6,7] We assume in the following that proper scaling has been done so that the $\{n_{il}\}$ represent the equivalent number of effectively independent samples.

Our objective is to construct the underlying unbiased free energy $G(x)$ that is most consistent with the observed simulation data. For this purpose we aim to determine the unbiased equilibrium probability distribution $\{p_l\}$ ($l = 1, \ldots, M$), representing the probability of finding the coordinate x in each bin when no biasing potential is applied. Then $G(x)$ is given by

$$G(x_l) = -k_B T \ln(p_l / \Delta_l), \tag{4}$$

where k_B is the Boltzmann constant and Δ_l is the width of bin l. Given the $\{p_l\}$, the expected probability distribution $\{p_{il}\}$ in simulation i is also known:[6]

$$p_{il} = f_i c_{il} p_l, \tag{5}$$

where c_{il} is determined by the biasing potential at the center of each bin:

$$c_{il} = \exp[-w_i(x_l)/k_B T], \tag{6}$$

and f_i is a normalization factor to ensure that $\{p_{il}\}$ sum to one:

$$f_i = 1 / \sum_l c_{il} p_l. \tag{7}$$

f_i is thus the reciprocal of the partition function of simulation i,

$$f_i^{-1} = \int p(x) \exp\left[-w_i(x)/k_B T\right] dx. \tag{8}$$

Given the probabilities $\{p_{il}\}$ for simulation i, the likelihood of having $\{n_{il}\}$ counts in the respective bins obeys the multinomial distribution:[6]

$$P_i(n_{i1}, \ldots, n_{iM} | p_{i1}, \ldots, p_{iM}) = \frac{(N_i)!}{\prod_l (n_{il})!} \prod_l (p_{il})^{n_{il}}. \tag{9}$$

The overall likelihood for jointly observing the counts in the S independent simulations is then given by

$$P(\{n_{il}\} | p_1, \ldots, p_M) = \prod_{i=1}^{S} P_i(n_{i1}, \ldots, n_{iM} | p_{i1}, \ldots, p_{iM}). \tag{10}$$

Substituting Eqs. (5) and (9) into Eq. (10) and then taking the logarithm, we obtain[6]

$$\ln P(\{n_{il}\} | p_1, \ldots, p_M) = -A(p_1, \ldots, p_M) + \text{const.}, \tag{11}$$

in which the negative log-likelihood A includes all the terms containing the $\{p_l\}$:

$$A(p_1, \ldots, p_M) = -\sum_{i=1}^{S} N_i \ln f_i - \sum_{l=1}^{M} M_l \ln p_l, \tag{12}$$

where M_l is the total count in the l -th bin, summed over all simulations:

$$M_l = \sum_{i=1}^{S} n_{il}. \tag{13}$$

Finding the maximum likelihood estimates of $\{p_l\}$ is then equivalent to the minimization of A. We note that general discussions of the maximum likelihood approach applied to biased sampling can be found in the statistical literature.[13–15]

To obtain the minimum of A, we take the derivative of A with respect to each individual p_l, noting that in Eq. (12) f_i is a function of the $\{p_l\}$ through Eq. (7):

$$\frac{\partial A}{\partial p_l} = \sum_{i=1}^{S} N_i f_i c_{il} - M_l / p_l. \tag{14}$$

By demanding these derivatives to be zero, we obtain

$$p_l = \frac{M_l}{\sum_i N_i f_i c_{il}} \tag{15}$$

for each bin. Eqs. (7) and (15) jointly form the set of coupled nonlinear WHAM equations, which are traditionally solved by iteratively substituting $\{p_l\}$ and $\{f_i\}$ to obtain a new set of values until self-consistency is achieved. This direct iteration method is numerically inefficient. In the following we introduce techniques with superior convergence rate that directly aim to minimize A, as given in Eq. (12). In addition, for a given problem, A can be used to compare the accuracy of different results, with those closer to the exact solution having smaller A values.

To reduce the dimensionality of the optimization problem, we rewrite the minimization of A with respect to the $\{p_l\}$ into an equivalent minimization with respect to the $\{f_i\}$. Since there are typically far fewer simulations than bins, $S < M$, this reformulation reduces the computational cost. Substitution of Eq. (15) into Eq. (12) results in a new function of the $\{f_i\}$ alone:

$$\tilde{A}(f_1, \ldots, f_S) = -\sum_{i=1}^{S} N_i \ln f_i - \sum_{l=1}^{M} M_l \ln \frac{M_l}{\sum_i N_i f_i c_{il}}. \tag{16}$$

By taking the gradient with respect to the $\{f_i\}$, it is then straightforward to show that the minimum of this new function coincides with the solution of the WHAM equations:

$$\frac{\partial \tilde{A}}{\partial f_i} = N_i \left(\sum_{l=1}^{M} \frac{M_l c_{il}}{\sum_j N_j f_j c_{jl}} - \frac{1}{f_i} \right) = N_i \left(\sum_{l=1}^{M} p_l c_{il} - \frac{1}{f_i} \right), \tag{17}$$

where we used Eq. (15) to rewrite the gradient in terms of the $\{p_l\}$. At the minimum of $\tilde{A}$ these derivatives are zero, and we thus recover Eq. (7). Solving the WHAM equations is equivalent to minimizing Eq. (16) with respect to the $\{f_i\}$, which is numerically more efficient than minimizing Eq. (12) with respect to the $\{p_l\}$ in typical cases.

One may also work on the logarithm of $\{f_i\}$ by a change of variables:

$$g_i = \ln f_i; \tag{18a}$$

$$f_i = \exp(g_i). \tag{18b}$$

According to Eq. (8), the $\{g_i\}$ actually correspond to the total free energies of the system in the presence of the biasing potential i. Using the $\{g_i\}$ as the free variables eliminates the possibility of having (unphysical) negative f_i values in the search of the minimum. The optimization function becomes:

$$\widehat{A}(g_1,\ldots,g_s) = -\sum_{i=1}^{S} N_i g_i - \sum_{l=1}^{M} M_l \ln \frac{M_l}{\sum_i N_i c_{il} e^{g_i}} \tag{19}$$

Furthermore, it can be shown[13] that $\widehat{A}(g_1,\ldots,g_s)$ is a convex function with nonnegative second derivatives everywhere, and thus has a single minimum.[13] The derivatives of $\widehat{A}$ with respect to the $\{g_i\}$ are given by

$$\frac{\partial \widehat{A}}{\partial g_i} = N_i \left(e^{g_i} \sum_{l=1}^{M} \frac{M_l c_{il}}{\sum_j N_j c_{jl} e^{g_j}} - 1 \right) = N_i \left(e^{g_i} \sum_{l=1}^{M} p_l c_{il} - 1 \right). \tag{20}$$

Because multiplying all $\{f_i\}$ by a constant factor, or adding a constant to all $\{g_i\}$, does not alter the free energy profile $G(x)$, we can set g_1 to zero without loss of generality, and minimize $\widehat{A}$ with respect to g_2, …, g_S.

In typical umbrella sampling setups, the histogram of a particular umbrella window only substantially overlaps with the histograms of neighboring windows. Consequently, the derivative $\partial\widehat{A}/\partial g_i$ for window i is predominantly determined by those values of $\{g_j\}$ with $|j - i|$ being small, and is hardly affected by distant windows. The derivative therefore mainly reflects the local consistency of the free energy across a small range of umbrella windows, but poorly indicates the real distance of g_i to its true value in the exact solution. To improve the performance of the optimization, one may use the incremental changes of the $\{g_i\}$ over adjacent windows as free variables in the minimization:

$$\Delta g_i = g_{i+1} - g_i; \tag{21a}$$

$$g_i = \sum_{j=1}^{i-1} \Delta g_j. \tag{21b}$$

The function $\widehat{A}$ to be minimized and its derivatives with respect to the $\{\Delta g_i\}$ (i=1, …, S −1) are then given by

$$\widehat{A}(\Delta g_1,\ldots,\Delta g_{s-1}) = -\sum_{i=2}^{S} N_i \sum_{j=1}^{i-1} \Delta g_j - \sum_{l=1}^{M} M_l \ln \frac{M_l}{\sum_i N_i c_{il} \exp\left(\sum_{j=1}^{i-1} \Delta g_j \right)}; \tag{22a}$$

$$\frac{\partial \widehat{A}}{\partial \Delta g_i} = \sum_{j=i+1}^{S} \frac{\partial \widehat{A}}{\partial g_j} = \sum_{j=i+1}^{S} N_j \left[\exp\left(\sum_{m=1}^{j-1} \Delta g_m \right) \sum_{l=1}^{M} \frac{M_l c_{jl}}{\sum_k N_k c_{kl} \exp\left(\sum_{m=1}^{k-1} \Delta g_m \right)} - 1 \right]. \tag{22b}$$

Our experimentation shows that the change of variables to $\{\Delta g_i\}$ makes the convergence considerably faster, especially when the number of umbrella windows is large.

