

## Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations

John D. Chodera,<sup>\*,†</sup> William C. Swope,<sup>‡</sup> Jed W. Pitera,<sup>‡</sup> Chaok Seok,<sup>§</sup> and Ken A. Dill<sup>||</sup>

Graduate Group in Biophysics and Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143, IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120, and Department of Chemistry, College of Natural Sciences, Seoul National University, Gwanak-gu, Shillim-dong, san 56-1 Seoul 151-747, Republic of Korea

Received November 21, 2005

**Abstract:** The growing adoption of generalized-ensemble algorithms for biomolecular simulation has resulted in a resurgence in the use of the weighted histogram analysis method (WHAM) to make use of all data generated by these simulations. Unfortunately, the original presentation of WHAM by Kumar et al. is not directly applicable to data generated by these methods. WHAM was originally formulated to combine data from independent samplings of the *canonical* ensemble, whereas many generalized-ensemble algorithms sample from mixtures of canonical ensembles at different temperatures. Sorting configurations generated from a parallel tempering simulation by temperature obscures the temporal correlation in the data and results in an improper treatment of the statistical uncertainties used in constructing the estimate of the density of states. Here we present variants of WHAM, STWHAM and PTWHAM, derived with the same set of assumptions, that can be directly applied to several generalized ensemble algorithms, including simulated tempering, parallel tempering (better known as replica-exchange among temperatures), and replica-exchange simulated tempering. We present methods that explicitly capture the considerable temporal correlation in sequentially generated configurations using autocorrelation analysis. This allows estimation of the statistical uncertainty in WHAM estimates of expectations for the canonical ensemble. We test the method with a one-dimensional model system and then apply it to the estimation of potentials of mean force from parallel tempering simulations of the alanine dipeptide in both implicit and explicit solvent.

### 1. Introduction

The difficulty of computing equilibrium averages for complex systems such as solvated biopolymers by Monte Carlo

or molecular dynamics simulation is well-known. Numerous minima and large free-energy barriers tend to slow exploration in phase space and trap the simulation in metastable regions of configuration space. This hampers the ability of the system both to equilibrate (reach the thermodynamically relevant region of phase space) and to sample sufficiently for estimates of ensemble averages to converge (reduce the statistical uncertainty in the estimate to an acceptable level) in finite computer time.

The emergence of a new class of simulation algorithms, termed *generalized-ensemble* algorithms,<sup>1</sup> has helped to

\* Corresponding author e-mail: jchodera@gmail.com.

† Graduate Group in Biophysics, University of California at San Francisco.

‡ IBM Almaden Research Center.

§ Seoul National University.

|| Department of Pharmaceutical Chemistry, University of California at San Francisco.

mitigate these problems. In a generalized-ensemble simulation, the probability distribution from which conformations are sampled is altered from a canonical distribution to one that will induce a broader sampling of the potential energy. Proper application should in principle allow the system to overcome energetic barriers and sample configuration space more thoroughly, at the expense of spending more time in high-energy regions that may be irrelevant at the temperature of interest. The particular method by which sampling is enhanced depends on the algorithm. In the *multicanonical algorithm* (MUCA),<sup>5–9</sup> conformations are sampled with a probability proportional to an approximation of the inverse potential energy density of states in an attempt to produce a random walk in the potential energy. In *simulated tempering* (ST),<sup>10,11,3</sup> a random walk between canonical ensembles at different temperatures is used to produce a random walk in energy, but an estimate of the free energy as a function of temperature is needed as input to ensure equal visitation of all temperatures. *Parallel tempering* (PT), a special case of the replica-exchange method (REM),<sup>12,4</sup> eliminates the need to know these free energies a priori by coupling temperature changes between pairs of a pool of simulated tempering simulations conducted in parallel. Several other algorithms and combinations thereof have also been proposed.<sup>3,13–15</sup>

In several of these algorithms, such as simulated tempering and parallel tempering, each replica generates configurations from a *mixed-canonical* distribution (a term coined in ref 16)—that is, a number of configurations are generated from the canonical distribution at each of several temperatures. To compute expectations over the canonical ensemble at a single temperature, either the configurations from all replicas that visit the temperature of interest must be collected and the remainder discarded (as in ref 17) or else a reweighting scheme must be used to properly weight the data generated at other temperatures. Fortunately, the weighted histogram analysis method (WHAM),<sup>2</sup> an extension of the single- and multiple-histogram methods introduced by Ferrenberg and Swendsen,<sup>18,19</sup> allows configurations generated from independent canonical simulations at different temperatures to be reweighted to compute expectations from the canonical ensemble at any temperature of interest. Okamoto and co-workers have applied this method to both replica-exchange simulated tempering (REST)<sup>3</sup> and parallel tempering<sup>4</sup> methods by reordering sampled configurations into pseudotrajectories, grouping configurations generated at a particular temperature together regardless of which replica they came from. Unfortunately, this permutation obscures the correlation among the stored configurations, causing the apparent correlation times for each pseudotrajectory to appear artificially shorter than the true correlation times within the independent replica trajectories. The permutation also introduces correlation *between* the pseudotrajectories, which is problematic because WHAM as presented in ref 2 is constructed to operate on *independent* canonical trajectories. Additionally, it is difficult to estimate the statistical uncertainty in the resulting estimate of the expectation from these pseudotrajectories, since standard autocorrelation analysis techniques<sup>20–23</sup> can no longer be applied. Recently, Gallicchio

et al.<sup>24</sup> have described a new method for computing expectations and uncertainties from canonical simulations at different temperatures based on Bayesian inference. While Bayesian approaches are usually superior to those based on first-order Taylor expansion methods for the propagation of uncertainties (of the sort we describe in this work), they are less suitable for treating highly correlated measurements where the functional form of the correlation is essentially unknown.

Here, we derive variants of WHAM that operate on replica trajectories that are not reordered or collected by temperature. It should be noted that even if simulation data have been stored to disk sorted by temperature, they can be permuted back to the original replica trajectories to perform the proposed analyses if information about the replica-to-temperature mapping or swapping was stored. Our presentation takes a careful approach to the correlation times involved, and we show under which conditions the almost universally omitted statistical inefficiency term that appears in all formulations of WHAM-like methods can be properly neglected. Finally, we show how the statistical uncertainty in the estimator for the configuration space average for some observable can be estimated by considering the effect of temporal correlation. The method is simple and inexpensive enough to employ in all cases where WHAM is used, and we hope all researchers using WHAM will report these statistical uncertainties in the future to assess both the significance and the degree of reproducibility of results from simulations.

This paper is organized as follows: In section 2, we present a derivation of the Kumar et al. WHAM for independent simulations sampling from the canonical ensemble. Careful attention is paid to the proper treatment of time correlation in estimating the statistical uncertainty in the histograms and the resulting estimator for the expectation, and a novel way of obtaining estimates for multiple observables is presented. In section 3, we derive analogues of the method for treating simulated and parallel tempering simulations, STWHAM and PTWHAM, while properly capturing the correlations among sequential configurations. In section 4, we validate our uncertainty estimates in a one-dimensional model system and demonstrate an application for biomolecular systems by estimating the potential of mean force and corresponding uncertainties from parallel tempering simulations of alanine dipeptide in implicit and explicit solvent. An illustrative efficient implementation of the method in Fortran 95 for use in the analysis of simulated and parallel tempering simulations can be found in the Supporting Information.

## 2. Independent Canonical Simulations

In this section, we review the derivation of WHAM for computing expectations from multiple independent simulations in the canonical ensemble. Conducting independent simulations at the same or different temperatures can reduce statistical uncertainty while obtaining perfect parallelism (after the initial time to reach equilibrium has been discarded). Some of these simulations might be conducted at a higher temperature than the temperature of interest to

promote greater sampling across barriers, for example. Sometimes, the expectation value of one or more observables is desired over a range of temperatures. Additionally, simulations started from different initial conditions can be used as a check of equilibration and convergence.<sup>25</sup> Below, we follow roughly the same approach as Kumar et al.<sup>2</sup> in deriving the WHAM equations, though our notation differs substantially, and we include a more detailed treatment of statistical uncertainty. Additionally, we arrive at a novel way of computing expectations of multiple observables and avoid the use of many-dimensional histograms. While the method presented in ref 2 has the full generality of treating simulations conducted with arbitrary biasing potentials, we focus on the case of independent canonical simulations at different temperatures, since variations on this approach will allow us to consider simulated and parallel tempering simulations in section 3. (For an informative treatment of the case of a multiple biasing potentials at a single temperature, as in the case of umbrella sampling, see ref 26.)

**2.1. Motivation and Definitions.** Suppose we have an observable  $A$  that is only a function of the Cartesian coordinates of the system  $\mathbf{q}$ , and we wish to estimate the expectation of  $A$  over the canonical ensemble at some temperature of interest  $T$ . Instead of this temperature  $T$ , we will generally refer to its corresponding inverse temperature  $\beta = (k_B T)^{-1}$ , where  $k_B$  is the Boltzmann constant. We denote the expectation of  $A$  over the canonical ensemble at inverse temperature  $\beta$  by  $\langle A \rangle_\beta$ , which can be written as

$$\langle A \rangle_\beta = \frac{\int d\mathbf{q} e^{-\beta U(\mathbf{q})} A(\mathbf{q})}{\int d\mathbf{q} e^{-\beta U(\mathbf{q})}} \quad (1)$$

where  $U(\mathbf{q})$  is the potential energy function of the system.

Further suppose we have carried out  $K$  independent simulations that sample from the canonical ensemble (using such techniques as Metropolis Monte Carlo<sup>33</sup> or thermally controlled molecular dynamics) at corresponding inverse temperatures  $\beta_1, \beta_2, \dots, \beta_K$ , some or all of which may be different from the temperature of interest. We denote the coordinates and potential energies sampled at a fixed time interval  $\Delta t$  from simulation  $k$  by the time series  $\{\mathbf{q}_{kn}, U_{kn}\}_{n=1}^{N_k}$ , where  $U_{kn} = U(\mathbf{q}_{kn})$  and  $N_k$  is the number of configurations collected from simulation  $k$ .

We first consider the probability density function from which the configurations are generated in simulation  $k$ . For a simulation sampling from the canonical distribution, the probability of generating a configuration with potential energy in the interval  $dU$  about  $U$  at inverse temperature  $\beta$  is given by

$$p(U|\beta) dU = [Z(\beta)]^{-1} \Omega(U) dU e^{-\beta U} \quad (2)$$

with the normalizing constant  $Z(\beta)$ , often referred to as the *configurational partition function*, chosen to ensure that  $p(U)$  integrates to unity. The quantity  $\Omega(U)$  is the *potential energy density of states*, and  $\Omega(U) dU$  represents the volume of configuration space with potential energy in the interval  $dU$  around  $U$ .

