

Theory of binless multi-state free energy estimation with applications to protein-ligand binding

Zhiqiang Tan, Emilio Gallicchio, Mauro Lapelosa, and Ronald M. Levy

Citation: *J. Chem. Phys.* **136**, 144102 (2012); doi: 10.1063/1.3701175

View online: <http://dx.doi.org/10.1063/1.3701175>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v136/i14>

Published by the [American Institute of Physics](#).

Additional information on *J. Chem. Phys.*

Journal Homepage: <http://jcp.aip.org/>

Journal Information: http://jcp.aip.org/about/about_the_journal

Top downloads: http://jcp.aip.org/features/most_downloaded

Information for Authors: <http://jcp.aip.org/authors>

ADVERTISEMENT



HAVE YOU HEARD?

Employers hiring scientists
and engineers trust
physicstodayJOBS



<http://careers.physicstoday.org/post.cfm>

Theory of binless multi-state free energy estimation with applications to protein-ligand binding

Zhiqiang Tan,^{1,a)} Emilio Gallicchio,^{2,a)} Mauro Lapelosa,^{2,b)} and Ronald M. Levy²

¹*Department of Statistics, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA*

²*BioMaPS Institute for Quantitative Biology and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA*

(Received 22 December 2011; accepted 21 March 2012; published online 9 April 2012)

The weighted histogram analysis method (WHAM) is routinely used for computing free energies and expectations from multiple ensembles. Existing derivations of WHAM require observations to be discretized into a finite number of bins. Yet, WHAM formulas seem to hold even if the bin sizes are made arbitrarily small. The purpose of this article is to demonstrate both the validity and value of the multi-state Bennett acceptance ratio (MBAR) method seen as a binless extension of WHAM. We discuss two statistical arguments to derive the MBAR equations, in parallel to the self-consistency and maximum likelihood derivations already known for WHAM. We show that the binless method, like WHAM, can be used not only to estimate free energies and equilibrium expectations, but also to estimate equilibrium distributions. We also provide a number of useful results from the statistical literature, including the determination of MBAR estimators by minimization of a convex function. This leads to an approach to the computation of MBAR free energies by optimization algorithms, which can be more effective than existing algorithms. The advantages of MBAR are illustrated numerically for the calculation of absolute protein-ligand binding free energies by alchemical transformations with and without soft-core potentials. We show that binless statistical analysis can accurately treat sparsely distributed interaction energy samples as obtained from unmodified interaction potentials that cannot be properly analyzed using standard binning methods. This suggests that binless multi-state analysis of binding free energy simulations with unmodified potentials offers a straightforward alternative to the use of soft-core potentials for these alchemical transformations.

© 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.3701175>]

I. INTRODUCTION

The weighted histogram analysis method (WHAM) (Ref. 1) has emerged as an effective, general method for computing free energies and expectations from multiple ensembles, for example, at different temperatures or with different biasing potentials.^{2,3} There are a variety of ways to derive and understand WHAM, including the self-consistency approach^{1,4} and the maximum likelihood approach.^{2,5,6} However, all existing derivations in the computational physics literature involve discretizing observations into a finite number of bins in order to construct proper histograms. On the other hand, it has been recognized that WHAM formulas remain mathematically defined even if the bin sizes are made arbitrarily small or equivalently if the actual data instead of their discretizations are used (e.g., Sec. 8.3.2 of Ref. 4). However no formal account exists in the chemical physics literature for whether and under what conditions such a binless extension is valid.

At the same time, there have been extensive developments in the mathematical and statistical fields of theory

and methods leading to essentially the binless extension of WHAM.^{7–10} Shirts and Chodera¹¹ presented the binless method as the result of making the optimal choice among a large class of estimators,¹⁰ and called it the multi-state Bennett acceptance ratio method (MBAR) by the fact that the method reduces to the optimal Bennett acceptance ratio (BAR) (Refs. 12 and 13) in the case of only two ensembles. In this article, we discuss two statistical arguments to derive MBAR equations, in parallel to the self-consistency and maximum likelihood derivations already known for WHAM. Disseminating these concepts to the chemical physics community is helpful to better appreciate the theoretical foundations of the method and to highlight the connections between MBAR and WHAM, building on the established familiarity and expertise of practitioners with the latter.