The target function $\widehat{A}$ can be minimized using a variety of numerical algorithms.[11] In the Newton-Raphson algorithm,[11] a quadratic expansion of the target function at the current search position is obtained from the local gradient and Hessian matrix, and the minimum of this approximate quadratic function is taken as the new search position. This algorithm was used on a similar problem,[16] and was found to be efficient yet slightly less reliable than the direct iteration method.[16] Some variants of the Newton-Raphson algorithm are both efficient and reliable, and are thus more widely adopted in practice. Here we test two such variants, the subspace trust region method and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method[11] with cubic line search, as implemented in the "fminunc" function in the Matlab[17] Optimization Toolbox. In the trust region method, a two-dimensional subspace is identified from the local curvature, and a trial move in the subspace is computed to minimize the approximate quadratic function, subject to the constraint that the move must lie within a given trust radius. If the trial move results in a higher value of the actual target function, the trust radius will be decreased to ensure that the quadratic expansion is sufficiently accurate. The quadratic approximation will become increasingly accurate upon approaching the true minimum, and therefore the iterations will converge rapidly. In the BFGS method, the Hessian matrix is not explicitly computed and is instead approximated and updated at each iteration step, and a line search is performed to locate the minimum along a given direction.

Error Estimation

Eq. (15) indicates that the statistical error in p_l obtained from WHAM can be formally attributed to two uncertainties: (1) in the observed counts in bin l, and (2) in the normalization factors $\{f_i\}$. Although the former uncertainty can be readily estimated, the latter is much more complicated to quantify. In typical umbrella sampling setups, only histograms of neighboring windows have substantial overlap, and the free energy difference between two distant windows can only be determined through a relay of intermediate windows. Consequently the uncertainties in the histograms of these intermediate windows contribute to the statistical errors in the $\{f_i\}$, and in turn to the errors in $G(x)$. In practice, the sampling within narrow windows is usually quite efficient, and the cumulative errors in the $\{f_i\}$ thus tend to dominate the error in the estimated free energy profile.

To estimate the statistical errors in the $\{f_i\}$, we first define the total free energy of the system in the presence of a harmonic biasing potential $K(x - r)^2 / 2$ centered at r:

$$G_r(r) = -k_B T \ln \int p(x) \exp\left[\frac{-k(x-r)^2}{2k_B T}\right] dx. \tag{23}$$

Using Eq. (8), we have

$$G_r(r_i) = k_B T \ln f_i = k_B T\, g_i, \tag{24}$$

with g_i defined in Eq. (18). Therefore the $\{f_i\}$ or $\{g_i\}$ are determined by the values of $G_r(r)$ at the discrete points $\{r_i\}$.

Furthermore, taking the derivative of $G_r(r)$, we have:

$$\frac{\partial}{\partial r}G_r(r)=\frac{K\int(r-x)\,p(x)\exp\left[\frac{-K(x-r)^2}{2k_BT}\right]dx}{\int p(x)\exp\left[\frac{-K(x-r)^2}{2k_BT}\right]dx}. \tag{25}$$

This derivatives at the $\{r_i\}$ are thus given by

$$\frac{\partial}{\partial r}G_r(r_i)=\langle K(r_i-x)\rangle_i=\langle F_i\rangle_i, \tag{26}$$

where $F_i = K(r_i - x)$ is the restraining force arising from the harmonic potential $w_i(x)$ in Eq. (3), and $\langle\rangle_i$ denotes the ensemble average in the presence of the biasing potential $w_i(x)$. The i -th umbrella simulation actually represents such an ensemble, and thus can be used to obtain the ensemble averages:

$$\langle F_i\rangle_i \approx K\left(r_i-\bar{x}_i\right)\equiv \bar{F}_i, \tag{27}$$

where $\bar{x}_i$ and $\bar{F}_i$ are the mean position of x and the mean restraining force in the simulation, respectively. If the biasing potentials are not harmonic, the estimators of the free energy gradient have to be modified appropriately (see Appendix B). We can integrate the gradients at the $\{r_i\}$ using the trapezoidal rule:

$$G_r(r_k)-G_r\left(r_j\right)\approx\sum_{i=j}^{k-1}\frac{1}{2}\left(\bar{F}_i+\bar{F}_{i+1}\right)(r_{i+1}-r_i),\ j<k. \tag{28}$$

Eqs. (24) and (28) show that the $\{f_i\}$ or $\{g_i\}$ can be obtained approximately by integrating the gradients of $G_r(r)$, or the mean restraining forces $\left\{\bar{F}_i\right\}$ from the simulations,[18] analogous to thermodynamic integration. This method can also be conveniently applied in multidimensional cases, and therefore has been adopted in conjunction with the string method[19] to obtain a free energy profile in a high-dimensional coordinate space. For the one-dimensional PMFs considered in this study, the method is expected to give similar results as WHAM does, if $G_r(r)$ varies smoothly as a function of r between windows. Whereas the systematic errors arising from discretization of the chosen coordinate are expected to be smaller in WHAM than in the gradient-based method, the statistical errors in the latter can nonetheless be taken as reasonable estimates of the statistical errors in the WHAM result. We expect the quality of this approximation to decrease when the windows are spaced too far compared to features in $G_r(r)$.

Because the $\{f_i\}$ or $\{g_i\}$ can be expressed as functions of the $\left\{\bar{F}_i\right\}$, their statistical errors can also be calculated from the variances of the $\left\{\bar{F}_i\right\}$, which according to Eq. (27) are determined by the variances of the mean coordinates $\left\{\bar{x}_i\right\}$:

$$\mathrm{var}\left(\bar{F}_i\right)=K^2\,\mathrm{var}\left(\bar{x}_i\right). \tag{29}$$

If the $\{r_i\}$ are equally spaced, i.e.,

$$r_{i+1}-r_i=\Delta r, \tag{30}$$

Eq. (28) can be simplified into

$$G_r(r_k)-G_r\left(r_j\right)=\Delta r\left(\frac{\bar{F}_j+\bar{F}_k}{2}+\sum_{i=j+1}^{k-1}\bar{F}_i\right). \tag{31}$$

The resulting variance of the estimator for the free energy difference over repeated measurements is then given simply by

$$\mathrm{var}\left[G_r(r_k)-G_r\left(r_j\right)\right]=(K\Delta r)^2\left[\frac{\mathrm{var}\left(\bar{x}_j\right)+\mathrm{var}\left(\bar{x}_k\right)}{4}+\sum_{i=j+1}^{k-1}\mathrm{var}\left(\bar{x}_i\right)\right]. \tag{32}$$

The corresponding standard deviation, given by the square root of the variance, is typically taken as the statistical error of the free energy difference above. Here we assumed that the errors in the $\left\{\bar{x}_i\right\}$ are uncorrelated. For some simulation protocols, the $\left\{\bar{x}_i\right\}$ of different windows are correlated and one should then include their covariances in addition to the variances. In Hamiltonian replica exchange simulations[20] the $\{x_i\}$ of different windows are statistically independent of each other in the joint probability distribution. For sufficiently long Hamiltonian replica exchange simulations that fully explore the combined phase space, the covariances of the $\left\{\bar{x}_i\right\}$ are thus expected to vanish, and Eq. (32) approximately holds in such cases.