While the Boltzmann factor  $e^{-\beta_k U}$  and normalization constant  $Z(\beta_k)$  differ for each simulation  $k$ , the density of states  $\Omega(U)$  is independent of temperature. Since the Boltzmann factor is a known function and the configurational partition function is simply a normalizing constant, knowledge of the density of states allows the potential energy probability density to be computed at *any* temperature. If the average of the observable  $A$  over all configurations with potential energy  $U$  is known, these can be combined to give the expectation at a desired inverse temperature  $\beta$

$$\langle A \rangle_\beta = \frac{\int dU \Omega(U) e^{-\beta U} A(U)}{\int dU \Omega(U) e^{-\beta U}} \quad (3)$$

where  $A(U)$  is defined as the average of  $A$  over all configurations with potential energy  $U$

$$A(U) \equiv \frac{\int d\mathbf{q} \delta(U(\mathbf{q}) - U) A(\mathbf{q})}{\int d\mathbf{q} \delta(U(\mathbf{q}) - U)} \quad (4)$$

It is easily seen that substituting this expression into eq 3 recovers the configuration space average in eq 1.

Our aim is to obtain the best estimate of the density of states and the expectation of the observable by combining information from several simulations. Since each simulation samples an energy range determined by its temperature, our final estimate of the density of states will be more accurate if we account for the different uncertainties in the estimate obtained from each simulation. We will therefore need a separate estimate of the density of states and its corresponding uncertainty from *each* simulation.

**2.2. Obtaining an Estimate of the Density of States from Each Simulation.** To obtain an estimate of the density of states from each simulation, we first need a way of mathematically expressing the form of the observed probability density function  $p(U)$ . While it may be possible to assume a particular functional form for this density, this would generally be inexact. A better approach is to use a *nonparametric density estimator* (see, for example, ref 27 for an overview) that makes no prior assumptions as to the true functional form of  $p(U)$ . Kumar et al.,<sup>2</sup> as Ferrenberg and Swendsen<sup>18,19</sup> earlier, chose a histogram-based estimator, in which the range of sampled energies is discretized into a set of nonoverlapping bins of equal width. While there are a number of more sophisticated smooth nonparametric estimators,<sup>28</sup> the histogram estimator is simpler and more efficient to apply.

Accordingly, we construct an estimate of the probability density function  $p(U|\beta)$  on a set of  $M$  points, labeled  $U_m$ , that span the sampled potential energy range and are spaced  $\Delta U$  apart. We denote the estimate of  $p(U|\beta)$  at  $U_m$  by  $p_m(\beta)$  and the corresponding estimate of the density of states  $\Omega(U_m)$  by  $\Omega_m$ .

$$p_m(\beta) \equiv p(U_m|\beta) = [Z(\beta)]^{-1} \Omega_m e^{-\beta U_m} \quad (5)$$

The normalization factor  $Z(\beta)$  can then be approximated by a discretized integration

$$Z(\beta) = \int dU \Omega(U) e^{-\beta U} \approx \sum_{m=1}^M \Delta U \Omega_m e^{-\beta U_m} \quad (6)$$

where we have made the assumption that the integrand,  $\Omega(U) e^{-\beta U}$ , does not change significantly over the bin width  $\Delta U$ . As the density of states  $\Omega(U)$  increases with  $U$  and the Boltzmann factor  $e^{-\beta U}$  decreases, their product is expected to vary less rapidly than either term individually.

We define  $\psi_m(U)$  as the indicator or characteristic function for the energy bin of width  $\Delta U$  centered about  $U_m$

$$\psi_m(U) = \begin{cases} 1 & \text{if } U \in [U_m - \Delta U/2, U_m + \Delta U/2) \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

and the time series defined by this indicator function as  $\{\psi_{mkn}\}_{n=1}^{N_k}$ , where  $\psi_{mkn} \equiv \psi_m(U_{kn})$ . We denote the count of configurations from simulation  $k$  that fall in energy bin  $m$ —the “histogram” from which the weighted histogram analysis method derives its name—by  $H_{mk}$ , and see that it can be computed by

$$H_{mk} = \sum_{n=1}^{N_k} \psi_{mkn} \quad (8)$$

We will also use the total number of configurations over all simulations that fall in energy bin  $m$ , which we term  $H_m$ :

$$H_m = \sum_{k=1}^K \sum_{n=1}^{N_k} \psi_{mkn} \quad (9)$$

Note that throughout our discussion, pairs of variables that only differ by the number of written subscripts, such as  $H_m$  and  $H_{mk}$ , represent similar quantities related in this way.

We can estimate  $p_m(\beta_k)$  by the number of configurations sampled from the simulation at temperature  $\beta_k$  with energies that fall in the bin centered about  $U_m$ :

$$p_m(\beta_k) \approx \frac{1}{\Delta U} \cdot \frac{H_{mk}}{N_k} \quad (10)$$

Equating this with the definition of  $p_m$  from eq 5 and rearranging terms, we can obtain an estimate of  $\hat{\Omega}_{mk}$ , the density of states at energy  $U_m$  from simulation  $k$ , which we will denote by  $\hat{\Omega}_{mk}$ :

$$\begin{aligned} \hat{\Omega}_{mk} &= \frac{1}{\Delta U} \cdot \frac{H_{mk}}{N_k} \cdot \frac{Z(\beta_k)}{e^{-\beta_k U_m}} \\ &= \frac{H_{mk}}{N_k \Delta U \exp[f_k - \beta_k U_m]} \end{aligned} \quad (11)$$

In the last step, we have replaced the partition function  $Z(\beta_k)$  by an exponentiated dimensionless free energy  $f_k \equiv -\ln Z(\beta_k)$ . Each independent simulation  $k$  contributes an estimate of the density of states  $\hat{\Omega}_{mk}$  for energy bin  $m$ . Each of these estimates in turn carries a statistical uncertainty  $\delta^2 \hat{\Omega}_{mk}$ , determined primarily by the number of uncorrelated samples of the energy bin. (Expressions for  $\delta^2 \hat{\Omega}_{mk}$  will be derived later in section 2.5.) We will combine these

individual estimates  $\hat{\Omega}_{mk}$  to produce a single optimal estimator  $\hat{\Omega}_m$  in a such way that the statistical uncertainty in the resulting estimate is minimized, giving more weight to the  $\hat{\Omega}_{mk}$  with smaller uncertainties. To do this, we must first briefly review the maximum-likelihood method for combining independent measurements with associated uncertainties into an optimal estimate and also consider the uncertainty in a mean computed from a set of correlated observations.

**2.3. Optimal Estimator from Independent Observations and Associated Uncertainties.** Suppose we have  $K$  independent observations or measurements of some random variable  $X$  denoted  $x_1, \dots, x_K$ , each with corresponding squared uncertainty  $\delta^2 x_k$ , defined by

$$\delta^2 x_k \equiv \langle (x_k - \langle x_k \rangle)^2 \rangle = \langle x_k^2 \rangle - \langle x_k \rangle^2 \quad (12)$$

where  $\langle \cdot \rangle$  here denotes the expectation over repeated measurements or experimental trials. We can then write  $\hat{X}$ , the optimal estimator for  $\langle X \rangle$  in the sense of minimizing  $\delta^2 \hat{X}$ , by a weighted sum of the individual estimates

$$\hat{X} = \frac{\sum_{k=1}^K [\delta^2 x_k]^{-1} x_k}{\sum_{k=1}^K [\delta^2 x_k]^{-1}} \quad (13)$$

Note that observations with smaller uncertainties get greater weight, and if all the uncertainties are equal, the weight is simply  $1/K$ , as would be expected.

The uncertainty in the resulting estimate is simply given by

$$\delta^2 \hat{X} = \left\{ \sum_{k=1}^K [\delta^2 x_k]^{-1} \right\}^{-1} \quad (14)$$

These are standard formulas that come from maximum likelihood considerations.<sup>29</sup>

**2.4. Statistical Uncertainty in the Estimator for Correlated Time Series Data.** We briefly review the estimation of statistical uncertainty for a time series of correlated measurements. (See ref 20 for an early exposition of this method as applied to the analysis of Monte Carlo simulations of spin systems, ref 21 for the analysis of molecular dynamics simulations, or ref 23 for a recent general illustration.)

Suppose we have a time series of correlated sequential observations of the random variable  $X$  denoted  $\{x_n\}_{n=1}^N$  that come from a stationary, time-reversible stochastic process. Our estimate for the expectation of  $X$  is given by the time average

$$\hat{X} = \frac{1}{N} \sum_{n=1}^N x_n \quad (15)$$

but the statistical uncertainty is more complicated than in the independent observation case

$$\delta^2 \hat{X} \equiv \langle (\hat{X} - \langle \hat{X} \rangle)^2 \rangle = \langle \hat{X}^2 \rangle - \langle \hat{X} \rangle^2$$

$$\begin{aligned}
&= \frac{1}{N^2} \sum_{n,n'=1}^N [\langle x_n x_{n'} \rangle - \langle x_n \rangle \langle x_{n'} \rangle] \\
&= \frac{1}{N^2} \sum_{n=1}^N [\langle x_n^2 \rangle - \langle x_n \rangle^2] + \frac{1}{N^2} \sum_{n \neq n'=1}^N [\langle x_n x_{n'} \rangle - \langle x_n \rangle \langle x_{n'} \rangle]
\end{aligned} \quad (16)$$

In the last step, we have split the sum into two sums—a term capturing the variance in the observations and a remaining term capturing the correlation between observations. Using the properties of stationarity and time-reversibility, we can further manipulate this to obtain

$$\begin{aligned}
\delta^2 \hat{X} &= \frac{1}{N} [\langle x_n^2 \rangle - \langle x_n \rangle^2] + \frac{2}{N} \sum_{t=1}^{N-1} \left( \frac{N-t}{N} \right) [\langle x_n x_{n+t} \rangle - \langle x_n \rangle \langle x_{n+t} \rangle] \\
&\equiv \frac{\sigma_x^2}{N} (1 + 2\tau) = \frac{\sigma_x^2}{N/g}
\end{aligned} \quad (17)$$

where the variance  $\sigma_x^2$ , statistical inefficiency  $g$ , and autocorrelation time  $\tau$  (in units of the sampling interval) are given by

$$\sigma_x^2 \equiv \langle x_n^2 \rangle - \langle x_n \rangle^2 \quad (18)$$

$$\tau \equiv \sum_{t=1}^{N-1} \left( 1 - \frac{t}{N} \right) C_t \quad (19)$$

$$g \equiv 1 + 2\tau \quad (20)$$

with the discrete-time normalized fluctuation autocorrelation function  $C_t$  defined as

$$C_t \equiv \frac{\langle x_n x_{n+t} \rangle - \langle x_n \rangle^2}{\langle x_n^2 \rangle - \langle x_n \rangle^2} \quad (21)$$

The quantity  $g \equiv (1 + 2\tau) \geq 1$  can be thought of as a *statistical inefficiency*, in that  $N/g$  gives the effective number of *uncorrelated* configurations contained in the time series. The statistical inefficiency will depend on the time interval at which configurations are collected for analysis; longer intervals will reduce the statistical inefficiency, which will approach unity as the sampling interval exceeds the correlation time. Practically, we use our best estimates for the variance  $\sigma_x^2$  and autocorrelation function  $C_t$  to compute an estimate of the statistical uncertainty  $\delta^2 \hat{X}$ .