To understand from a theoretical perspective the binless formulation of WHAM, an important quantity to consider is the measure of states, a non-negative measure from which the density of states is defined as the (Radon-Nikodym) derivative with respect to the counting or Lebesgue measure.¹⁴ From this perspective, the validity of MBAR as binless WHAM can be seen as follows. The measure of states can be consistently estimated in the sense that integrals of the density of states can be estimated with standard errors inversely proportional to the squared root of the sample size, even though the density of states, in general, cannot be pointwise estimated at the usual

^{a)} Authors to whom correspondence should be addressed. Electronic addresses: ztan@stat.rutgers.edu and emilio@biomaps.rutgers.edu.

^{b)} Present address: Department of Chemical and Biological Engineering, Drexel University, Philadelphia, Pennsylvania 19104, USA.

rate of standard errors. Examples of integrals include the partition function or the probability that the value of a system observable falls into a given bin.

We also provide a number of analytically and computationally useful results on MBAR from the statistical literature. The maximum likelihood derivation shows that the MBAR estimators can be obtained by minimizing a convex objective function, equivalent to solving a system of self-consistent equations. Various fast and reliable numerical algorithms have been developed for such optimization problems. For example, the trust region algorithm is globally convergent at the second order.¹⁵ Computing MBAR estimates by these optimization algorithms can be more effective than by algorithms in current use;¹¹ relevant comparisons have recently been reported in the context of solving WHAM equations.⁶

Statistical large-sample theory gives not only conditions under which the MBAR estimates are consistent and asymptotically normal but also formulas for asymptotic variance matrices, as the sample size grows to infinity. Although the theory can be applied to correlated data,⁷ the variance formulas are much simplified if the observations from each ensemble are independent.^{10,16} These formulas can be used for variance estimation provided that observations are subsampled to be approximately independent. Alternatively, as also done here, block bootstrapping¹⁷ can be used to estimate statistical uncertainties taking into account data correlations.

We illustrate the advantages of MBAR, based on the sampled values directly without binning, over conventional WHAM, with binning, on the calculation of absolute protein-ligand binding free energies by alchemical transformations. These calculations take various forms¹⁸ but they all consist of collecting samples from simulations distributed along a suitable thermodynamic path connecting the coupled and uncoupled states of the ligand-receptor complex. The path is parameterized by a progress parameter λ whereby, for example, $\lambda = 0$ corresponds to the uncoupled state and $\lambda = 1$ to the coupled state. The progress parameter λ , in turn, dials the parameters of a hybrid potential in such a way that at $\lambda = 1$ it represents the bound complex and at $\lambda = 0$ the ligand and receptor are not interacting.^{19–23}

In typical applications, the binding free energy is computed from the free energy differences between neighboring λ -states using only data collected at these states using pairwise exponential or more accurate BAR free energy estimators.^{13,24,25} These and analogous binding free energy estimators are notoriously affected by end point numerical instabilities near $\lambda = 0$, when the ligand and the receptor are nearly uncoupled. Under these conditions conformations are generated in which receptor and ligand atoms interpenetrate each other yielding very large interaction energies. These cause instabilities which are difficult to overcome unless specialized soft core potentials are employed.^{22,26–28}

Multi-state free energy estimation methods such as WHAM and MBAR (Refs. 3 and 11) are beginning to be employed in binding free energy calculations. The general idea behind these methods is to efficiently extract information from all of the intermediate states so as to achieve binding free energy estimates with smaller statistical variance. One example in this class of methods is the binding energy

distribution analysis method (BEDAM),^{29,30} which is employed here. The method is based on the analysis of samples of the binding energy of the complex (defined as the change in the effective potential energy of the complex with implicit solvation for bringing receptor and ligand from infinite separation to the bound conformation) without internal conformational rearrangements. In BEDAM, the end point problem with unmodified potentials is manifested with the occurrence near $\lambda = 0$ of large binding energy values spread over an extremely wide range, which, as we will show, makes the application of binning-based methods such as WHAM unfeasible. Binless methods such as MBAR do not suffer from the same issues and are shown to be able to treat data sets of this kind. This observation opens the possibility that using binless multi-state inference methods such as MBAR in conjunction with standard functional forms for the interactions potentials could be as effective as using modified soft-core potentials to circumvent the end point problem of binding free energy calculations.

II. THEORY AND METHODS

A. Setup

Consider a generalized ensemble whose Boltzmann probability density function is

$$\frac{1}{Z_\theta} e^{-\theta^T u(x)}, \quad (1)$$

where u is a column vector of d generalized energy functions of the configuration x of the system, θ is a column vector, also of length d , of corresponding coefficients, and

$$Z_\theta = \int e^{-\theta^T u(x)} dx \quad (2)$$

is the generalized configurational partition function in physics or the normalizing constant in statistics. Throughout, a superscript T denotes transpose so that for two vectors a and b each of length d ,

$$a^T b = \sum_{k=1}^d a_k b_k, \quad (3)$$

where a_k and b_k are vector elements, gives the inner product of a and b .