If the harmonic springs are relatively stiff (with large K), as is often the case in umbrella sampling simulations, $p(x)$ would vary little within a narrow umbrella window. In such cases we have

$$\int p(x)\exp\left[\frac{-K(x-r_i)^2}{2k_BT}\right]dx\;\approx\;p(r_i)\int\exp\left[\frac{-K(x-r_i)^2}{2k_BT}\right]dx=p(r_i)\sqrt{\frac{2\pi k_BT}{K}}. \tag{33}$$

$G_r(r)$ in Eq. (23) can then be simplified into

$$G_r(r_i)\approx G(r_i)+const. \tag{34}$$

Therefore under the stiff spring approximation,[21] the $\{G_r(r_i)\}$ directly represent the PMF, and thus also share the same statistical errors as the PMF.

The variance $\mathrm{var}\left(\bar{x}_i\right)$ can be estimated from the trajectory of x in simulation i. For a correlated trajectory, this variance depends on the autocorrelation function of the data. By block averaging, however, $\mathrm{var}\left(\bar{x}_i\right)$ can be estimated without explicitly computing the autocorrelation function.[12] In this method the time series $x_{i\alpha}$ of the x - coordinate in window i is divided into n blocks of sufficiently large size m, so that the averages of the n blocks are independent of each other, and then $\mathrm{var}\left(\bar{x}_i\right)$ is estimated from the variance of these averages:[12]

$$\mathrm{var}\left(\bar{x}_i\right)=\frac{1}{n(n-1)}\sum_{b=0}^{n-1}\left(\frac{1}{m}\sum_{\alpha=bm+1}^{(b+1)m}x_{i\alpha}-\bar{x}_i\right)^2. \tag{35}$$

For small n, this estimate tends to be noisy; for large n, the blocks are correlated. While n between 5 and 10 are typically used in practice, for accurate estimates it is advisable to plot Eq. (35) as a function of $\log(n)$ to identify a plateau[12] at small n and use that for the estimate of $\mathrm{var}\left(\bar{x}_i\right)$.

From the variances of $\bar{x}$ in each individual simulation, the uncertainty in the $\{g_i\}$ (assuming g_1=0) can be obtained via Eqs. (24) and (32):

$$\mathrm{var}\left(g_j\right)=\left(\frac{K\Delta r}{k_BT}\right)^2\left[\frac{\mathrm{var}\left(\bar{x}_1\right)+\mathrm{var}\left(\bar{x}_j\right)}{4}+\sum_{i=2}^{j-1}\mathrm{var}\left(\bar{x}_i\right)\right],\ j>1. \tag{36}$$

This estimate is the central result of the second part of this paper, and clearly shows the accumulation of the statistical error over the intermediate umbrella windows. Under the stiff spring approximation discussed earlier, $G_r(r_i)$ or $\{g_i\}$ directly represent the PMF (Eq. 34), and therefore Eq. (36) also gives the error estimate for the PMF, thus leading to Eq. (1) except for minor differences arising from the first and last windows.

Additionally, the effective number of independent samples in the trajectory, N_i, can be estimated from the variance of $\bar{x}_i$:

$$N_i\approx \mathrm{var}\left(x_i\right)/\mathrm{var}\left(\bar{x}_i\right)\approx n_i/\left(2\tau_i+1\right), \tag{37}$$

where $\mathrm{var}(x_i)$ is the variance of the coordinate x in trajectory i, and n_i the length of the trajectory. We note that with this relation and $\mathrm{var}(x_i)\approx k_BT/K$ (ignoring additional curvature effects) one can rewrite Eq. (1) in terms of the correlation time τ_i :

$$\mathrm{var}\left[G(x)\right]\approx k_BTK(\Delta r)^2\sum_i\frac{2\tau_i+1}{n_i}. \tag{38}$$

WHAM Consistency Tests

The free energy profile $G(x)$ obtained from WHAM is a function of the reaction coordinate x, with all coordinates orthogonal to x integrated out. The determination of an accurate $G(x)$ requires that these orthogonal coordinates be sufficiently sampled in the simulations. In each individual simulation, although the reaction coordinate x is restrained to a reference position, the subspace formed by the orthogonal coordinates may feature multiple metastable states or local minima, and infrequent transitions between these states will result in long autocorrelation in the trajectory. If the simulation is not long in comparison to the corresponding correlation time τ_i, the true error in $\bar{x}_i$ will be underestimated by block averaging or other methods based on the autocorrelation function. In extreme cases, the system may stay in a single state without ever visiting other states in the subspace orthogonal to x during the entire simulation i, resulting in a severe underestimation of the errors in $\bar{x}_i$ and the free energy. In such cases the trajectory of the single simulation alone will show no indication of the insufficient sampling; however, if other states are visited in the simulations of neighboring windows, the problem can be inferred by a comparison of the simulation results. If different states of the orthogonal coordinates are sampled in two adjacent umbrella windows, the two simulations will normally produce inconsistent probability distributions of x. The problem of insufficient sampling of the orthogonal degrees of freedom, therefore, can be detected by checking the consistency of the histograms in neighboring windows, as described below.

Consider two simulations in adjacent umbrella windows centered at r_1 and r_2. From the simulation trajectories, we obtain the probability distributions $p_1(x)$ and $p_2(x)$, and aim to test whether they are consistent with a common underlying unbiased probability distribution $p(x)$. In principle, we may reconstruct $p(x)$ from either $p_1(x)$ or $p_2(x)$, and then compare the two resulting distributions for consistency. In practice, however, the $p(x)$ obtained from $p_i(x)$ is only accurate in the vicinity of r_i, and will bear increasingly large statistical uncertainties at increasing distances from r_i. Indeed, in the WHAM equations $p_1(x)$ and $p_2(x)$ primarily contribute to the shape of $p(x)$ around the region between r_1 and r_2, and the comparison should therefore also be focused on this range. For this purpose, we imagine a virtual umbrella window at the midpoint $r^* = (r_1 + r_2)/2$, with the biasing potential given by $w*(x) = \frac{K}{2}(x - r*)^2$ as in Eq. (3). The probability distribution of x in this virtual umbrella window can be estimated from either $p_1(x)$ or $p_2(x)$:

$$p_i^*(x) = \frac{p_i(x)\exp\{[w_i(x) - w*(x)]/k_BT\}}{\int_{-\infty}^{\infty} p_i(x)\exp\{[w_i(x) - w*(x)]/k_BT\}\,dx}, \quad i=1,2. \tag{39}$$

If $p_1(x)$ and $p_2(x)$ indeed arise from a common unbiased probability distribution, $p_1^*(x)$ and $p_2^*(x)$ should be identical within statistical uncertainties. Furthermore, $p_1^*(x)$ and $p_2^*(x)$ are peaked in the region where $p_1(x)$ and $p_2(x)$ have maximum overlap, and could be compared on the same footing. A significant disagreement of $p_1^*(x)$ and $p_2^*(x)$ would then indicate that the two simulations are probably sampling different free energy surfaces.

Several statistical methods can be applied to test $p_1^*(x)$ and $p_2^*(x)$. In this study we adopt a simple protocol based on the Kolmogorov-Smirnov test. The method uses the maximum difference between the two respective cumulative distribution functions to quantify the deviation between $p_1^*(x)$ and $p_2^*(x)$. For two histograms arising from a common probability distribution, the expected value of this maximum difference asymptotically approaches zero following the inverse square root of the sample size, as the latter approaches infinity.

Therefore we define an inconsistency coefficient, $\theta_{1,2}$, as an empirical measure for the discrepancy between $p_1^*(x)$ and $p_2^*(x)$:

$$\theta_{1,2}=\sqrt{\frac{N_1N_2}{N_1+N_2}}\max_x\left|\int_{-\infty}^{x}\left[p_1^*(x)-p_2^*(x)\right]dx\right|. \tag{40}$$

Note that N_1 and N_2 above should be the effective number of independent samples, which can be estimated from Eq. (37) for correlated data. An abnormally high $\theta_{1,2}$ indicates inconsistency between the results of the two simulations. For any pair of adjacent umbrella windows i and $i+1$, we can similarly calculate an inconsistency coefficient $\theta_{i,i+1}$. We note that other methods, such as Pearson's χ^2 test, can also be used to examine $p_1^*(x)$ and $p_2^*(x)$.