**2.5. Optimal Estimate of the Density of States.** We now construct an optimal estimator of the density of states  $\Omega_m$  from the individual estimates obtained from the  $K$  independent canonical simulations. From the results of section 2.3, we can write this estimator and its corresponding uncertainty as

$$\hat{\Omega}_m = \frac{\sum_{k=1}^K [\delta^2 \hat{\Omega}_{mk}]^{-1} \hat{\Omega}_{mk}}{\sum_{k=1}^K [\delta^2 \hat{\Omega}_{mk}]^{-1}} \quad (22)$$

$$\delta^2 \hat{\Omega}_m = \left\{ \sum_{k=1}^K [\delta^2 \hat{\Omega}_{mk}]^{-1} \right\}^{-1} \quad (23)$$

The results of section 2.4 show us how to write the  $\delta^2 \hat{\Omega}_{mk}$ , the uncertainty in our estimate of the density of states for energy bin  $m$  from simulation  $k$ . In eq 11 above, we see that this uncertainty comes only from  $\delta^2 H_{mk}$ , the uncertainty in the histogram count for the energy bin, since all other terms are known with certainty:

$$\delta^2 \hat{\Omega}_{mk} = \frac{\delta^2 H_{mk}}{\{N_k \Delta U \exp[f_k - \beta_k U_m]\}^2} \quad (24)$$

$H_{mk}$ , the histogram count from simulation  $k$ , can be written as a time average of the indicator function  $\psi_m$  over the correlated configurations collected from the simulation:

$$H_{mk} = N_k \cdot \frac{1}{N_k} \sum_{n=1}^{N_k} \psi_{mkn} \quad (25)$$

We can use the result of section 2.4 above to obtain an expression for  $\delta^2 H_{mk}$ , the uncertainty in the histogram count:

$$\begin{aligned}
\delta^2 H_{mk} &= N_k^2 \cdot \frac{\sigma_{mk}^2}{N_k/g_{mk}} \\
&= g_{mk} N_k (\langle \psi_{mk}^2 \rangle - \langle \psi_{mk} \rangle^2) \\
&= g_{mk} N_k \langle \psi_{mk} \rangle (1 - \langle \psi_{mk} \rangle) \\
&= g_{mk} \langle H_{mk} \rangle \left( 1 - \frac{\langle H_{mk} \rangle}{N_k} \right)
\end{aligned} \quad (26)$$

where, because  $\psi_m(U)$  is an indicator function (eq 7),  $[\psi_m(U)]^2 = \psi_m(U)$ . If the histograms are sparsely populated, a reasonable assumption if there are a sufficient number of histogram bins spanning the energy range sampled by each simulation, then  $\langle H_{mk} \rangle / N_k \ll 1$ , and we can further simplify this to

$$\delta^2 H_{mk} \approx g_{mk} \langle H_{mk} \rangle \quad (27)$$

The statistical inefficiency  $g_{mk}$  here reflects the number of configurations required for an uncorrelated sampling of the energy bin. This will, in general, depend on the bin index, bin width, and temperature. This dependence was omitted in the original Kumar et al. presentation.<sup>2</sup> At higher temperatures, the correlation time, and hence the statistical inefficiency, is expected to be smaller as the simulation can move through configuration space more easily. The structure of the energy landscape may cause the simulation to be stuck in certain regions of configuration space for different times, hence the dependence on energy bin index is also potentially important.

The expectation  $\langle H_{mk} \rangle$  should be replaced by our best estimate of the histogram count for energy bin  $m$  at temperature  $\beta_k$ , which could be obtained from our yet-to-be-determined optimal estimate of the density of states  $\hat{\Omega}_m$ :

$$\begin{aligned} \langle H_{mk} \rangle &= N_k p_m(\beta_k) \Delta U \\ &\approx N_k \Delta U \hat{\Omega}_m \exp[f_k - \beta_k U_m] \end{aligned} \quad (28)$$

Substituting this expression back into eqs 27 and 24, we obtain

$$\begin{aligned} \delta^2 \hat{\Omega}_{mk} &= \frac{g_{mk} N_k \Delta U \hat{\Omega}_m \exp[f_k - \beta_k U_m]}{\{N_k \Delta U \exp[f_k - \beta_k U_m]\}^2} \\ &= \frac{\hat{\Omega}_m}{g_{mk}^{-1} N_k \Delta U \exp[f_k - \beta_k U_m]} \end{aligned} \quad (29)$$

Using eq 29 for the uncertainty in the density of states associated with simulation  $k$  and eq 22 for the best estimate of the density of states, along with eq 11, we obtain

$$\hat{\Omega}_m = \frac{\sum_{k=1}^K g_{mk}^{-1} H_{mk}}{\sum_{k=1}^K g_{mk}^{-1} N_k \Delta U \exp[f_k - \beta_k U_m]} \quad (30)$$

Everything in the above expression can be easily evaluated, except for the  $f_k$ , which depends on the  $\hat{\Omega}_m$  through

$$f_k = -\ln \sum_{m=1}^M \hat{\Omega}_m \Delta U e^{-\beta_k U_m} \quad (31)$$

The  $f_k$  may therefore be solved for self-consistency by iteration of eqs 30 and 31 starting from an arbitrary choice, such as  $f_k = 0$ .

The statistical uncertainty in  $\hat{\Omega}_m$  is given by eq 14:

$$\delta^2 \hat{\Omega}_m = \frac{\hat{\Omega}_m}{\sum_{k=1}^K g_{mk}^{-1} N_k \Delta U \exp[f_k - \beta_k U_m]} \quad (32)$$

We note that the relative uncertainty in this estimate is given by

$$\frac{\delta^2 \hat{\Omega}_m}{\hat{\Omega}_m^2} = \left[ \sum_{k=1}^K g_{mk}^{-1} H_{mk} \right]^{-1} \quad (33)$$

which is equal to  $H_m^{-1}$ , the inverse of the total number of configurations from all simulations in energy bin  $m$ , if all  $g_{mk}$  are unity. This is reasonable, since the uncertainty in our estimate for  $\Omega_m$  should diminish as more independent samples are collected in energy bin  $m$ .

**2.6. Estimating an Observable at the Temperature of Interest.** Using the estimate of the density of states obtained above, we can obtain an estimate for the expectation of any configuration function  $A(\mathbf{q})$  at an arbitrary temperature by writing analogous equations to eqs 3 and 4 where we have discretized the energy  $U$

$$\langle A \rangle_\beta \approx \frac{\sum_{m=1}^M \hat{\Omega}_m \Delta U e^{-\beta U_m} A_m}{\sum_{m=1}^M \hat{\Omega}_m \Delta U e^{-\beta U_m}} \quad (34)$$

where

$$A_m = \frac{\int d\mathbf{q} A(\mathbf{q}) \psi_m(U(\mathbf{q}))}{\int d\mathbf{q} \psi_m(U(\mathbf{q}))} \quad (35)$$

$A_m$ , the mean of observable  $A$  over all configurations with potential energies consistent with energy bin  $m$ , can be best approximated by pooling configurations from *all*  $K$  simulations that have energies in bin  $m$

$$\hat{A}_m = H_m^{-1} \sum_{k=1}^K \sum_{n=1}^{N_k} \psi_{mkn} A_{kn} \quad (36)$$

where  $H_m$  (eq 9) is the total count of configurations in energy bin  $m$  from all simulations. Substituting this expression for  $A_m$  into eq 34 above produces an estimator  $\hat{A}(\beta)$  for  $\langle A \rangle_\beta$

$$\begin{aligned} \hat{A}(\beta) &= \frac{\sum_{m=1}^M \hat{\Omega}_m \Delta U e^{-\beta U_m} A_m}{\sum_{m=1}^M \hat{\Omega}_m \Delta U e^{-\beta U_m}} \\ &= \frac{\sum_{m=1}^M \hat{\Omega}_m e^{-\beta U_m} [H_m^{-1} \sum_{k=1}^K \sum_{n=1}^{N_k} \psi_{mkn} A_{kn}]}{\sum_{m=1}^M \hat{\Omega}_m e^{-\beta U_m} [H_m^{-1} \sum_{k=1}^K \sum_{n=1}^{N_k} \psi_{mkn}]} \\ &= \frac{\sum_{k=1}^K \sum_{n=1}^{N_k} w_{kn}(\beta) A_{kn}}{\sum_{k=1}^K \sum_{n=1}^{N_k} w_{kn}(\beta)} \end{aligned} \quad (37)$$

where we have defined the per-configuration weights  $w_{kn}(\beta)$  by

$$w_{kn}(\beta) = \sum_{m=1}^M \psi_{mkn} H_m^{-1} \hat{\Omega}_m e^{-\beta U_m} \quad (38)$$

where only one term of the sum will contribute—the bin  $m$  containing the energy  $U_{kn}$ —due to the presence of the indicator function  $\psi_{mkn}$ . Note that we only need to compute the weight  $w_{kn}$  up to a constant of proportionality because this constant drops out in the normalized sum in eq 37.

This relationship is significant in that we now have an expression for the canonical expectation of observable  $A$  in terms of a weighted sum over *all* of the data. These weights are determined by the temperature of interest from the WHAM equations and are simple functions of the count of

configurations with energies falling in a particular energy bin. The weights  $w_{kn}(\beta)$  can be computed once for the temperature of interest and then used to calculate expectations of many observables.

It should be noted that our estimate of  $\langle A \rangle_\beta$  will only be reasonable if the inverse temperature of interest  $\beta$  lies near or within the range of inverse temperatures sampled by the canonical simulations—the uncertainty in the estimate will increase as the temperature of interest deviates from the sampled range of temperatures (see refs 30 and 31 for an examination of this issue).

**2.7. Statistical Uncertainty of the Estimator for the Expectation.** If the observable of interest has a long correlation time compared to fluctuations in the potential energy (e.g., if the observable is a function of the large scale molecular conformation), then it is possible that the density of states  $\Omega_m$  and dimensionless free energies  $\{f_k\}$  may be sufficiently well-converged that they are not dominant contributors to the uncertainty in the estimate of the observable of interest. Instead, the long time-correlation in the observable means that there are many fewer effectively independent observations of the observable than stored configurations. We may then use the following procedure.