The foregoing notation is suitable to accommodate various applications. For example, the canonical ensemble at inverse temperature $\beta = 1/k_B T$ and potential energy function $U(x)$, is recovered by setting $d = 1$, $\theta = \beta$, and $u(x) = U(x)$ in Eq. (3). Similarly, the isothermal grand-canonical ensemble for a neat substance is recovered with $d = 2$, $\theta = (\beta, \beta\mu)$, and $u(x) = (U(x), N)$, where μ is the chemical potential and N the number of particles, so that $\theta^T u(x) = \beta(U(x) + \mu N)$. (Note that in this case the system configuration x includes atomic coordinates as well as the number of particles N , and Eq. (2) includes a summation over N .) A variety of ensembles commonly used in molecular simulations can also be accommodated by this notation. For example, each replica of a temperature replica exchange simulation is a canonical ensemble at the corresponding temperature as described above.

Free energy perturbation and “umbrella sampling” setups are obtained by setting the potential energy vector as $u(x) = (U_0(x), \omega_1(x), \dots, \omega_d(x))$, where $U_0(x)$ is the reference potential and $\omega_k(x)$ is the perturbation or umbrella potential in window k , and by setting the coefficient vector in window k as $\theta_k = (\beta, 0, \dots, 0, \beta, 0, \dots, 0)$, in which all elements are zero except for the first (corresponding to reference potential U_0) and the $(k+1)$ th element corresponding to the perturbation potential $\omega_k(x)$. For the binding free energy application illustrated below, we adopt the latter setup but with a simplified notation afforded by the particular linear form, $\omega_k(x) = \lambda_k b(x)$, of the perturbation (see Sec. III).

The notation introduced above is also useful to obtain compact expressions for thermodynamic observables. For example, the distribution and expectation of some observable $c(x)$ under Eq. (1) can be obtained in compact form (see, for example, Eq. (22)) by formally including $c(x)$ as a component of the generalized energy vector $u(x)$ with the corresponding coefficient in θ set to zero, so as to leave the physical system energy $\theta^T u(x)$ unchanged. In the following, we will implicitly assume that the generalized energy vector $u(x)$ includes components related to system observables.

Assume that simulations are conducted at m coefficient vectors θ_j ($j = 1, \dots, m$) and with the same energy vector $u(x)$. (Note that in this notation the dimensionality, d , of the θ and u vectors and the number of simulations, m , are, in general, distinct; for example, for temperature replica exchange $d = 1$ while m is the number of replicas.) Denoted by $\{x_{ji}: i = 1, \dots, n_j\}$ the set of configurations of size n_j obtained from the j th simulation, and denoted by $u_{ji} = u(x_{ji})$ the corresponding generalized energy vectors, which, as discussed above, may also include system observables. The total sample size is $n = \sum_{j=1}^m n_j$. Typically, the low-dimensional vectors u_{ji} are stored, instead of the high-dimensional, full configurations x_{ji} . For example, in the case of free energy perturbation calculations, $u_{ji} = (U_0(x_{ji}), \omega_1(x_{ji}), \dots, \omega_d(x_{ji}))$ contains the value of the perturbation potential, $\omega_j(x_{ji})$, corresponding to the same window as the observed conformation, x_{ji} , as well as values of the perturbation potential, $\omega_k(x_{ji})$, $k \neq j$, for all other windows for the same conformation. This specification of $u(x)$ well captures the type of data manipulations needed in multi-state inference methods such as WHAM and, as will be seen, the binless extension of WHAM.

Under Eq. (1), the induced probability density function of $u(x)$ at θ is of the form

$$\frac{1}{Z_\theta} \Omega(u) e^{-\theta^T u}, \quad (4)$$

where $\Omega(u)$, formally defined as

$$\Omega(u) = \int \delta(u(x) - u) dx \quad (5)$$

is a generalized density of states, which does not depend on θ . The partition function Z_θ can also be determined from $\Omega(u)$ as

$$Z_\theta = \int \Omega(u) e^{-\theta^T u} du. \quad (6)$$

The density function (1) and relationship (2) are replaced by Eqs. (4) and (6), respectively, when the data are reduced from x_{ji} to u_{ji} ($i = 1, \dots, n_j; j = 1, \dots, m$).