In addition to the pair-wise consistency test above, one can also check the agreement between the actual observed histograms and the consensus histograms predicted from the WHAM results. Given the $\{p_l\}$ and $\{f_i\}$ from the WHAM calculations, the consensus histogram in each umbrella window, denoted by $\widehat{p_i}(x_l)=\widehat{p_{il}}$, can be obtained according to Eq. (5) and then compared to the observed probability distribution $p_i(x_l)$ in the same window. Among various possible methods, in this study we adopt the relative entropy, denoted by η_i, as a metric for the consistency between $\widehat{p_i}(x)$ and $p_i(x)$:

$$\eta_i=\int p_i(x)\ln\frac{p_i(x)}{\widehat{p_i}(x)}dx=\sum_{l=1}^{M}\frac{n_{il}}{N_i}\ln\frac{n_{il}}{N_i\widehat{p}_{il}}. \tag{41}$$

η_i is guaranteed to be non-negative, with smaller η_i values indicating better agreement between the two probability distributions.

Optimal Allocation of Computational Resources in Umbrella Sampling

Assume that one sets up umbrella sampling simulations aiming to calculate the free energy difference between the first and last umbrella windows. For simulation i, the variance of the mean coordinate $\mathrm{var}\left(\bar{x}_i\right)$ asymptotically depends on the simulation length n_i as $\mathrm{var}\left(\bar{x}_i\right)=v_i/n_i$, where v_i is a constant for window i. To obtain an estimate of the $\{v_i\}$, one could first simulate each umbrella window for a short length n'. We caution that these initial simulations should be sufficiently long to avoid significant systematic errors, which can be indicated by the inconsistency coefficients as mentioned earlier. Under the constraint of a fixed total length of all simulations $\Sigma_i n_i = N$ imposed by available computational resources, we seek to optimally allocate the $\{n_i\}$ for the individual simulations to minimize the statistical error, which according to Eq. (1) is determined by $\Sigma_i v_i/n_i$. Utilizing the Cauchy-Schwarz inequality $\left(\Sigma_i x_i^2\right)\left(\Sigma_i y_i^2\right)\geq\left(\Sigma_i x_i y_i\right)^2$, we have

$$\left(\sum_i\frac{v_i}{n_i}\right)\left(\sum_i n_i\right)\geq\left(\sum_i\sqrt{v_i}\right)^2. \tag{42}$$

The two sides above are equal if and only if $n_i=\lambda\sqrt{v_i}$, with λ a normalization constant. This relation can also be derived by applying Lagrange multipliers. Therefore in the optimal allocation scheme one should sample each umbrella window with the simulation length proportional to $\sqrt{v_i}\propto\left[\mathrm{var}\left(\bar{x}_i\right)\right]^{1/2}$, based on the variances of the mean position estimated

from trial runs of equal length n'. We note that whereas the prescription above may serve as a general guideline, in practice other technical factors also need to be considered. In the replica exchange scheme, e.g., all simulations must have the same length. In such cases, one could either first simulate all windows with replica exchange and then perform extension runs for those windows that require further sampling, or introduce additional windows in regions of poor statistics.

RESULTS

In this section, we present two tests for the numerical algorithms and the error estimation scheme discussed in the Methods section. In the first test, we analyze data from a practical application, a set of trial molecular dynamics (MD) simulations in our study of ion conduction through a protein channel.[22] Our focus in this test is on the performance of different numerical algorithms to solve the WHAM equations. In the second test, we carry out Monte-Carlo (MC) simulations on a simple potential designed to reproduce some common problems (such as hysteresis) in practical applications. Moreover, with the true free energy profile known, the absolute errors can be calculated, allowing us to assess the quality of the estimated statistical errors.

Umbrella Sampling of Na^+ in an Ion Channel

As described in ref. 22, a total of 153 umbrella windows with a uniform spacing of 0.5 Å were employed to determine the free energy for passage of a Na^+ ion through the transmembrane pore of the GLIC channel.[22] In each window an umbrella potential with a spring constant of 10 kcal/mol/Å^2 was applied on the z -coordinate of the Na^+ ion along the membrane normal direction, and a lateral restraint on the xy plane was applied to confine the ion in the bulk region.[22] For the present study, we performed new calculations in which we implemented Hamiltonian replica exchange[20] between neighboring windows. Here we analyze data from a set of trial MD simulations of 1 ns in each window. We construct histograms with a uniform bin width of 0.02 Å for each of the 1-ns trajectories as the input for the WHAM calculation. We note that the resulting free energy has a considerable entropic component, due to a significant variation in the lateral area accessible to the ion at different z positions. The mean squared deviation of the ion in the xy plane from the axis varies from ~0.6 Å^2 at the narrowest portion of the pore to ~28 Å^2 in the bulk region.

We first compare the performance of the superlinear (trust region and BFGS) algorithms and the traditional direct iteration method in solving the WHAM equations. In all methods the iterations start with given initial values of the probabilities $\{p_l\}$ or the normalization factors $\{f_i\}$. Although it is common practice to assign a constant initial value to all $\{p_l\}$ or $\{f_i\}$, it was shown that using more accurate estimates of the free energy as initial values can significantly speed up the convergence in the direct iteration method.[23] In fact, approximate $\{f_i\}$ or $\{g_i\}$ can be calculated by integrating the gradients, or the mean restraining forces (Eqs. 24, 31). Here we test each method, first for constant initial values, and then by using the coarse-grained profile determined from the mean forces. In our implementation of the direct iteration method, the convergence is deemed achieved when the relative change in $\{p_l\}$ for any l by a WHAM iteration is smaller than a given threshold δ. We test each case with a larger (10^{-3}) and a smaller (10^{-6}) δ value.

As shown in Table 1, the trust region and BFGS numerical algorithms yield more accurate results than the traditional direct iteration method, indicated by smaller values of the target function A. In fact, the direct iteration method with a convergence threshold of $\delta = 10^{-3}$ leads to a rather inaccurate free energy profile (Fig. 1A, *blue dashed* curve), with an error of more than 2 k_BT, when starting with uniform initial values. When the gradient-based estimates of the $\{f_i\}$ are used as initial values, the direct iteration method indeed converges

faster, with the threshold of $\delta = 10^{-3}$ almost satisfied already at the start of the iterations. To achieve better accuracy with a more stringent threshold of $\delta = 10^{-6}$, however, a large number of iterations is still required, taking a computational time at least an order of magnitude longer than the trust region and BFGS methods. The example also demonstrates that apparently small variations between WHAM iterations, on the order of δ, do not necessarily mean that convergence at that level of accuracy has been achieved. At the end of the WHAM iterations in our test, the free energies change by less than $10^{-6}\, k_BT$ in a single iteration, although their absolute errors are still on the order of 0.01 k_BT, implying that to gain one more significant digit in the result, one needs to perform thousands of iterations. In contrast, the convergence rate of superlinear algorithms increases when approaching the final solution; in our case only ~500 more equivalent iterations are required in the trust region method when the termination tolerance is decreased from 10^{-6} to $\sim 2\times 10^{-16}$ (the minimum relative change that can be represented by a double-precision floating-point number).

The results of the free energy calculation are shown in Fig. 1. The coarse-grained free energy profile (*red* curve, Fig. 1A) obtained by integrating the mean forces (Eq. 31) indeed closely overlaps with that (*blue solid* curve, Fig. 1A) obtained by solving the WHAM equations. The statistical uncertainties in the average position $\bar{x}_i$ (Fig. 1B) are largest in the region between $x = -20$ Å and $x = 0$ Å, which is actually the narrow part of the channel where the ion is coupled to the motions of the protein side chains. The flat baselines at the two ends of the free energy curves (Fig. 1A) represent the bulk water regions at the two sides of the channel, respectively. Although the MD simulations were performed under 3D periodic boundary conditions, the umbrella windows do not span the entire length of the unit cell and the windows at the two ends are still too far from each other to have an overlap across the periodic boundary. Nonetheless, in the absence of a membrane potential, the two levels at both ends of the ideal PMF should match, although in a calculated PMF they may not match exactly due to the statistical errors in the finite sampling. Overall, the calculated statistical errors of the free energy shown in Fig. 1A appear to be reasonable, and offer a faithful estimate of the uncertainty in the baseline difference. However, as the true free energy profile is not known in this case, we could not thoroughly calibrate the errors. To systematically examine the validity of our error estimation method, in the following we design a model potential that allows us to obtain the absolute errors of the calculation.