We can rewrite the estimator  $\hat{A}_\beta$  as a ratio of two random quantities  $X$  and  $Y$

$$\hat{A}_\beta = \frac{\hat{X}}{\hat{Y}} \quad (39)$$

where

$$\begin{aligned} \hat{X} &\equiv \sum_{k=1}^K \sum_{n=1}^{N_k} w_{kn}(\beta) A_{kn} \\ \hat{Y} &\equiv \sum_{k=1}^K \sum_{n=1}^{N_k} w_{kn}(\beta) \end{aligned} \quad (40)$$

Applying standard error propagation techniques for a function of random variables (see, e.g. ref 32), which amounts to a first-order Taylor series expansion of  $\hat{A}$  about  $\langle \hat{X} \rangle / \langle \hat{Y} \rangle$ , we can estimate the uncertainty in  $\hat{A}$  as

$$\delta^2 \hat{A} = \left[ \frac{\hat{X}}{\hat{Y}} \right]^2 \left[ \frac{\delta^2 \hat{X}}{\hat{X}^2} + \frac{\delta^2 \hat{Y}}{\hat{Y}^2} - 2 \frac{\delta \hat{X} \delta \hat{Y}}{\hat{X} \hat{Y}} \right] \quad (41)$$

Here, the cross-term  $\delta \hat{X} \delta \hat{Y} \equiv \langle (\hat{X} - \langle \hat{X} \rangle)(\hat{Y} - \langle \hat{Y} \rangle) \rangle$  is non-zero only if the random variables  $X$  and  $Y$  are correlated, in which case the term involving it in the equation above serves to reduce the uncertainty in the estimate of the ratio  $\hat{A}$ .

Recognizing that  $X$  and  $Y$  include contributions from  $K$  statistically independent simulations, we can collect these terms and write (dropping the hats)

$$\begin{aligned} X &\equiv \sum_{k=1}^K N_k X_k \\ Y &\equiv \sum_{k=1}^K N_k Y_k \end{aligned}$$

$$\begin{aligned} X_k &\equiv \frac{1}{N_k} \sum_{n=1}^{N_k} w_{kn} A_{kn} \\ Y_k &\equiv \frac{1}{N_k} \sum_{n=1}^{N_k} w_{kn} \end{aligned} \quad (42)$$

where the argument  $\beta$  has been omitted for notational convenience. Because the  $K$  individual simulations are *independent*, the uncertainties required in eq 41 are given by

$$\begin{aligned} \delta^2 X &= \sum_{k=1}^K N_k^2 \delta^2 X_k \\ \delta^2 Y &= \sum_{k=1}^K N_k^2 \delta^2 Y_k \\ \delta X \delta Y &= \sum_{k=1}^K N_k^2 \delta X_k \delta Y_k \end{aligned} \quad (43)$$

These uncertainties involve the correlated data of simulation  $k$  and can be estimated by standard correlation analysis methods<sup>21,23</sup> or by block transformation methods,<sup>22</sup> though the latter method requires some modification to estimate the uncertainty cross-term  $\delta X_k \delta Y_k$ .

To compute the uncertainties by correlation analysis methods as in section 2.4, we first define new observables  $x_{kn} = w_{kn} A_{kn}$  and  $y_{kn} = w_{kn}$  and compute the uncertainties

$$\begin{aligned} \delta^2 X_k &= \frac{\sigma_{k,x;x}^2}{N_k / g_{k,x;x}} \\ \delta^2 Y_k &= \frac{\sigma_{k,y;y}^2}{N_k / g_{k,y;y}} \\ \delta X_k \delta Y_k &= \frac{\sigma_{k,x;y}^2}{N_k / g_{k,x;y}} \end{aligned} \quad (44)$$

These uncertainties involve (co)variances of the type  $\sigma_{k,x;y}^2$ , estimated for each replica by

$$\hat{\sigma}_{k,x;y}^2 = \frac{1}{N_k - 1} \sum_{n=1}^{N_k} (x_{kn} - \hat{X}_k)(y_{kn} - \hat{Y}_k) \quad (45)$$

The statistical inefficiencies of the form  $g_{k,x;y}$  are computed by

$$g_{k,x;y} \equiv 1 + 2\tau_{k,x;y} \quad (46)$$

$$\tau_{k,x;y} \equiv \sum_{t=1}^{N_k-1} \left( 1 - \frac{t}{N_k} \right) C_{kt,x;y} \quad (47)$$

with the correlation function for simulation  $k$  computed by taking advantage of stationarity and time-reversibility:

$$C_{kt,xy} \approx \frac{1}{2\hat{\sigma}_{k,xy}^2} \frac{1}{(N_k - t)} \times \sum_{n=1}^{N_k - t} [(x_{kn} - \hat{X}_k)(y_{kn+t} - \hat{Y}_k) + (y_{kn} - \hat{Y}_k)(x_{kn+t} - \hat{X}_k)] \quad (48)$$

See section 5.2 for a discussion on efficiently computing the integrated correlation time  $\tau$  from  $\hat{C}_{kt,xy}$ .

### 3. Simulated and Parallel Tempering

**3.1. WHAM for Simulated Tempering (STWHAM).** In a simulated tempering simulation,<sup>10,3</sup> a single simulation is conducted in which configurations are sampled from a *mixed-canonical* ensemble.<sup>16</sup> In practice, a simulation algorithm that samples from the canonical ensemble is used to generate configurations, and at regular intervals attempts are made to change the temperature among a discrete set of choices  $\beta_1, \dots, \beta_L$ . The probability of accepting a proposed temperature change is given by the Metropolis-like criterion

$$P(\beta_i \rightarrow \beta_r) = \min\{1, \exp[-(\beta_r - \beta_i)U + (a_r - a_i)]\} \quad (49)$$

where the constants  $\{a_i\}_{i=1}^L$  are specified beforehand and chosen, often by tedious exploratory simulations, to attempt to achieve near-equal visitation of each temperature and, hopefully, potential energy. The optimal choice of  $\{a_i\}$  is given by the dimensionless free energies  $\{f_k\}$  in the equation above, and proposed temperature changes are usually between neighboring temperatures because the exchange probability diminishes with increased temperature separation. Use of the above criterion for accepting or rejecting proposed temperature changes ensures that, if the configurations were originally distributed from the equilibrium distribution at the old temperature, they are also distributed from the canonical distribution at the new temperature.

As a result of this procedure, the system spends a fraction of time in each of a number of different temperatures. Since we know the number of times each temperature was visited, we can write the probability density for energy bin  $m$  as a weighted sum of the canonical probability density functions at these different temperatures

$$p_m = \sum_{l=1}^L \frac{N_l}{N} \cdot \frac{\Omega_m e^{-\beta_l U_m}}{Z(\beta_l)} \quad (50)$$

where  $N_l/N$  is the fraction of configurations generated at inverse temperature  $\beta_l$  over the course of the simulation. As above, we introduce the Helmholtz free energy  $f_l \equiv -\ln Z(\beta_l)$ , which allows us to write

$$\begin{aligned} p_m &= \sum_{l=1}^L \frac{N_l}{N} \Omega_m \exp[f_l - \beta_l U_m] \\ &= \Omega_m \sum_{l=1}^L (N_l/N) \exp[f_l - \beta_l U_m] \end{aligned} \quad (51)$$

We can approximate  $p_m$  as before using our histogram count,  $H_m$ , the number of configurations with potential energy in the bin centered about  $U_m$ :

$$p_m \approx \frac{1}{\Delta U} \cdot \frac{H_m}{N} \quad (52)$$

Rearranging and including our definition of  $f_l$ , we obtain the coupled set of equations for estimating the density of states

$$\hat{\Omega}_m = \frac{H_m}{\sum_{l=1}^L N_l \Delta U \exp[f_l - \beta_l U_m]} \quad (53)$$

$$f_l = -\ln \sum_{m=1}^M \hat{\Omega}_m \Delta U e^{-\beta_l U_m} \quad (54)$$

These equations are similar to eqs 30 and 31 for the canonical ensemble WHAM if the configurations are grouped by the temperature at which they were generated but lacking statistical inefficiency terms since we are not combining data from multiple simulations.

The uncertainty in  $\hat{\Omega}_m$  is then given by

$$\begin{aligned} \delta^2 \hat{\Omega}_m &= \frac{\delta^2 H_m}{\left\{ \sum_{l=1}^L N_l \Delta U \exp[f_l - \beta_l U_m] \right\}^2} \\ &= \frac{g_m \langle H_m \rangle}{\left\{ \sum_{l=1}^L N_l \Delta U \exp[f_l - \beta_l U_m] \right\}^2} \end{aligned} \quad (55)$$

where, as in eq 27, we assume the histograms are sparsely populated and introduce the statistical inefficiency  $g_m$  to estimate the histogram uncertainty.

The estimate for the expectation of the total histogram count in energy bin  $m$  is given by the sampling probability

$$\langle H_m \rangle = N \Delta U p_m \approx N \Delta U \hat{\Omega}_m \sum_{l=1}^L (N_l/N) \exp[f_l - \beta_l U_m] \quad (56)$$

which gives the final estimate for the uncertainty as

$$\delta^2 \hat{\Omega}_m = \frac{\hat{\Omega}_m}{\sum_{l=1}^L g_m^{-1} N_l \Delta U \exp[f_l - \beta_l U_m]} \quad (57)$$

Following the approach in section 2.6, we can again write the estimator in the form of a weighted sum over configurations

$$\hat{A}(\beta) = \frac{\sum_{n=1}^N w_n(\beta) A_n}{\sum_{n=1}^N w_n(\beta)} \quad (58)$$

$$w_n(\beta) = \sum_{m=1}^M \psi_{nm} H_m^{-1} \hat{\Omega}_m e^{-\beta U_m} \quad (59)$$

where again, only one term contributes to the sum in the expression for the weight  $w_n$ . The statistical uncertainty in this estimate, as in section 2.7, can be computed by eq 41, where  $X$  and  $Y$  are now given by

$$\begin{aligned} X &\equiv \frac{1}{N} \sum_{n=1}^N w_n A_n \\ Y &\equiv \frac{1}{N} \sum_{n=1}^N w_n \end{aligned} \quad (60)$$

These uncertainties are simply computed as in eqs 44–48, without the subscript  $k$  as there is only one simulation instead of many. The quantities  $\hat{X}$  and  $\hat{Y}$  no longer correspond to canonical averages, since they are the expectations over the simulated tempering trajectory which spends a different amount of time at each of the  $L$  temperatures—it is a mixed canonical average. The sample mean over the trajectory provides the best estimator for these quantities.