B. From WHAM to binless WHAM

The WHAM, first proposed by Ferrenberg and Swendsen,¹ can be used to compute various quantities of interest. The method involves constructing a histogram, $N_j(u)$, from each sample $\{u_{ji}: i = 1, \dots, n_j\}$, where $N_j(u)$ indicates the number of observations falling into a bin about u , for example, an interval or a rectangle if $u(x)$ is 1 or two-dimensional. Then $\Omega(u)$ is estimated by

$$\hat{\Omega}(u) \Delta u = \frac{\sum_{r=1}^M N_r(u)}{\sum_{r=1}^M n_r \hat{Z}_{\theta_r}^{-1} e^{-\theta_r^T u}}, \quad (7)$$

where the partition function estimators ($\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m}$) are defined by self-consistency according to Eq. (6)

$$\hat{Z}_{\theta_k} = \sum_u \frac{\sum_{r=1}^m N_r(u)}{\sum_{r=1}^m n_r \hat{Z}_{\theta_r}^{-1} e^{(\theta_k - \theta_r)^T u}} \quad (k = 1, \dots, m), \quad (8)$$

where the summation \sum_u is taken over all possible bins centered at u of size Δu . The estimators ($\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m}$) are determined up to a multiplicative constant. It is customary to pick a reference value, for example, Z_{θ_1} , and then estimate the ratios ($Z_{\theta_2}/Z_{\theta_1}, \dots, Z_{\theta_m}/Z_{\theta_1}$) from Eq. (8).

Again by relationship (6), the partition function Z_θ at any other parameter value is estimated by

$$\hat{Z}_\theta = \sum_u \frac{\sum_{r=1}^m N_r(u)}{\sum_{r=1}^m n_r \hat{Z}_{\theta_r}^{-1} e^{(\theta - \theta_r)^T u}}. \quad (9)$$

Furthermore, let $h(u)$ be a function of u , for example, a component of u , and denote by $\langle h \rangle_\theta$ the expectation of $h(u)$ under Eq. (4), that is, the expectation of $h(u(x))$ under Eq. (1). From Eqs. (4) and (7), the WHAM estimate \hat{h}_θ for $\langle h \rangle_\theta$ is

$$\hat{h}_\theta = \frac{1}{\hat{Z}_\theta} \sum_u h(u) \frac{\sum_{r=1}^m N_r(u)}{\sum_{r=1}^m n_r \hat{Z}_{\theta_r}^{-1} e^{(\theta - \theta_r)^T u}}. \quad (10)$$

This estimator depends on ($\hat{Z}_\theta, \hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m}$) up to a multiplicative constant, that is, only depends on the ratios ($\hat{Z}_\theta/\hat{Z}_{\theta_1}, \hat{Z}_{\theta_2}/\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m}/\hat{Z}_{\theta_1}$). It is interesting to note that the summation over bins in Eq. (10) can be equivalently expressed in terms of a weighted average over observations

$$\hat{h}_\theta = \sum_{ji} h(u_{ji}^b) F_{ji}(\theta), \quad (11)$$

where u_{ji}^b is a representative generalized energy of the bin containing u_{ji} , F_{ji} is the “WHAM weight” of u_{ji} that, by comparing Eqs. (10) and (11), is defined as

$$F_{ji}(\theta) = \frac{\hat{Z}_\theta^{-1}}{\sum_{r=1}^m n_r \hat{Z}_{\theta_r}^{-1} e^{(\theta - \theta_r)^T u_{ji}^b}} = \frac{1}{\hat{Z}_\theta} e^{-\theta^T u_{ji}^b} G_{ji} \quad (12)$$

and

$$G_{ji} = \frac{1}{\sum_{r=1}^m n_r \hat{Z}_{\theta_r}^{-1} e^{-\theta_r^T u_{ji}^b}} \quad (13)$$

is the θ -independent component of the WHAM weight $F_{ji}(\theta)$ for each observation.

Equation (11) states that the expectation value of any observable can be obtained by attaching a statistical weight $F_{ji}(\theta)$ to each observation u_{ji} which depends on the bin to which it is assigned. An obvious simplification is to express the WHAM estimate of $\langle h \rangle_\theta$ and the WHAM weights [Eqs. (11) and (12)] in terms of the actual observations u_{ji} rather than their closest bin representatives u_{ji}^b . This idea, which has been noted before without formal justification in the computational physics literature,^{4,31} leads naturally to a binless extension of WHAM. A closely related formalism has been developed in statistics for computing normalizing constants.⁷⁻¹⁰ Although this method can be derived by various statistical arguments, it is essentially an extension of WHAM without binning data. Below we give a formal derivation of the binless method by importance weighting and self-consistency.