Umbrella Sampling of a Model Potential with Hysteresis

As illustrated in Fig. 2A and discussed earlier, when the reaction coordinate x is fixed at a given value, a multidimensional system may still populate different metastable states separated by high barriers in directions orthogonal to x. Insufficient sampling of the relevant orthogonal coordinates (or "solvent degrees of freedom") may give rise to systematic errors in the obtained free energy and result in hysteresis. Here, we reduce the problem further and assume fast relaxation in the orthogonal degrees of freedom (y) in each well of Fig. 2A, but slow transitions between the wells. Analogous to the Marcus theory of electron transfer, we can then collapse motion in y and treat the problem as a transition between two one-dimensional surfaces, as shown in Fig. 2B. In addition to the continuous dimensionless reaction coordinate x, our model includes an orthogonal y coordinate, which takes discrete values of 1 or 2, representing two metastable states with different potentials $E_i(x)$:

$$H(x,y) = \begin{cases} E_1(x) & \text{if } y=1 \\ E_2(x) & \text{if } y=2 \end{cases}, \tag{43}$$

For simplicity, we use harmonic potentials

$$E_i(x) = (x - a_i)^2/2, i = 1, 2, \tag{44}$$

in which $a_1 = -4$ and $a_2 = 4$. The free energy as a function of x is then

$$G(x) = -k_B T \ln\left[\sum_y e^{-H(x,y)/k_B T}\right] = -k_B T \ln\left[e^{-E_1(x)/k_B T} + e^{-E_2(x)/k_B T}\right], \tag{45}$$

as shown in Fig. 2C.

To sample the free energy, we carried out MC simulations in a total of 21 umbrella windows, covering the range from $r_0 = -5$ to $r_{20} = 5$ with a uniform spacing of $\Delta r = 0.5$. In each simulation, an umbrella potential with a spring constant of $K = 17\ k_B T$ was applied. At every MC step we made a random choice for either a y -move with a probability of 10^{-4}, or otherwise an x -move. In an attempted y -move, the y -coordinate is switched to the alternative value (i.e., from 1 to 2 or vice versa). In an attempted x -move, the x -coordinate is moved by a random displacement from a uniform distribution in [−0.24, 0.24]. In either case the energy at the new coordinates is calculated, and the move is accepted or rejected according to the Metropolis criterion. The center, or reference position, of each window was used as the initial x -coordinate for the corresponding simulation. The initial y -coordinate was set to 1 for the first 12 windows (at $r_0 = -5$ to $r_{11} = 0.5$), and to 2 for the remaining 9 windows (at $r_{12} = 1$ to $r_{20} = 5$).

In Fig. 3, we examine the WHAM results and the error estimation from simulation data of various lengths measured by the number of MC steps. We construct histograms with a uniform bin width of 0.05 for the individual simulations, and then solve the WHAM equations by the minimization technique described in Methods. The resulting coarse-grained free energy profiles $G_r(x)$ (*red* curves, Fig. 3A), calculated by integrating the local gradients (Eq. 31), closely match the WHAM results $G(x)$ (*blue* curves, Fig. 3A). Therefore, we expect that the errors in $G(x)$ can be determined from the errors in the mean x -coordinate $\bar{x}_i$ in each simulation window, as described in Methods.

For short simulations ($N = 10{,}000$), we find large systematic deviations in the free energy, and significantly underestimated statistical errors. The reason for the deviations is that in these short simulations the y -coordinate typically remains unchanged and undergoes no transition to the alternative state during the entire simulation. This lack of equilibration gives rise to particularly large errors in the windows at $r_{10} = 0$, with equal expected probabilities of the two states with $y = 1$ and 2, and at $r_{11} = 0.5$, which sampled a different state ($y = 1$) than the expected most probable state ($y = 2$) because of the biased initial condition. Consequently both the barrier height and the relative difference between the two local minima deviate considerably from those in the true free energy profile (*green* curve, Fig. 3A). As a result of the inadequate statistics in y, the errors $\sigma(\bar{x}_i) = \sqrt{\mathrm{var}(\bar{x}_i)}$ in $\bar{x}_i$ are significantly underestimated for the two windows, which in turn leads to an underestimation of the errors in the free energy (Fig. 3A, $N = 10{,}000$).

In principle, the discontinuity in y can be identified by a direct examination of the y -coordinate in neighboring windows. In practice, however, this is not always possible in simulations of complex biomolecular systems with an enormous number of degrees of freedom, in which the states with slow transitions in the orthogonal subspace can probably

only be described by collective order parameters. In such cases, the inconsistency coefficient, obtained from the reaction coordinate x, can still be readily used to detect the problem. For example, in Fig. 3C (N =10,000), the inconsistency coefficient θ for the two problematic windows at r_{11} =0.5 and r_{12} =1 is found to be abnormally high, indicating that the two corresponding simulations were not sampling a common unbiased distribution of x. Indeed, the y -coordinate in these two simulations stays at 1 and 2, respectively, thus resulting in the sampling of different potential surfaces.

As the simulations were extended to N =100,000 and N =1,000,000, multiple transitions of the y -coordinate occur in the windows near r_{10} =0. As a result, the autocorrelation functions of x decay slowly, and the block averaging method properly reports the large statistical uncertainty of $\bar{x}_i$ in these windows, as shown in Fig. 3B. Consequently, the estimated statistical error in the calculated free energy (Fig. 3A) is now consistent with the actual absolute errors. The large variances in $\bar{x}_i$ for the windows near r_{10} =0 result in small and more reasonable values of N_i, the effective numbers of independent samples (Eq. 37) in these simulations. When these proper N_i values are used, the inconsistency coefficient θ (Eq. 40) for the two windows r_{11} =0.5 and r_{12} =1 is no longer high (Fig. 3C), indicating that the discrepancy between the two histograms is not abnormally large in comparison to expectations. Indeed, the infrequent transitions in the y -coordinate are already reflected in the estimated variances of $\bar{x}_i$, and are thus accounted for in the error estimation of the free energy.

As shown in Fig. 3D, the errors in the observed histograms are also reflected in the consistency with respect to the consensus histograms. Insufficient sampling of the orthogonal coordinate in an umbrella window usually results in higher values of the relative entropy (η) in this window and in its neighboring windows. For example, for N =10,000, as mentioned earlier, major inconsistencies occur near r_{11} =0.5 and r_{12} =1, and the η values are indeed higher in the corresponding windows. For N =100,000 and N =1,000,000, relatively large η values occur near the central window at r_{10} =0 which bears the largest uncertainty. Overall, the η values decrease with longer simulation length N, indicating improved consistency with more extensive sampling. We note that for a system characterized by two local minima on the energy surface that are partially overlapping in projection, the resulting free energy will typically feature a barrier at an intermediate location. In this barrier region, the sampling is more challenging because this intermediate region tends to be of high energy relative to the minima even within one well and, more seriously, both minima need to be visited with the correct proportionality. Consequently, for such systems the barrier region generally bears larger uncertainty in the free energy calculation, consistent with the results for our model system here.

To further test the quality of the error estimation, we repeated the simulations 100 times with identical initial coordinates as described above and different random seeds. When the sampling size N is relatively small, the results are biased by the particular initial condition and the systematic error in the free energy is significant due to hysteresis. As shown in Fig. 4, in this case the estimated errors are smaller than the absolute errors with respect to the exact free energy. When N becomes larger, however, the systematic error arising from a particular initial condition becomes insignificant in comparison to the statistical error, and the latter becomes the main contributor to the actual error. When N is large enough so that all umbrella windows get well equilibrated, the statistical errors obtained from Eq. (32) indeed offer a faithful estimate for the absolute errors, as shown in Fig. 4. The results here further confirm that an accurate estimate of the errors is only possible after sufficient equilibrations (i.e., in the asymptotic limit, where statistical errors dominate).