Here, the statistical inefficiency  $g_m$  appearing in eq 57 and the inefficiencies required in applying eqs 44–48 are computed from the correlation functions computed over the simulated tempering trajectory, which includes unphysical jumps in temperature. It is worth noting that expressions for  $\langle A \rangle_\beta$  given in formulations by Okamoto and co-workers (e.g. eq 24 of ref 14) instead contain a statistical inefficiency for each temperature. In principle, one could account for a temperature-dependent statistical inefficiency, since one might expect correlation times to be different at each temperature, but, in practice, the limited number of configurations sampled between temperature changes is likely too short to allow temperature-dependent correlation times to be computed. Additionally, a temperature-dependent treatment does not account for the correlation between configurations sampled before and after a temperature swap. The derivation presented here assumes the statistical inefficiency  $g_m$  depends only on the energy bin  $m$ , which causes these factors to cancel out of our estimator for  $\langle A \rangle_\beta$  in eqs 58 and 59.

**3.2. WHAM for Parallel Tempering or Independent Simulated Tempering Simulations (PTWHAM).** In a parallel tempering (or replica-exchange among temperatures) simulation,<sup>12,4</sup> it was recognized that the constants  $a_k$  needed in the simulated tempering simulation to ensure equal sampling of temperatures could be eliminated if multiple simulated tempering simulations were conducted in parallel and the temperature changes of two simulations were coupled together into a temperature swap between the replicas. In practice, a number  $K$  of replicas are simulated independently at inverse temperatures  $\beta_1, \dots, \beta_K$  using some simulation method that samples from the canonical distribution. At given intervals, an attempt is made to exchange the temperatures of two replicas  $i$  and  $j$ , with the exchange accepted with probability

$$\begin{aligned} P_{\text{exch}} &= \min\{1, \exp[-(\beta_j - \beta_i)U_i + (a_j - a_i)]\} \times \\ &\quad \min\{1, \exp[-(\beta_i - \beta_j)U_j + (a_i - a_j)]\} = \\ &\quad \min\{1, \exp[-(\beta_j - \beta_i)(U_i - U_j)]\} \end{aligned} \quad (61)$$

where  $\beta_i$  is the current inverse temperature of replica  $i$ , and  $U_i$  is the corresponding potential energy.

Because of this exchange procedure, each replica executes a more or less random walk in temperature, eliminating the need to perform exploratory simulations to determine the parameters  $\{a_i\}_{i=1}^K$  required for simulated tempering. Each replica simulation is nearly independent, as the correlation between configurations of different replicas introduced by the exchange of temperatures is minimal. The dominant contribution to statistical uncertainties will almost certainly be due to the variance and temporal correlation in the value of the observable of interest within each replica, which reduces the effective number of independent samples. We can therefore analyze a parallel tempering simulation as a set of *independent* simulated tempering simulations, each with a number  $L$  of accessible temperatures, with  $L$  equal to the number of replicas  $K$ . Below, we derive an analogue of the Kumar et al. WHAM procedure for the treatment of  $K$  independent simulated tempering simulations (replicas) each capable of visiting  $L$  temperatures, allowing this method to also treat simulations generated by procedures such as REST.<sup>3</sup> We make use of the sampling distribution for simulated tempering described above and properly account for the correlation within each replica, eliminating the need to artificially reorder configurations from parallel tempering simulations by temperature.

We can use the simulated tempering eqs 53 and 57 above to write the estimator and uncertainty for the density of states obtained from each replica  $k$  as

$$\hat{\Omega}_{mk} = \frac{H_{mk}}{\sum_{l=1}^L N_{kl} \Delta U \exp[f_l - \beta_l U_m]} \quad (62)$$

$$\delta^2 \hat{\Omega}_{mk} = \frac{\hat{\Omega}_m}{g_{mk}^{-1} \sum_{l=1}^L N_{kl} \Delta U \exp[f_l - \beta_l U_m]} \quad (63)$$

where we have added the index  $k$  to denote the replica from which the data are generated.  $H_{mk}$  therefore denotes the number of configurations sampled with potential energy in energy bin  $m$  from replica  $k$ , and  $N_{kl}$  denotes the number at temperature  $\beta_l$  from replica  $k$ .  $g_{mk}$  is the statistical inefficiency computed from replica  $k$  for energy bin  $m$ .

Again using the optimal combination rule of eq 13, we obtain the optimal estimate for the density of states

$$\hat{\Omega}_m = \frac{\sum_{k=1}^K g_{mk}^{-1} H_{mk}}{\sum_{k=1}^K g_{mk}^{-1} \sum_{l=1}^L N_{kl} \Delta U \exp[f_l - \beta_l U_m]} \quad (64)$$

and the statistical uncertainty from eq 14:

$$\delta^2 \hat{\Omega}_m = \left\{ \frac{\sum_{k=1}^K \left[ g_{mk}^{-1} \sum_{l=1}^L N_{kl} \Delta U \exp[f_l - \beta_l U_m] \right]}{\hat{\Omega}_m} \right\}^{-1}$$

$$= \frac{\hat{\Omega}_m}{\sum_{k=1}^K g_{mk}^{-1} \sum_{l=1}^L N_{kl} \Delta U \exp[f_l - \beta_l U_m]} \quad (65)$$

We can rewrite eq 64 as

$$\hat{\Omega}_m = \frac{H_m^{\text{eff}}}{\sum_{l=1}^L N_{ml}^{\text{eff}} \Delta U \exp[f_l - \beta_l U_m]} \quad (66)$$

where  $H_m^{\text{eff}} \equiv \sum_{k=1}^K g_{mk}^{-1} H_{mk}$  is the effective number of independent samples in energy bin  $m$  from all replicas, and  $N_{ml}^{\text{eff}} \equiv \sum_{k=1}^K g_{mk}^{-1} N_{kl}$  is an effective number of independent samples at temperature  $\beta_l$  from all replicas.

To compute the estimator of the expectation for an observable  $A$ , we apply the same technique in section 2.6 above and write the expectation as a weighted sum over configurations

$$\hat{A}(\beta) = \frac{\sum_{k=1}^K \sum_{n=1}^{N_k} w_{kn}(\beta) A_{kn}}{\sum_{k=1}^K \sum_{n=1}^{N_k} w_{kn}(\beta)} \quad (67)$$

where the weights are given by

$$w_{kn}(\beta) \equiv \sum_{m=1}^M \psi_{mkn} H_m^{-1} \hat{\Omega}_m e^{-\beta U_m} \quad (68)$$

As in eqs 38 and 59, the sum over  $m$  reduces to a single term, the one with energy bin index appropriate for configuration  $n$  of replica  $k$ .  $A_{kn}$  is the value of the observable  $A$  for configuration  $n$  of replica  $k$  and  $H_m = \sum_{k=1}^K \sum_{n=1}^{N_k} \psi_{mkn}$ , the total number of configurations from all replicas with potential energy in bin  $m$ .

Again, if the observable of interest has a correlation time that is long compared to fluctuations in the potential energy, we may compute the dominant contribution to the statistical uncertainty  $\delta^2 A(\beta)$  by eqs 41–48, with the important distinction that  $k$  now indexes the *replicas*, rather than the temperatures. The correlation times are, as in the simulated tempering case, computed over the nonphysical replica trajectories; because the replicas perform random walks in temperature, these times are likely to be shorter than the correlation time for this observable computed from a canonical simulation at the lowest temperature. These replica correlation times properly capture the correlation between successive snapshots generated by a sampling method like Metropolis Monte Carlo or molecular dynamics, and their use in estimating the uncertainty is the primary novel result of this paper. Collecting configurations from all replicas into pseudotrajectories of constant temperature, as suggested in

previous attempts to apply the method to parallel tempering simulations,<sup>4</sup> would give correlation times that are erroneously short and make the incorrect assumption that these pseudotrajectories are statistically independent.

## 4. Applications

**4.1. One-Dimensional Model Potential.** To validate the STWHAM and PTWHAM methods described above for estimating expectations and corresponding uncertainties, we consider a one-dimensional model potential where canonical expectations can be computed directly and a large quantity of simulation data can be obtained in order to verify our uncertainty formulas. We use an asymmetric double well potential, given by

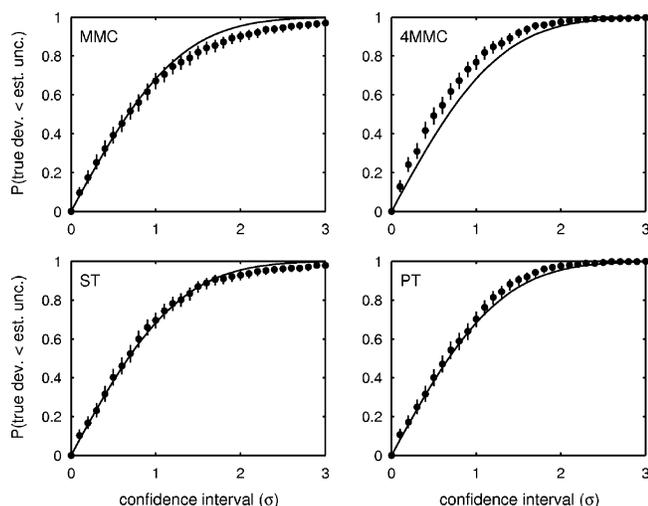
$$U(q) = (q - 1)^2(q + 1)^2 + 0.1q \quad (69)$$

All simulations utilize the Metropolis Monte Carlo method<sup>33</sup> with the trial displacement  $\Delta q$  uniformly distributed on the interval  $[-0.2, +0.2]$  to generate a series of configurations which are sampled every 10 move attempts, resulting in highly correlated data. In the following simulations, we estimate the expectation  $\langle q \rangle_{\beta^*}$  at  $\beta^* = 4$ , where the integrated correlation time of  $q$  is rather long—approximately 130 samples. The initial conformation was chosen uniformly on the interval  $[-1.8, +1.8]$  and the first  $10^5$  steps discarded to equilibration.

Four types of simulations were performed: a standard canonical Metropolis Monte Carlo (MMC) simulation at  $\beta = \beta^* = 4$ , as described above; a set of four independent canonical (4MMC) simulations with inverse temperatures  $\beta$  exponentially spaced in the range  $1-4$  ( $\beta \approx \{4, 2.52, 1.59, 1\}$ ); a simulated tempering simulation (ST) with the same four possible temperatures and analytically computed optimal weights; and a parallel tempering (PT) simulation with replicas at the same four temperatures.