To understand binless WHAM, it is useful to introduce the concept of the measure G defined by

$$dG = \Omega(u) du, \quad (14)$$

that is, $G(A) = \int_A \Omega(u) du$ for every measurable set A of u . Informally, Eq. (14) says that for an infinitesimal bin about u of size du , the weight assigned under G is $\Omega(u) du$. Thereafter G is called the measure of states. The concept of measure can be used to reformulate the ideas developed above. Denote by F_θ the probability distribution of $u(x)$ under (1), that is, the probability distribution with density function (4). Then, from Eqs. (4) and (14), F_θ is related to G as

$$dF_\theta = \frac{1}{Z_\theta} e^{-\theta^T u} \Omega(u) du = \frac{1}{Z_\theta} e^{-\theta^T u} dG, \quad (15)$$

that is, $F_\theta(A) = Z_\theta^{-1} \int_A e^{-\theta^T u} dG$ for every measurable set A of u . For an infinitesimal bin about u of size du , the probability assigned under F_θ is the density function (4) times du and hence is $Z_\theta^{-1} e^{-\theta^T u}$ times the weight assigned under G . The partition function Z_θ by Eq. (6) can then be expressed as

$$Z_\theta = \int e^{-\theta^T u} dG. \quad (16)$$

See, for example, Ref. 14 for discussion of measure-theoretic concepts.

The pooled data $\{u_{ji}: i = 1, \dots, n_j, j = 1, \dots, m\}$ can be regarded as an approximate sample from the mixture distribution, F_* , whose components are $(F_{\theta_1}, \dots, F_{\theta_m})$ with proportions $(n_1/n, \dots, n_m/n)$. (Note that the pooled data are not strictly an independent and identically distributed sample from F_* , which would involve randomly selecting a distribution F_{θ_r} with probability n_r/n ($r = 1, \dots, m$), simulating one observation from F_{θ_r} and then repeating this process for n times. The numbers of observations from $(F_{\theta_1}, \dots, F_{\theta_m})$ would be random, instead of being fixed at (n_1, \dots, n_m) . To highlight main ideas, this difference is ignored in the derivation below. The resulting estimators are, however, evaluated in Sec. II D without making this simplification.) Then, in anal-

ogy with Eq. (15), F_* is related to G as

$$dF_* = \left\{ \sum_{r=1}^m \frac{n_r}{n} Z_{\theta_r}^{-1} e^{-\theta_r^T u} \right\} \Omega(u) du = \left\{ \sum_{r=1}^m \frac{n_r}{n} Z_{\theta_r}^{-1} e^{-\theta_r^T u} \right\} dG. \quad (17)$$

For an infinitesimal bin about u of size du , the probability assigned under F_* is the expression in the curly bracket times the weight assigned under G . Dividing both sides of Eq. (17) by the quantity in the curly brackets gives

$$dG = \left\{ \sum_{r=1}^m \frac{n_r}{n} Z_{\theta_r}^{-1} e^{-\theta_r^T u} \right\}^{-1} dF_*. \quad (18)$$

For an infinitesimal bin about u of size du , the weight assigned under G is the inverse of the quantity in the curly brackets times the probability assigned under F_* .

Relationship (18) can be used for estimating G from the pooled data by importance weighting. Recall that the pooled data form an approximate sample from F_* . Then F_* can be estimated by the empirical distribution \hat{F}_* for which each observation u_{ji} is assigned the probability n^{-1} . By Eq. (18), the resulting estimator \hat{G} is a discrete measure for which each observation u_{ji} is assigned the weight

$$\hat{G}(u_{ji}) = \frac{1}{\sum_{r=1}^m n_r \hat{Z}_{\theta_r}^{-1} e^{-\theta_r^T u_{ji}}}, \quad (19)$$

where $(\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m})$ are defined by self-consistency according to Eq. (16)

$$\begin{aligned} \hat{Z}_{\theta_k} &= \sum_{j=1}^m \sum_{i=1}^{n_j} e^{-\theta_k^T u_{ji}} \hat{G}(u_{ji}) \\ &= \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{1}{\sum_{r=1}^m n_r \hat{Z}_{\theta_r}^{-1} e^{(\theta_k - \theta_r)^T u_{ji}}} \quad (k = 1, \dots, m). \end{aligned} \quad (20)$$

Formulas (19) and (20) provide a binless extension of Eqs. (7) and (8) in WHAM.