DISCUSSION

In this study, we demonstrated that superior numerical optimization algorithms, such as the trust region and BFGS methods, can solve the WHAM equations with higher accuracy and significantly faster speed than the traditional direct iteration method. In the direct iteration method, the incremental change in each WHAM iteration can be smaller by orders of magnitude than the actual distance to the solution, resulting in slow convergence rates and potentially misperceptions of false convergence. In contrast, common numerical minimization algorithms take into account the curvature of the target function and make a more informed move in each iteration, thus requiring significantly fewer iterations to find the solution. Furthermore, as the search approaches the optimal solution the convergence rate of these algorithms will speed up, and their performance gain over the direct iteration method will therefore be more significant when results of higher precision are desired. In our tests we found that the BFGS method is more efficient than the trust region method, mainly because BFGS constructs an approximate Hessian matrix and thus avoids the costly computation of the second derivatives. If the number of the umbrella windows is very large, the nonlinear conjugate gradient method is another viable choice, as it avoids the construction of the Hessian matrix altogether.

One underlying reason for the poor convergence of the iterative solution to the WHAM equations is that the free energies of each umbrella run need to be adjusted globally, but the iteration operates locally. In our case of ion translocation through a membrane channel, the WHAM iterations changed the free energies of windows at one end of the channel only slowly relative to the windows at the other end. Even in a globally not yet converged state, each window was already well-matched with its neighbors, and iterations had to maintain that matching. The standard WHAM iteration thus faces similar numerical challenges as, say, local update schemes in path integral simulations or in polymer simulations. The use of minimizers successfully addresses this problem by using what amounts to gradient-based multiparticle moves in the simulation analogs.

The coupled nonlinear WHAM equations also make it difficult to directly estimate the statistical errors in the obtained free energies, especially for the free energy difference across multiple umbrella windows. To address this problem, a coarse-grained free energy profile can be calculated by integrating the mean restraining forces corresponding to the free energy gradients. The statistical errors in the coarse-grained free energy can be easily determined from the variances of the average reaction coordinate in each individual window. The coarse-grained free energy is a good approximation for the normalization factors in WHAM and, in the stiff-spring case, the desired PMF itself. On this basis, one can use the error obtained from the coarse free energy profile to estimate the error of the fine WHAM profile. In the WHAM-derived free energy profile, the statistical uncertainty in the difference between two distant positions can thus be readily estimated (Eq. 32), with an explicit expression of the error accumulation over the intermediate umbrella windows. Moreover, in this scheme it is straightforward to decompose the statistical errors into the contributions of each individual window; to further improve the accuracy of the PMF, one may focus on the windows with predominant contributions to the errors and extend the sampling in these windows.

The error estimation above relies on a proper estimate for the variance of the average reaction coordinate in each window, which can only be achieved when the other orthogonal degrees of freedom are also sufficiently sampled. Insufficient equilibration or sampling of the orthogonal coordinates is a common problem in practical applications of free energy methods, and may result in significant underestimation of the errors. This problem typically manifests itself as inconsistencies between neighboring (and more distant) histograms of the

umbrella windows, as different states of the orthogonal coordinates normally correspond to different probability distributions also of the WHAM coordinate x. The inconsistency coefficients introduced in this study can help identify discontinuities in the orthogonal coordinates between neighboring umbrella windows, when data in each individual window alone give no hint of such a problem, as demonstrated in our tests. If inconsistency is detected, one may thoroughly examine all relevant coordinates in the involved windows for potential discontinuity,[24] or simply extend the simulations for a better equilibration. It is therefore advisable to perform these consistency tests in addition to the error estimation in umbrella sampling simulations.

As alternatives to WHAM, some other methods aim to directly determine the normalization factors $\{f_i\}$ or equivalently, the free energy factors $\{g_i\}$, without constructing histograms. The multistate Bennett acceptance ratio (MBAR) method,[16] e.g., obtains the optimal $\{g_i\}$ by solving a set of coupled nonlinear equations, and provides a formula for the statistical uncertainty in the $\{g_i\}$.[16] In this study, we introduce other routes to determine the $\{g_i\}$, by minimizing the function $\widehat{A}$ in Eq. (19) and, more approximately, by using free energy gradients in Eq. (28). It was shown that the MBAR formulism is equivalent to the WHAM equations as the bin width in WHAM approaches zero.[16] We also demonstrate in this study that WHAM and our gradient-based method give very similar results for the same dataset. Given such similarities, we therefore expect the statistical errors estimated by our method and by MBAR to be comparable. However, we have not performed a direct comparison because a full implementation of the MBAR error estimate would require manipulation of matrices of dimensions $n\times k$, where n is the total number of data points summed over all k simulations. At least for one-dimensional umbrella sampling with harmonic bias functions, our error estimation scheme is thus much simpler to implement than that of MBAR, and clearly reveals the contribution of each umbrella window to the accumulated error.

A potential drawback of WHAM is the somewhat artificial choice of the bin width for the histograms and the associated discretization error.[25] The theoretical estimate[25] of such error in our test cases is well below 0.05 k_BT; indeed, we find that when using different bin widths the WHAM results typically differ by less than 0.01 k_BT, much smaller than the magnitude of other errors discussed in this study. However, we note that if the simulations are extended by orders of magnitude, other errors could in principle be dramatically reduced such that the discretization becomes the major error source. In such cases one can simply use a smaller bin width to reduce the error. Nonetheless, WHAM has practical advantages in certain applications. For example, the number of bins is typically much smaller than the number of the original data, and does not increase with the data size. Therefore the reduction of the data into low-dimensional histograms can have significant benefit in the speed and memory requirement when dealing with very large datasets. Also, WHAM directly provides the unbiased free energy profile, which is typically the main objective of umbrella sampling simulations. Furthermore, the histograms permit a more detailed examination of the data. It was noted that an exchange of data between different simulations will not change the MBAR result.[14,16] In contrast, with the histograms one can further check whether the data in each simulation are consistently distributed to detect potential problems, as demonstrated in this study. Overall, in practical applications, major errors are almost always due to imperfect data arising from the finite sampling and the initial conditions. Having proper diagnostics of such errors and indicators of the resulting inconsistencies provides a basis for the development of optimized and refined free energy sampling strategies.

Acknowledgments

We thank Dr. Edina Rosta for helpful discussions regarding Eq. (38). This research was supported by the Intramural Research Programs of the NIDDK, NIH, and utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov).

APPENDIX A

In this appendix, the estimate of the statistical error in Eqs. 1, 32, and 36 is generalized to higher-dimensional umbrella sampling in a space spanned by N order parameters x_α, sampled in S simulations with added harmonic bias potentials $\Sigma_{\alpha=1}^{N} K_\alpha \left(x_\alpha - r_\alpha^i\right)^2/2$ for simulation i. From each simulation, we obtain an estimate of the N components of the free energy gradient from the mean restraining forces, $\bar{F}_\alpha^i = K_\alpha \left(r_\alpha^i - \bar{x}_\alpha\right)$. As in Eq. 28, we want to combine these gradients into an estimate of the relative free energies $G_i = k_B T g_i$ of the restrained simulations. However, in multiple dimensions the numerically estimated free energy gradients can be integrated through multiple paths, the results of which may not match exactly, due to both statistical errors in the estimated gradients and discretization errors of the integration. A path-independent result can be obtained by minimizing the squared deviation of the free energy differences $G_j - G_i$ between adjacent windows i and j from the corresponding averaged gradients,

$$\chi^2 = \sum_{(i,j)} \left[G_j - G_i - \frac{1}{2} \sum_\alpha \left(\bar{F}_\alpha^j + \bar{F}_\alpha^i \right) \left(r_\alpha^j - r_\alpha^i \right) \right]^2 \tag{A1}$$

where the sum is over pairs (i,j) of simulations with nearby restraining centers r_α^i and r_α^j, such that the linear approximation in Eq. A1 is applicable. For restraining points on a rectangular grid, minimization of χ^2 results in a discrete Poisson equation with boundary condition $G_1 = 0$ (where we arbitrarily chose $i = 1$ as reference of the free energies, without loss of generality). In general, setting the derivatives of χ^2 with respect to the G_i to zero results in a set of linear equations that can be written in matrix form as $\mathbf{MG} = \mathbf{BF}$, where $\mathbf{M}$ amounts to a discrete version of the Laplace operator, $\mathbf{G}$ is the vector of (unknown) free energies G_i, and $\mathbf{B}$ is a matrix that produces the appropriate linear combination of the restraining forces $\bar{F}_\alpha^i$ forming the vector $\mathbf{F}$, consistent with Eq. A1. The formal solution then is $\mathbf{G} = \mathbf{M}^{-1}\mathbf{BF}$. From the errors in the mean positions of the order parameters, we can again estimate the uncertainties in the corresponding restraining forces, $\mathrm{var}\left(\bar{F}_\alpha^i\right) = K_\alpha^2 \mathrm{var}_i\left(\bar{x}_\alpha\right)$, and more generally estimate their covariances $\mathrm{cov}\left(\bar{F}_\alpha^i, \bar{F}_\gamma^i\right) = K_\alpha K_\gamma \mathrm{cov}_i\left(\bar{x}_\alpha, \bar{x}_\gamma\right)$. The covariances of the estimated free energies are then obtained as linear combinations of the covariances in the restraining forces,