All simulations were conducted for  $5 \times 10^7$  steps each (per replica, if multiple replicas are used), generating  $5 \times 10^6$  samples (per replica). The data were then divided into 500 sequential blocks of  $10^4$  configurations (per replica) each, whose expectations were verified to be statistically independent by computing the correlation between expectations in neighboring blocks. The standard deviation of the set of expectations computed from each block is indicative of the statistical uncertainty in simulations of a single block length— $10^4$  samples (per replica)—and the difference between the mean of these estimates and the expectation computed from the potential directly is indicative of the bias. Expectations and uncertainties for each block were computed using the code appearing in listing 1 of the Supporting Information.

To assess the performance of the uncertainty estimate for each block, we compute the fraction of blocks for which the true magnitude of the deviation from the mean of the block expectations is smaller than a multiplicative constant  $\sigma$  times the estimated uncertainties, for  $\sigma \in [0.1, 3]$ . This fraction is related to a confidence interval if compared to the error function Gaussian integral (Figure 1). For example, for our uncertainty estimates to be meaningful, we expect the difference between the true mean and our estimate to be within one standard deviation  $\sigma$  approximately 66% of the



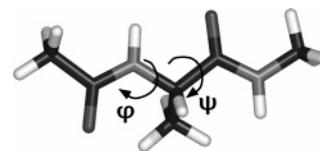
**Figure 1.** Confidence curves for Metropolis Monte Carlo simulations on the 1D model potential. The fraction of statistically independent blocks for which the true uncertainty (the deviation of the estimated expectation over the block from the mean of the block estimates) is less than a multiplier of the predicted  $1\sigma$  uncertainty (here plotted as the independent variable is shown). The solid curve shows the fraction expected to fall within the interval for the normal distribution. Ideally, the curves would coincide. The results are shown for (MMC) a single Metropolis Monte Carlo simulation at  $\beta = 4$ ; (4MMC) a set of four independent canonical simulations spanning the range  $\beta = 1-4$ ; (ST) a simulated tempering simulation spanning  $\beta = 1-4$ ; (PT) a parallel tempering simulation with four replicas spanning  $\beta = 1-4$ . Uncertainties, with 95% confidence intervals shown here as vertical bars, were computed as described in Appendix B.

time. It is readily apparent from the figure that the computed uncertainty estimate computed for each block is in fact quite good. Additionally, the bias is small—less than 10% of the magnitude of the statistical uncertainty in the cases studied (data not shown).

#### 4.2. Alanine Dipeptide in Implicit and Explicit Solvent.

To illustrate the utility and verify the correctness of the PTWHAM procedure described above for simulations of biological interest, we demonstrate their use in the analysis of parallel tempering simulations of alanine dipeptide in implicit and explicit solvent. A similar strategy to the 1D model system described above was adopted, with a long simulation partitioned into short blocks (here, 2 ns/replica per block) whose expectations were verified to be statistically independent by the same procedure described above.

Using the LEaP program from the AMBER7 molecular mechanics package,<sup>34</sup> a terminally blocked alanine peptide (sequence ACE-ALA-NME, see Figure 2) was generated in the extended conformation. For the explicit solvent system, the peptide was solvated with 431 TIP3P water molecules<sup>35</sup> in a truncated octahedral simulation box whose dimensions were chosen to ensure a minimum distance to the box boundaries from the initial extended peptide configuration of 7 Å. Peptide force field parameters were taken from the parm96 parameter set.<sup>36</sup> For the implicit solvent simulation, the Generalized Born method of Tsui and Case<sup>37</sup> (corre-



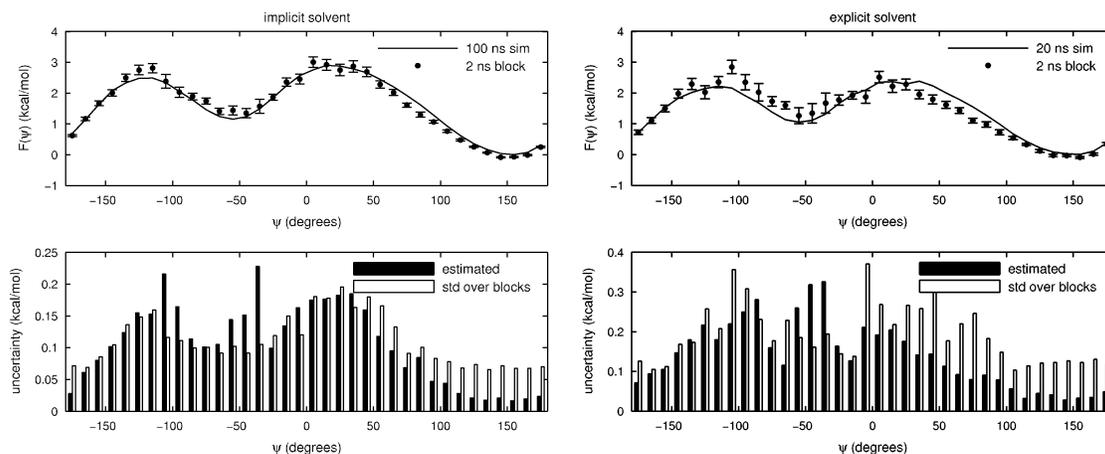
**Figure 2.** Terminally blocked alanine peptide with  $(\phi, \psi)$  torsions labeled.

sponding to the flag `igb=1`) was employed with radii from AMBER6, along with a surface area penalty term of the default  $5 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ . Covalent bonds to hydrogen were constrained with SHAKE using a tolerance of  $10^{-8} \text{ \AA}$ .<sup>38</sup> Long-range electrostatics for the explicit solvent simulation were treated by the particle-mesh Ewald (PME) method<sup>39</sup> with default settings.

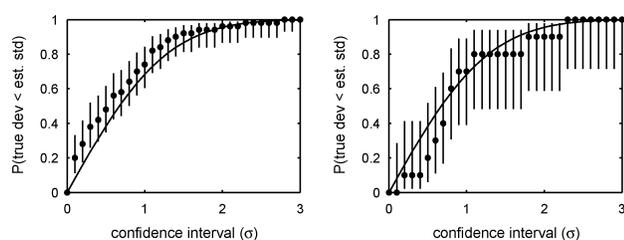
Each system was first subjected to 50 steps of steepest descent energy minimization, followed by 1000 steps of conjugate gradient optimization. To equilibrate the explicit solvent system to the appropriate volume, a 100 ps molecular dynamics simulation was performed with the temperature adjusted to 300 K and the pressure to 1 atm by the Berendsen weak-coupling algorithm<sup>40</sup> with temperature and pressure relaxation time constants of 1 and 0.2 ps, respectively. The simulation box was fixed at the final size obtained from this equilibration step, with a volume of  $13\,232 \text{ \AA}^3$ , in all subsequent simulations.

A parallel tempering (or replica-exchange among temperatures) molecular dynamics simulation<sup>4</sup> was conducted using a parallel Perl wrapper for the sander program. [A copy of this Perl wrapper to perform replica-exchange simulations using AMBER7 and AMBER8 can be obtained from URL <http://www.dillgroup.ucsf.edu/~jchodera/code/rer>.] Replica temperatures were exponentially distributed over the range 273–600 K, with 10 replicas required for the implicit solvent simulation (yielding an exchange acceptance probability between neighboring temperatures of approximately 75%) and 40 replicas for the explicit solvent simulation (yielding an acceptance probability of approximately 50%). All momenta were reassigned from the Maxwell-Boltzmann distribution at the appropriate replica temperature after each exchange attempt. Between exchanges, constant-energy, constant-volume molecular dynamics was carried out for the explicit solvent simulation, while the implicit solvent simulation utilized Langevin dynamics with a friction coefficient of  $95 \text{ ps}^{-1}$  to mimic the viscosity of water. All dynamics utilized a 2 fs time step. The algorithm used to select pairs of replicas for temperature exchange attempts starts from the highest-temperature replica and attempts to swap the configuration for the next-lowest temperature replica using a Metropolis-like criteria and proceeds down the temperatures in this manner. On the next iteration, swapping attempts start from the lowest temperature and proceed upward, and this alternation in direction is continued in subsequent pairs of iterations.

Starting all replicas from the minimized or volume-equilibrated configuration described above, 100 iterations were conducted with 1 ps between exchange attempts to equilibrate the replicas to their respective temperatures. This equilibration run was followed by a production run with 20



**Figure 3.** Potential of mean force in  $\psi$  for implicit and explicit solvent parallel tempering simulations. Left: implicit solvent; right: explicit solvent. Upper panels: The potential of mean force in the  $\psi$  torsion angle at 300 K. The solid line shows the PMF estimated from the entire simulation, while the filled circles show the estimated PMF uncertainty using the method described in the text for a single 2 ns/replica block. Lower panels: The computed uncertainties for the same 2 ns block (filled bars) along with the average uncertainty expected for a simulation 2 ns/replica in length, estimated from the standard deviation of the PMFs computed from all nonoverlapping blocks of length 2 ns in the full simulation (open bars). All uncertainties are shown as one standard deviation.



**Figure 4.** Confidence curves for implicit and explicit solvent parallel tempering simulations. As in Figure 1, the fraction of statistically independent 2 ns blocks for which the true uncertainty is less than a multiplier of the predicted  $1\sigma$  uncertainty is shown. The observable used is an indicator function for the  $\alpha_R$  configuration. Left: implicit solvent (statistics over 50 blocks); right: explicit solvent (statistics over 10 blocks).

ps between exchange attempts, giving a total of 100 ns/replica for the implicit solvent production run and 20 ns/replica for the explicit solvent run. Solute configurations and potential energies were saved from the production run every 1 ps. Expectations and uncertainties were again estimated using listing 1 appearing in the Supporting Information.

Over 2 ns blocks of simulation time (containing 2000 configurations/replica in each block), we computed the probability of the peptide occupying the  $\alpha_R$  conformation at 300 K, with  $\alpha_R$  here defined as  $-105 \leq \phi < 0$  and  $-124 \leq \psi < 28$ . This corresponds to configurations that would be classified as right-handed alpha-helical. To validate the uncertainty estimates, confidence curves of the type description in section 4.1 were computed and are shown in Figure 4. Though the confidence intervals are larger because the data contain fewer independent blocks, the uncertainty estimates are still good indicators of the expected deviation from the true expectation.

The potential of mean force (PMF) for the  $\psi$  torsion angle at 300 K was also computed and is shown in Figure 3. The

computed PMF and uncertainty for a representative block is depicted in the top panel, along with the PMF computed using the entire trajectory. The deviations of the block PMFs from the whole-simulation estimate fall within the  $1\sigma$  uncertainty bars to the expected degree. In the lower panel, the uncertainties computed from the representative block are compared to the standard deviation of the PMF computed from all blocks, which should be indicative (to within an order of magnitude) of the uncertainty expected from a simulation of the block length. These too compare favorably.