By again relationship (16), the partition function Z_θ at any other parameter value is estimated by

$$\begin{aligned} \hat{Z}_\theta &= \sum_{j=1}^m \sum_{i=1}^{n_j} e^{-\theta^T u_{ji}} \hat{G}(u_{ji}) \\ &= \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{1}{\sum_{r=1}^m n_r \hat{Z}_{\theta_r}^{-1} e^{(\theta - \theta_r)^T u_{ji}}}. \end{aligned} \quad (21)$$

The expectation $\langle h \rangle_\theta$ is by definition $Z_\theta^{-1} \int h(u) e^{-\theta^T u} dG$ and hence estimated by

$$\begin{aligned} &\frac{1}{\hat{Z}_\theta} \sum_{j=1}^m \sum_{i=1}^{n_j} h(u_{ji}) e^{-\theta^T u_{ji}} \hat{G}(u_{ji}) \\ &= \frac{1}{\hat{Z}_\theta} \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{h(u_{ji})}{\sum_{r=1}^m n_r \hat{Z}_{\theta_r}^{-1} e^{(\theta - \theta_r)^T u_{ji}}}. \end{aligned} \quad (22)$$

Formulas (21) and (22) provide a binless extension of Eqs. (9) and (10) in WHAM. In addition, we see that the

WHAM weights (13), identified heuristically earlier, coincide (except for the difference between u_{ji}^b vs. u_{ji}) with the discrete measure with weights (19) derived from the statistical theory sketched out above. Therefore, the binless formulation of WHAM, while it appears straightforward, is nevertheless rooted on fundamental statistical concepts.

It is worth emphasizing that the binless method, like WHAM, can be used not only to estimate partition functions Z_θ and equilibrium expectations $\langle h \rangle_\theta$, but also to estimate equilibrium distributions F_θ . Recall that $u(x)$ is in general a vector of multiple components and F_θ is the joint distribution of those components under Eq. (1). By relationship (15), F_θ is estimated by a discrete distribution \hat{F}_θ on the pooled data with probabilities

$$\hat{F}_\theta(u_{ji}) = \frac{\hat{Z}_\theta^{-1}}{\sum_{r=1}^m n_r \hat{Z}_{\theta_r}^{-1} e^{(\theta - \theta_r)^\top u_{ji}}}. \quad (23)$$

In other words, F_θ is approximated by attaching weight (23) to each observation u_{ji} in the pooled data, where the weights sum up to 1 by Eq. (21). As a result of this approximation, the marginal distribution of $h(u(x))$ under Eq. (1) is approximated by attaching the same weight (23) to $h(u_{ji})$ for each u_{ji} in the pooled data. Then the expectation $\langle h \rangle_\theta$ is approximated as before (Eq. (22)) by a weighted average of the form $\sum_{j=1}^m \sum_{i=1}^{n_j} h(u_{ji}) \hat{F}_\theta(u_{ji})$.

The above approximation to the marginal distribution of $h(u(x))$ under Eq. (1) can be visualized as a weighted histogram with suitable bins. The height of each bin is the sum of $\hat{F}_\theta(u_{ji})$ such that $h(u_{ji})$ falls into the bin for u_{ji} in the pooled data. The histogram can be normalized into a probability density plot, where the height of each bin is divided by the bin size. If $\theta = \theta_k$ for some k , this weighted histogram based on the pooled data provides a better approximation than the raw histogram of $h(u_{ki})$ based on the observations u_{ki} from F_{θ_k} only. On the other hand, a comparison of these two histograms can be used to assess goodness of simulations. A substantial discrepancy between the two histograms suggest that the quality of simulations is questionable, that is, the simulated data are actually not distributed according to Eq. (4).

C. Maximum likelihood

We describe a derivation of binless WHAM by the method of nonparametric maximum likelihood taking G as an infinite-dimensional unknown parameter.⁹ The likelihood of the j th sample from F_{θ_j} is by Eq. (15)

$$L_j = \prod_{i=1}^{n_j} \left\{ \frac{1}{Z_{\theta_j}} e^{-\theta_j^\top u_{ji}} G(u_{ji}) \right\}, \quad (24)$$

where $G(u_{ji})$ is the mass assigned to the singleton u_{ji} under G , and $Z_{\theta_j} = \int e^{-\theta_j^\top u} dG$, a functional of G , by Eq. (16). The likelihood of the pooled sample is then $L = \prod_{j=1}^m L_j$. The method of nonparametric maximum likelihood is to find \hat{G} which maximizes the likelihood L among all possible non-negative measures including discrete measures.