$$\mathrm{cov}\left(G_i, G_j\right) = \sum_{k=1}^{S} \sum_{\gamma=1}^{N} \sum_{\alpha=1}^{N} c_{ik\alpha} c_{ij\gamma}\, \mathrm{cov}\left(\bar{F}_\alpha^k, \bar{F}_\gamma^k\right), \tag{A2}$$

where $c_{ik\alpha} = (\mathbf{M}^{-1}\mathbf{B})_{i,k\alpha}$. Here we assumed that there are no correlations between the results of different simulations i. If we assume further that the restraining forces are uncorrelated, we obtain an estimate of the uncertainty in the free energies,

$$\mathrm{var}\,(G_i) = \sum_{k=1}^{S}\sum_{\alpha=1}^{N} c_{ik\alpha}^2\,\mathrm{var}\left(\bar{F}_\alpha^k\right). \tag{A3}$$

In practice, it may be simpler to circumvent the matrix computations and instead calculate $c_{ik\alpha}$ directly by repeated minimization of the quadratic target function χ^2, in what amounts to a variational construction of the Green's function $\mathbf{M}^{-1}$. Specifically, if all $\bar{F}_\alpha^i$ are set to zero in Eq. A1, except $\bar{F}_\gamma^k$, then numerical minimization of the resulting χ^2 with respect to the G_i (as above with the reference G_1=0) directly produces the required solutions $c_{ik\gamma} = G_i$. By repeated minimization of χ^2 for all k and γ, one can thus build up the entire set of coefficients entering the error formulas Eqs. A2 and A3. We note that for one-dimensional umbrella sampling, with the (i,j) sum in Eq. A1 restricted to nearest neighbors, one recovers the results of the main text, in particular Eqs. 32 and 36.

APPENDIX B

In this appendix we show that our analysis of the statistical errors can be similarly applied to umbrella sampling with more general biasing potentials $w(x,r)$, such as those with nonuniform spring constants, or anharmonic potentials. In the general case we define the coarse-grained free energy

$$G_r(r) = -k_B T \ln \int p(x) \exp\left[\frac{-w(x,r)}{k_B T}\right] dx, \tag{A4}$$

whose derivative is

$$\frac{\partial}{\partial r} G_r(r) = \frac{\int \left[\frac{\partial}{\partial r} w(x,r)\right] p(x) \exp\left[\frac{-w(x,r)}{k_B T}\right] dx}{\int p(x) \exp\left[\frac{-w(x,r)}{k_B T}\right] dx}. \tag{A5}$$

We define

$$F(x,r) \equiv \frac{\partial}{\partial r} w(x,r). \tag{A6}$$

The derivative of $G_r(r)$ at r_i is then given by

$$\frac{\partial}{\partial r} G_r(r_i) = \langle F(x,r_i)\rangle_i \approx \bar{F}_i, \tag{A7}$$

in which $\bar{F}_i$ is obtained by averaging $F(x,r_i)$ in simulation i with biasing potential $w(x,r_i)$. Then the difference of $G_r(r)$ between any two windows is again given approximately by Eq. (28), with the statistical uncertainty accordingly determined from the estimated variances of the $\left\{\bar{F}_i\right\}$.

For the harmonic biasing potentials $w(x,r) = K(x - r)^2 / 2$ discussed in the main text, we have $F = \frac{\partial}{\partial r} w(x, r) = -\frac{\partial}{\partial x} w(x, r)$, and $\bar{F}_i$ thus coincides with the average restraining force of the biasing potential.

REFERENCES

1. Torrie GM, Valleau JP. Chem Phys Lett. 1974; 28(4):578–581.
2. Ferrenberg AM, Swendsen RH. Phys Rev Lett. 1989; 63(12):1195–1198. [PubMed: 10040500]
3. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. J Comput Chem. 1992; 13(8): 1011–1021.
4. Boczko EM, Brooks CL 3rd. Science. 1995; 269(5222):393–396. [PubMed: 7618103]
5. Hub JS, de Groot BL, van der Spoel D. J Chem Theory Comput. 2010; 6(12):3713–3720.
6. Gallicchio E, Andrec M, Felts AK, Levy RM. J Phys Chem B. 2005; 109(14):6722–6731. [PubMed: 16851756]
7. Chodera JD, Swope WC, Pitera JW, Seok C, Dill KA. J Chem Theory Comput. 2007; 3(1):26–41.
8. Rosta E, Nowotny M, Yang W, Hummer G. J Am Chem Soc. 2011; 133(23):8934–8941. [PubMed: 21539371]
9. Allen TW, Andersen OS, Roux B. Biophys J. 2006; 90(10):3447–3468. [PubMed: 16500984]
10. Bartels C, Karplus M. J Comput Chem. 1997; 18(12):1450–1462.
11. Heath, MT. Scientific Computing: An Introductory Survey. McGraw-Hill; New York: 2002.
12. Flyvbjerg H, Petersen HG. J Chem Phys. 1989; 91(1):461–466.
13. Pollard, D. NSF-CBMS Regional Conference Series in Probability and Statistics. 1990.
14. Kong A, McCullagh P, Meng XL, Nicolae D, Tan Z. J Roy Stat Soc Ser B (Stat Method). 2003; 65:585–604.
15. Tan ZQ. J Amer Statistical Assoc. 2004; 99(468):1027–1036.
16. Shirts MR, Chodera JD. J Chem Phys. 2008; 129(12):124105. [PubMed: 19045004]
17. MATLAB. MathWorks
18. Hummer G, Szabo A. Acc Chem Res. 2005; 38(7):504–513. [PubMed: 16028884]
19. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G. J Chem Phys. 2006; 125(2):24106. [PubMed: 16848576]
20. Fukunishi H, Watanabe O, Takada S. J Chem Phys. 2002; 116(20):9058–9067.
21. Park S, Schulten K. J Chem Phys. 2004; 120(13):5946–5961. [PubMed: 15267476]
22. Zhu F, Hummer G. Proc Natl Acad Sci U S A. 2010; 107(46):19814–19819. [PubMed: 21041674]
23. Bereau T, Swendsen RH. J Comput Phys. 2009; 228(17):6119–6129.
24. Rosta E, Woodcock HL, Brooks BR, Hummer G. J Comput Chem. 2009; 30(11):1634–1641. [PubMed: 19462398]
25. Kobrak MN. J Comput Chem. 2003; 24(12):1437–1446. [PubMed: 12868109]

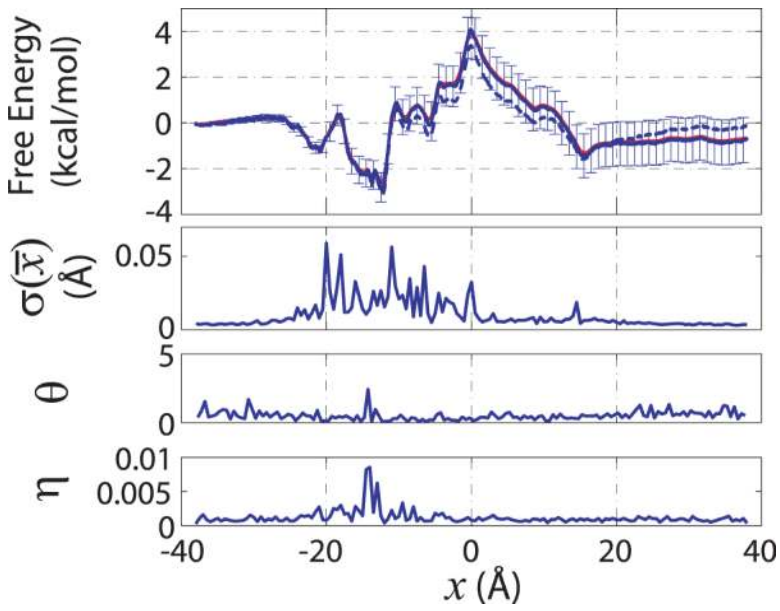

Figure 1.