It is important to note that our neglect of the uncertainty in the dimensionless free energies,  $\{f_i\}$ , is only reasonable if the correlation time of the observable of interest is much longer than that of the potential energy. When this condition is satisfied, the dominant contribution to the uncertainty in the computed expectation of the observable is due to the small number of effectively independent samples of this observable present in the simulation data. To demonstrate that this is the case for systems of interest, we have assessed the relative contribution of the neglected uncertainty in the  $\{f_i\}$  to the uncertainty of the estimated probability of the  $\alpha_R$  conformation of the alanine dipeptide system considered here. The resulting contribution is 10 times smaller than the uncertainty due to the time correlation treated above for the explicit solvent system and 100 times smaller for the implicit solvent system. [The impact of the uncertainty in the  $\{f_i\}$  on the uncertainty in the estimated observable was computed in the following manner: We first computed estimates of the  $\{f_i\}$  over all uncorrelated 2 ns/replica blocks of simulation data to form a pool of dimensionless free energies that represent the typical uncertainty in a simulation of this length. Next, for each 2 ns/replica block, we computed the standard deviation in the estimated  $\alpha_R$  probability when all  $\{f_i\}$  in this pool were substituted into the WHAM equations. The mean of this standard deviation over all blocks then provides an estimate of the magnitude of the impact of typical

uncertainties in the  $\{f_i\}$  on the observable of interest.] However, if the observable has a correlation time comparable to that of the potential energy (e.g., if the expectation of the potential energy itself is of interest), then the uncertainty due to imperfect knowledge of the  $\{f_i\}$  can be comparable to the uncertainty due to the correlation in the observable. In these cases where correlation times are comparable, an algorithm that combines our approach with the T-WHAM method of Gallicchio et al.,<sup>24</sup> which explicitly treats the uncertainty in the  $\{f_i\}$  when the potential energy samples are uncorrelated, may provide a superior estimate of the uncertainty in the estimate of the observable.

We further note that pathological cases may arise where simulations at neighboring temperatures may have poor energy overlap, resulting in large uncertainties in some of the  $\{f_i\}$ . Fortunately, these cases are easily detected by examination of the exchange acceptance rates between neighboring temperatures, where they will be conspicuously low, and detectable early in the simulation. Such cases are easily remedied by adjusting the temperature spacing or by the addition of more replicas at intermediate temperatures.

## 5. Practical Considerations

Several issues of great importance to successful implementation of the algorithm have received little discussion in the literature.

**5.1. Choice of Bin Width and Number of Bins.** There is a bias-variance tradeoff in the choice of energy histogram width. As the energy bin width increases, the uncertainty in our histogram estimator for  $p_m(\beta)$ , the probability density for energy bin  $m$ , decreases. At the same time, one expects the resulting estimate of the density of states  $\Omega_m$  to become increasingly biased, especially considering the dependence of  $p(U)$  on the rapidly varying exponential Boltzmann factor  $e^{-\beta U}$ . Because of this, a reasonable assumption might be that the bin width  $\Delta U$  should be chosen such that  $\Delta U \ll k_B T$ . However, if the bin size is too small, the uncertainty in the estimate for the  $p_m(\beta)$  will be large. One possibility might be to use a data based choice of histogram bin width, as in Wand,<sup>41</sup> which uses concepts from nonparametric density estimation in attempting to minimize the mean integrated square error (MISE) to the true probability density.

For the alanine dipeptide simulations described in section 4.2 above, we find that the estimated probability of occupying the  $\alpha_R$  region of conformation space is largely insensitive to the number of bins used to discretize the sampled potential energy range. In fact, the variation in the computed expectation is well within the statistical uncertainty over the range of 50–5000 bins (corresponding to a range of bin widths of 0.5–50  $k_B T$ ).

**5.2. Computing Integrated Correlation Times.** Estimating the correlation time  $\tau$ , defined above in eqs 19 and 21, can be difficult when one is confronted with noisy correlation functions. While ensuring trajectories are many times longer than the longest correlation times is necessary for an accurate estimate, even if this is achieved, performing the straightforward sum over the entirety of the correlation function  $C_t$  as in eq 19 is almost always a poor choice, as the uncertainty

in the computed correlation function grows approximately linearly with the lag time  $t$ .<sup>42</sup> Even for trajectories many times longer than the correlation length, this sum will be dominated by contributions from the noisy tail, likely resulting in large errors or even negative values for the computed correlation time  $\tau$ . Janke proposes a self-consistent approach where the summation is performed only out to lag times of  $6\tau$ , after which the correlation function is assumed to be negligible.<sup>23</sup> Evertz contends that this approach produces incorrect results,<sup>43</sup> instead proposing an exponential fit to the tail of the correlation function and use of this fit to evaluate the summand when the correlation function is dominated by noise. Neither solution is both stable and straightforward to apply, so we instead truncate the sum when the normalized fluctuation correlation function  $C_t$  first crosses zero, since it is likely unphysical for the correlation function to be negative for most observables. [Velocity autocorrelation functions, where there is often a clear negative peak at short times, are an obvious exception.] The zero crossing is an indication that the statistical uncertainty dominates the signal and that the remainder of the correlation function should be considered indistinguishable from zero.

For most systems and observables, the correlation function will decay rapidly at first and then slowly, approximately exponentially for large  $t$ . To avoid the expense of computing  $C_t$  at each value of  $t$  while still obtaining reasonably accurate integrated correlation times for observables with very different decay time scales, we use an adaptive integration scheme in which the correlation function  $C_t$  is computed only at times  $t_i = 1 + i(i-1)/2$ , where  $i = 1, 2, 3, \dots$ . In computing the correlation time  $\tau$ , the sum in eq 19 is now performed only over the  $t_i$  terms, with each term weighted by  $t_{i+1} - t_i$ , with  $t_1 = 1$ . This approach ensures high time resolution at small  $t$  when  $C_t$  is likely to be rapidly changing but avoids the expense of computing  $C_t$  at every  $t$  in the slowly decaying tails. We find the accuracy of this approach to be acceptable—differences typically amount to at most 10%.

**5.3. Neglect of Bin Statistical Inefficiencies  $g_{kn}$ .** It is often assumed or stated without justification that the energy bin statistical inefficiencies  $g_{mk}$  appearing in eqs 30 and 64, representing the number of snapshots required for a statistically independent sampling of the energy bin, are all equal or equal to unity.<sup>2,4,1</sup> All  $g_{mk}$  will be equal to unity only if the  $\{\psi_{mkn}\}_{n=1}^{N_k}$  are uncorrelated. To test this assumption, we have computed the statistical inefficiencies for the systems mentioned in section 4 above. To our knowledge, this is the first time a test of this claim has been reported in the literature. Indeed, for the explicit solvent system studied in section 4.2 above, we find large differences in the statistical inefficiencies for the same replica but different energy bins, sometimes differing up to two orders of magnitude. Similarly, for the same energy bin, the statistical inefficiencies from different replicas can differ by up to 2 orders of magnitude.

In the limit that our parallel tempering simulation is very long—long enough for each replica to execute an unrestricted random walk through all temperatures and explore all relevant regions of configuration space—each replica can be considered to be equivalent. In this case, the statistical inefficiencies should be independent of replica index  $k$ , and

we can write  $g_m$  as the statistical inefficiency for energy bin  $m$ . Applied to eqs 64 and 65, this yields a new set of expressions:

$$\hat{\Omega}_m = \frac{\sum_{k=1}^K H_{mk}}{\sum_{k=1}^K \sum_{l=1}^L N_{kl} \Delta U \exp[f_l - \beta_l U_m]} \quad (70)$$

$$\delta^2 \hat{\Omega}_m = \frac{\hat{\Omega}_m}{g_m^{-1} \sum_{k=1}^K \sum_{l=1}^L N_{kl} \Delta U \exp[f_l - \beta_l U_m]} \quad (71)$$

The quantity  $\sum_{k=1}^K N_{kl}$ , simply represents the total number of configurations stored from any replica at temperature  $\beta_l$ . For a parallel tempering simulation, where each temperature must be populated by exactly one replica at all times, this is simply  $N$ , the total number of configurations stored per replica. Additionally,  $H_m \equiv \sum_{k=1}^K H_{mk}$  is the total number of configurations over all replicas (and hence temperatures) with energy in bin  $m$ . This gives

$$\Omega_m = \frac{H_m}{N \sum_{l=1}^L \exp[f_l - \beta_l U_m]} \quad (72)$$

which is identical to the WHAM result for independent canonical simulations for the case where all  $g_{mk}$  are identical or unity. If there is no correlation in the data—that is, all configurations are independent—it does not matter whether we apply this analysis to the original data or collect the configurations by temperature and apply WHAM equations for independent canonical simulations. This expression is in fact identical to the one used in many published works that have previously attempted to use the weighted histogram analysis method for the analysis of parallel tempering simulations, such as refs 4 and 1.

Under these same assumptions, the correlation functions for any observable  $A$  should also be identical for each replica. One can therefore average estimates of the unnormalized correlation functions  $\langle A_n A_{n+1} \rangle$  over the replicas and use optimal estimates of the mean and variance computed over all of the replicas to obtain an optimal estimate of the statistical inefficiencies and uncertainties. An implementation illustrating this procedure is provided in the Supporting Information as listing 2.

Note, however, that the above assumptions of equivalence cannot be made in cases where the replicas are clearly *inequivalent*, such as in a simulated tempering replica-exchange (STREM) simulation.<sup>3,14</sup> In that case, the expression above will only be recovered if the time between samples is so long that all the  $g_{mk}$  are unity.

**5.4. The Statistical Inefficiency for the Cross-Correlation Term,  $g_{k,wA,w}$ .** In computing the statistical inefficiency for the cross-correlation term, uncertainties in the computed integrated autocorrelation times due to insufficient data or approximations may kill the cross-correlation term to

dominate and the estimate of the square uncertainty of  $\hat{A}$  (eq 41) to be negative. Clearly, this should not be allowed to occur, as the squared-uncertainty should be a strictly positive quantity.

The statistical inefficiencies  $g$  should obey the following relation (derived in Appendix A):

$$g_{x,y} \leq (g_x g_y)^{1/2} \frac{\sigma_x \sigma_y}{|\sigma_{x,y}|^2} \quad (73)$$

This is often violated when the correlation function is noisy and can lead to negative estimates of the squared uncertainty when the cross-correlation term dominates. In these cases, we find it best to simply limit  $g_{x,y}$  to its maximum allowed value computed from the right-hand side of eq 73. Since the autocorrelation times are usually shorter than the cross-correlation time, it is believed that these estimates will be better than the integrated cross-correlation time.