There are two steps to find the maximum likelihood estimator \hat{G} . First, it is sufficient to restrict our search to discrete

measures supported on the set of pooled data $\{u_{ji}: j = 1, \dots, n_j, j = 1, \dots, m\}$. If a positive mass is assigned under G to any set outside the pooled data, then relocating the mass evenly to each observation in the pooled data only increases L . Second, for a discrete measure G , put $w_{ji} = G(u_{ji})$. The likelihood at G is

$$L = \prod_{j=1}^m \prod_{i=1}^{n_j} \left\{ \frac{1}{Z_{\theta_j}} e^{-\theta_j^\top u_{ji}} w_{ji} \right\}, \quad (25)$$

where $Z_{\theta_r} = \sum_{j=1}^m \sum_{i=1}^{n_j} e^{-\theta_r^\top u_{ji}} w_{ji}$ for $r = 1, \dots, m$. Taking the log of the likelihood gives

$$\log L = \left\{ \sum_{j=1}^m \sum_{i=1}^{n_j} \log w_{ji} - \sum_{j=1}^m n_j \log Z_{\theta_j} \right\} - \sum_{r=1}^m \sum_{i=1}^{n_j} \theta_r^\top u_{ji}. \quad (26)$$

The term outside the curly brackets does not depend on w_{ji} and can be ignored. Taking the partial derivative of $\log L$ with respect to w_{ji} gives

$$\frac{1}{w_{ji}} - \sum_{r=1}^m n_r \frac{e^{-\theta_r^\top u_{ji}}}{Z_{\theta_r}} = 0 \quad (27)$$

or

$$w_{ji} = \frac{1}{\sum_{r=1}^m n_r Z_{\theta_r}^{-1} e^{-\theta_r^\top u_{ji}}}, \quad (28)$$

which leads to the basic formulas (19) and (20). Furthermore, substituting the expression of w_{ji} into the term inside the curly bracket in Eq. (26) yields

$$- \sum_{j=1}^m \sum_{i=1}^{n_j} \log \left\{ \sum_{r=1}^m n_r Z_r^{-1} e^{-\theta_r^\top u_{ji}} \right\} - \sum_{r=1}^m n_r \log Z_r, \quad (29)$$

which is a function of (Z_1, \dots, Z_m) only. This function multiplied by $-n^{-1}$ and then subtracted by $\log n$ gives the function κ below (Eq. (31)).

It is interesting that the maximum likelihood estimator \hat{G} is always a discrete measure, even though the actual measure G is not. This discrete approximation of G by \hat{G} serves precisely our computational purpose. A complication is that even though there is a general statistical theory to justify the method of maximum likelihood with a finite-dimensional unknown parameter, the validity of the estimators obtained by the method of nonparametric likelihood need to be established on a case-by-case basis. Fortunately, a statistical theory of binless WHAM has been rigorously developed in statistics, and is reviewed in Sec. II D.

The foregoing derivation takes the measure of states G as the underlying unknown parameter. Equivalently, the method of nonparametric maximum likelihood can be applied with a reparameterization taking F_{θ_0} as the unknown parameter for some fixed, reference value θ_0 . By Eq. (15), F_θ is related to F_{θ_0} as

$$dF_\theta = \frac{Z_{\theta_0}}{Z_\theta} e^{-(\theta - \theta_0)^\top u} dF_{\theta_0}. \quad (30)$$

By invariance of maximum likelihood under reparameterization, the resulting estimator of F_{θ_0} is the same as Eq. (23)

with θ set to θ_0 . Furthermore, the formulas (20)–(22) remain the same as before. This derivation is essentially an extension of the derivation of WHAM by Bartels and Karplus, and Gallicchio *et al.*^{2,5}

D. Statistical theory

As seen from Sec. II B, the estimators in binless WHAM are similar to those in WHAM, but based on the actual data without binning. While this construction seems heuristically easy, a central issue is to evaluate statistical and computational properties of binless WHAM. We point out a number of useful results which demonstrate the usefulness of the binless formulation of WHAM, by drawing on related statistical work. Although there are results applicable to correlated data,⁷ we assume for simplicity that $\{u_{ji}; i = 1, \dots, n_j; j = 1, \dots, m\}$ are independent.