Umbrella sampling of a Na^+ ion in an ion channel. The chosen reaction coordinate is the z -coordinate of the ion, but is denoted by x here, as used consistently throughout this article. (**A**) Free energy profiles calculated from the WHAM equations (*blue solid* curve) and from the integration of the mean forces (*red* curve, which is largely overlapping with the blue curve). The *blue dashed* curve represents the result from the direct iteration calculation with a convergence threshold of $\delta = 10^{-3}$ and uniform initial values, as explained in Table 1 and in the text. All curves are vertically aligned to have a value of zero at $x = -38$ Å. The plotted error bars are for the free energy difference with respect to the first umbrella window at $x = -38$ Å, as determined according to Eq. (32). (**B**) Standard deviation $\sigma\left(\bar{x}_i\right) = \left[\mathrm{var}\left(\bar{x}_i\right)\right]^{1/2}$ in the average position $\bar{x}_i$ in each window, estimated from the block averaging method[12] as described in the text. (**C**) Inconsistency coefficient π, defined in Eq. (40), between every pair of adjacent umbrella windows. (**D**) Relative entropy η, defined in Eq. (41), between the consensus and the observed histograms.

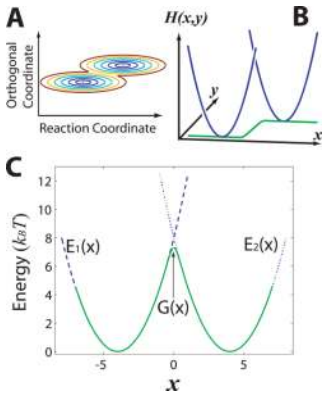

Figure 2.

Model free energy surface. (**A**) Schematic illustration of a typical situation in free energy calculations, with degrees of freedom orthogonal to the reaction coordinate exhibiting multiple local minima. (**B**) Realization of the scheme in (A) with a simple quantitative model, as defined in Eq. (43), with continuous motion in the *x* direction and discrete hopping in the *y* direction between two 1D surfaces (*blue*). The *green* curve plots the ensemble average of *y* for a given *x*. (**C**) The potential energies for the two states defined in Eq. (44), and the corresponding free energy (*green* curve) calculated from Eq. (45).

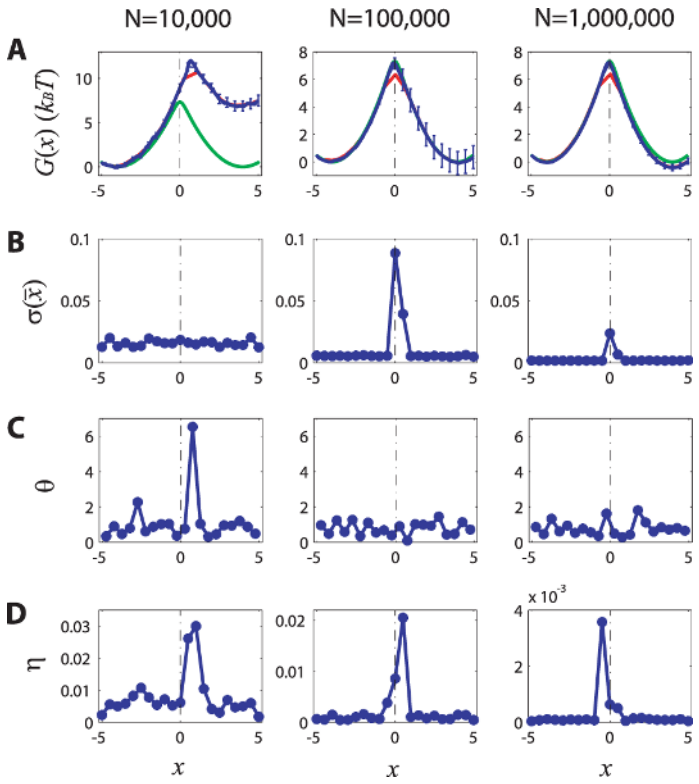

Figure 3.

Umbrella sampling MC simulations for different sampling sizes, plotted in a manner similar to Fig. 1. The three columns show results for simulations with N =10,000, N =100,000, and N =1,000,000 MC steps, respectively. (**A**) The free energy profiles calculated from the WHAM equations (*blue* curves) and from the integration of the mean forces (*red* curves, Eq. 31), and the ideal (analytical) free energy profile (*green* curves, as in Fig. 2C). The free energy curves are vertically aligned to have the same value at x =−5. The plotted error bars are for the free energy difference with respect to the first umbrella window at x =−5. (**B**) The standard deviation in the average position $\bar{x}_i$ in each window, estimated from the block averaging method.[12] (**C**) The inconsistency coefficients π for η all pairs of adjacent umbrella windows. (**D**) The relative entropy between the consensus and the observed histograms.

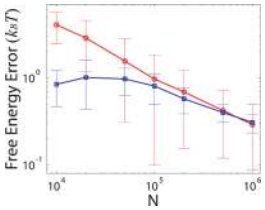

Figure 4.

Comparison of the estimated errors and the absolute errors, with respect to the free energy difference between the two minima at $x = \pm 4$ in the model potential. The umbrella simulations with 1,000,000 MC steps are repeated 100 times, and the first N data from each trajectory, with N ranging from 10,000 to 1,000,000, are used for analysis. From each dataset, the estimated standard deviation calculated from Eq. (32) and the absolute error (with respect to the ideal value, 0) are calculated. The two curves represent the mean (with the standard deviation shown as error bars) of the estimated (*blue squares*) and absolute (*red circles*) errors from the 100 datasets.

Table 1

Performance benchmarks for the direct iteration, trust region, and BFGS numerical algorithms. Each algorithm was tested with two sets of initial values, from a uniform assignment or from a gradient-based estimation of the free energy (Eqs. 24 and 31), respectively, as explained in the text. In addition, the direct iteration algorithm was tested with a convergence threshold (δ) of 10^{-3} or 10^{-6}, respectively. For each test, we list the calculation time (in seconds) and the iteration count (n_{iter}), represented by the number of WHAM iterations (using Eqs. 7 and 15) for the direct iteration algorithm or the number of evaluations of the target function A and its gradients (Eqs. 15–22) for the trust region and BFGS algorithms. To quantify the accuracy of the algorithms, the target function A is calculated for each result, and the minimum $A_{\min}$ (5.2×10^{8}) is identified. The relative differences $\Delta A = A - A_{\min}$ are then listed. Moreover, for each obtained free energy profile $G(x)$, the deviation $\Delta G = G(x) - G_0(x)$ from the best result $G_0(x)$ (corresponding to $A_{\min}$) is calculated, and $\Delta G_{\max} = \max[\Delta G(x)] - \min[\Delta G(x)]$, the maximum deviation in the relative free energy, is listed in the table. The algorithms are implemented in Matlab,[17] and all tests were run on a 2.2 GHz AMD Opteron processor.

			Time	n_{iter}	ΔA	$\Delta G_{\max}$ (K_BT)
Uniform Initial Values	Direct Iteration	$\delta = 10^{-3}$	20 s	1,316	1.5×10^{4}	2.2
		$\delta = 10^{-6}$	532 s	36,117	0.28	0.017
	Trust Region		7.4 s	1,071	3.4×10^{-7}	1.1×10^{-6}
	BFGS		1.4 s	11	0	0
Estimated Initial Values	Direct Iteration	$\delta = 10^{-3}$	0.68 s	48	197	0.14
		$\delta = 10^{-6}$	217 s	16,000	0.27	0.017
	Trust Region		8.3 s	1,377	3.4×10^{-7}	1.1×10^{-6}
	BFGS		1.4 s	10	6.8×10^{-7}	2.0×10^{-6}