## 6. Conclusion

We have presented extensions of the weighted histogram analysis method, STWHAM and PTWHAM, for the analysis of one or more independent simulated tempering or parallel tempering simulations. The method provides not only estimators of canonical expectations but also estimators for the statistical uncertainties in the resulting estimates. We hope that, with the availability of the provided example code, workers using these simulation techniques will provide uncertainty estimates so that the statistical significance of results obtained from them can be assessed. We have shown that the estimator for the expectation has small bias and produces excellent uncertainty estimates for both a 1D model system and a solvated biomolecular system in implicit and explicit solvent.

While other workers had attempted to apply WHAM to simulated or parallel tempering data in the past,<sup>3,4</sup> the key advance here is the consideration of the correlated nature of the configurations sampled by each replica as it performs a pseudorandom walk in temperature, allowing a proper assessment of the true number of independent samples present in the data. This produces correct optimal estimators and makes possible the estimation of statistical uncertainties. This method can be extended to the analysis of other generalized-ensemble simulations, such as the multicanonical method (MUCA),<sup>5–9</sup> by consideration of replica correlation times as the system samples various energy levels biased by the estimate of the density of states. Still, it is important to point out that, while we consider the contribution from time-correlation of the observable to the uncertainty estimate, we currently neglect the contribution of the uncertainty in the per-configuration weights (which originates from the uncertainty in the density of states) to the estimate of the expectation—we assume it is negligible and await a more complete treatment of the uncertainty in cases where it is not.

## Appendix A: Relation for Statistical Inefficiencies

Consider a random process where we make a series of  $N$  time-correlated measurements of two (possibly correlated)

observables  $X$  and  $Y$ , resulting in the time series  $\{x_n, y_n\}_{n=1}^N$ . We estimate the quantity  $Z = \langle X \rangle \langle Y \rangle$  from our sample means and wish to compute the uncertainty in our estimate, defined as

$$\delta^2 Z \equiv \sigma_Z^2 = \langle (Z - \langle Z \rangle)^2 \rangle = \langle Z^2 \rangle - \langle Z \rangle^2 \quad (74)$$

To first order about  $\langle X \rangle \langle Y \rangle$ , this uncertainty is given by

$$\sigma_Z^2 = \left[ \frac{\langle X \rangle}{\langle Y \rangle} \right]^2 \left[ \frac{\sigma_X^2}{\langle X \rangle^2} + \frac{\sigma_Y^2}{\langle Y \rangle^2} - 2 \frac{\sigma_{X;Y}}{\langle X \rangle \langle Y \rangle} \right] \quad (75)$$

where  $\sigma_{X;Y}$  denotes the (not necessarily positive) covariance of the expectations of  $X$  and  $Y$ . The Schwartz inequality requires that this covariance obey the relation

$$|\sigma_{X;Y}^2| \leq \sigma_X \sigma_Y \quad (76)$$

(see, for example, ref 44). Given this, we note

$$\left| \frac{\sigma_{X;Y}}{\sigma_X \sigma_Y} \right| \leq 1 \quad (77)$$

Correlation analysis gives estimators for these quantities as

$$\begin{aligned} \sigma_X^2 &= \sigma_x^2 / (g_x^{-1} N) \\ \sigma_Y^2 &= \sigma_y^2 / (g_y^{-1} N) \\ \sigma_{X;Y}^2 &= \sigma_{x;y}^2 / (g_{x;y}^{-1} N) \end{aligned} \quad (78)$$

where  $\sigma_x^2$  denotes the sample variance of the observations  $\{x_n\}_{n=1}^N$  and  $g$  denotes the statistical inefficiency obtained from the autocorrelation time, i.e.

$$g_x = 1 + 2\tau_x \geq 1 \quad (79)$$

Combining eqs 77 and 78 gives

$$\frac{|\sigma_{x;y}^2|}{\sigma_x \sigma_y} \cdot \frac{g_{x;y}}{(g_x g_y)^{1/2}} \leq 1 \quad (80)$$

where we have moved the statistical inefficiencies and variances out of the absolute value, as they are always positive. Finally, we obtain an upper bound for  $g_{x;y}$ :

$$g_{x;y} \leq (g_x g_y)^{1/2} \frac{\sigma_x \sigma_y}{|\sigma_{x;y}^2|} \quad (81)$$

In using numerical methods to estimate the statistical inefficiencies  $g$  from finite trajectories, this inequality may not hold, sometimes leading to negative squared uncertainties. Limiting the estimated  $g_{x;y}$  by capping it at this value will prevent this from occurring.

## Appendix B: Uncertainty Estimates for Confidence Curves

To estimate the uncertainties in Figures 1 and 4, a Bayesian inference scheme was used. Since the expectations computed from each block are independent, the number of blocks  $n$  that fall inside the given scaled deviation  $\sigma$  is described by

a binomial distribution with parameter  $\theta$ , true (unknown) probability that the blocks fall within the given deviation  $\sigma$ :

$$P(n|\theta) = \frac{N!}{n!(N-n)!} \theta^n (1-\theta)^{N-n} \quad (82)$$

We can write the posterior distribution for the probability  $p$  given the observed number of blocks  $n$  within the given deviation using Bayes' rule

$$p(\theta|n) \propto p(n|\theta) p(\theta) \quad (83)$$

where  $p(\theta)$  is the prior distribution for the parameter  $\theta$ . If we choose the prior  $p(\theta)$  to be a Beta distribution with hyperparameters  $\alpha$  and  $\beta$ , given by

$$p(\theta|\alpha,\beta) = [B(\alpha,\beta)]^{-1} x^{\alpha-1} (1-x)^{\beta-1} \quad (84)$$

where  $B(\alpha,\beta)$  is the beta function, then the posterior  $p(\theta|n)$  will also be a Beta distribution with parameters  $n + \alpha$  and  $(N - n) + \beta$ , as the Beta distribution is a conjugate prior to the Binomial distribution. We take the hyperparameters  $\alpha$  and  $\beta$  to be unity to make the prior distribution uniform, resulting in a posterior  $\theta \sim \text{Beta}(n + 1, N - n + 1)$ . A 95% central confidence interval, corresponding to the location where the cumulative distribution function for the Beta distribution reaches the values of 0.025 and 0.0975, was plotted.

**Acknowledgment.** J.D.C. was supported by a Howard Hughes Medical Institute and an IBM predoctoral fellowship. W.S. acknowledges support from NSF MRSEC Center on Polymer Interfaces and Macromolecular Assemblies DMR - 0213618, and K.D. acknowledges the support of NIH grant GM34993. The authors would like to thank Libusha Kelly, Bosco K. Ho, and M. Scott Shell (University of California, San Francisco) for critical reading of this manuscript as well as the anonymous referee who raised an excellent point regarding our neglect of the uncertainty in the dimensionless free energies. This manuscript was strengthened as a result of their input.

**Supporting Information Available:** Source code Listings 1 and 2 for reference implementations of PTWHAM in Fortran 90/95, assuming inequivalent and equivalent replicas, respectively, as well as an archive containing a compilable version of Listing 1 and a test data set. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Okamoto, Y. *J. Mol. Graphics Modell.* **2004**, *22*, 425–439.
- (2) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (3) Mitsutake, A.; Okamoto, Y. *Chem. Phys. Lett.* **2000**, *332*, 131–138.
- (4) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (5) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249–253.
- (6) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9–12.

- (7) Hansmann, U. H. E.; Okamoto, Y. *J. Comput. Chem.* **1993**, *14*, 1333–1338.
- (8) Nakajima, N.; Nakamura, H.; Kidera, A. *J. Phys. Chem. B* **1997**, *101*, 817–824.
- (9) Yaşar, F.; Çelik, T.; Berg, B. A.; Meirovitch, H. *J. Comput. Chem.* **2000**, *21*, 1251–1261.
- (10) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451–458.
- (11) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776–1783.
- (12) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
- (13) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *J. Chem. Phys.* **2003**, *118*, 6664–6675.
- (14) Mitsutake, A.; Okamoto, Y. *J. Chem. Phys.* **2004**, *121*, 2491–2504.
- (15) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **2000**, *2329*, 261–270.
- (16) Fischer, A.; Cordes, F.; Schütte, C. *J. Comput. Chem.* **1998**, *1999*, 1689–1697.
- (17) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 225–234.
- (18) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1998**, *61*, 2635–2638.
- (19) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.
- (20) Müller-Krumbhaar, H.; Binder, K. *J. Stat. Phys.* **1973**, *8*, 1–23.
- (21) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637–649.
- (22) Flyvbjerg, H.; Petersen, H. G. *J. Chem. Phys.* **1989**, *91*, 461–466.
- (23) Janke, W. Statistical analysis of simulations: Data correlations and error estimation. In *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*; Grotendorst, J., Marx, D., Murmatsu, A., Eds.; John von Neumann Institute for Computing: 2002; Vol. 10.
- (24) Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. *J. Phys. Chem. B* **2005**, *109*, 6722–6731.
- (25) Gelman, A.; Rubin, D. B. *Stat. Sci.* **1992**, *7*, 457–472.
- (26) Souialle, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40–57.
- (27) Silverman, B. W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: New York, 1986.
- (28) Willard, K. E.; Connelly, D. P. *Comput. Biomed. Res.* **1992**, *25*, 17–28.
- (29) Cowan, G. *Statistical Data Analysis*; Oxford University Press: 1998.
- (30) Ferrenberg, A. M.; Landau, D. P.; Swendsen, R. H. *Phys. Rev. E* **1995**, *51*, 5092–5099.
- (31) Newman, M. E. J.; Palmer, R. G. *J. Stat. Phys.* **1999**, *97*, 1011–1026.
- (32) Taylor, J. R. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd ed.; University Science Books: 1996; Chapter 3.11, pp 73–79.
- (33) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (34) Case, D. A. et al. *AMBER7*; 2002.
- (35) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (36) Kollman, P. A.; Dixon, R.; Cornell, W.; Vox, T.; Chipot, C.; Pohorille, A. The development/application of a ‘minimalist’ organic/biochemical molecular mechanic force field using a combination of *a b* initio calculations and experimental data. In *Computer Simulation of Biomolecular Systems*; Wilkinson, A., Weiner, P., van Gunsteren, W. F., Eds.; Kluwer/Escom: 1997; Vol. 3.
- (37) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489–2498.
- (38) Ryckaert, J.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (39) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (40) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (41) Wand, M. P. *Am. Stat.* **1997**, *51*, 59.
- (42) Zwanzig, R.; Ailawadi, N. K. *Phys. Rev.* **1969**, *182*, 280–283.
- (43) Evertz, H. G. *Adv. Phys.* **2003**, *52*, 1–66.
- (44) Casella, G.; Berger, R. L. *Statistical inference*; Duxbury Press: Belmont, CA, 1998.