First, the estimators $(\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m})$ are defined by Eq. (20), a system of nonlinear equations. Remarkably, an equivalent characterization is that $\log(\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m})$ are jointly a minimizer of the criterion function^{7,16}

$$\begin{aligned} \kappa(\log z_1, \dots, \log z_m) &= \frac{1}{n} \sum_{j=1}^m \sum_{i=1}^{n_j} \log \left\{ \sum_{r=1}^m \frac{n_r}{n} z_r^{-1} e^{-\theta_r^T u_{ji}} \right\} \\ &\quad + \sum_{r=1}^m \frac{n_r}{n} \log z_r. \end{aligned} \quad (31)$$

See Sec. II C above for the derivation of κ by maximum likelihood. The function κ is invariant under translation: $\kappa(a + \log z_1, \dots, a + \log z_m) = \kappa(\log z_1, \dots, \log z_m)$ for an arbitrary constant a , in agreement with the fact that $\log(\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m})$ are only determined up to an additive constant. Moreover, κ is bounded from below, by application of Jensen's inequality to the log of the term in the curly brackets

$$\begin{aligned} \log \left\{ \sum_{r=1}^m \frac{n_r}{n} z_r^{-1} e^{-\theta_r^T u_{ji}} \right\} &\geq \sum_{r=1}^m \frac{n_r}{n} \log \{ z_r^{-1} e^{-\theta_r^T u_{ji}} \} \\ &= - \sum_{r=1}^m \frac{n_r}{n} \log z_r - \sum_{r=1}^m \frac{n_r}{n} \theta_r^T u_{ji}. \end{aligned} \quad (32)$$

Finally, if one of $(\log z_1, \dots, \log z_m)$ is fixed, for example $\log z_1 = 0$, then κ is strictly convex.¹⁶ The convexity can be directly shown by the fact that

$$\sum_{r=1}^m \frac{n_r}{n} z_r^{-1} e^{-\theta_r^T u_{ji}} \quad (33)$$

is convex, and consequently the log of this term is also convex in $(\log z_1, \dots, \log z_m)$. Therefore, $\log(\hat{Z}_{\theta_2}/\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m}/\hat{Z}_{\theta_1})$ can be obtained as a unique minimizer of $\kappa(0, \log z_2, \dots, \log z_m)$. This approach of minimizing a convex function can be more effective than solving the system of nonlinear equations (20) by the self-consistency or the Newton-Raphson algorithm.¹¹ See Appendix A for details.

Second, the estimators $(\hat{Z}_{\theta_2}/\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m}/\hat{Z}_{\theta_1})$ are always consistent (that is, converge in probability to the true

values) and asymptotically normally distributed as the sample size n_j tends to infinity and n_j/n is fixed for each j .^{10,16} The connectedness condition required for the general result of Gill *et al.*¹⁶ and Tan¹⁰ is satisfied here because the weighting function $e^{-\theta^T u}$ is positive. Moreover, the estimator $\hat{Z}_\theta/\hat{Z}_{\theta_1}$ is consistent and asymptotically normally distributed provided that the variance under F_* of the density ratio of F_θ over F_* is finite

$$\int \left\{ \sum_{r=1}^m \frac{n_r}{n} z_{\theta_r}^{-1} e^{(\theta - \theta_r)^T u} \right\}^{-2} dF_* < \infty. \quad (34)$$

Similarly, the estimator of $\langle h \rangle_\theta$ is consistent and asymptotically normally distributed provided that the variance under F_* of $h(u)$ times the density ratio of F_* over F_* is finite

$$\int h^2(u) \left\{ \sum_{r=1}^m \frac{n_r}{n} z_{\theta_r}^{-1} e^{(\theta - \theta_r)^T u} \right\}^{-2} dF_* < \infty. \quad (35)$$

These conditions require that the mixture ‘‘umbrella’’ distribution F_* should provide sufficient coverage of F_θ , so that observations from F_* can be weighted by the density ratio of F_θ over F_* to estimate F_θ . Therefore, interpolation is in general valid, but extrapolation needs to be considered more carefully. For example, for the application in Sec. III, it is important to obtain observations from the end thermodynamic states, in addition to intermediate states, in order to estimate the free energy differences between them. Obtaining observations from thermodynamic states however close to the end states, but not at end states, would require extrapolation whereby condition (34) would be difficult to verify.

Third, the asymptotic variance matrix of $(\hat{Z}_{\theta_2}/\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m}/\hat{Z}_{\theta_1})$ and $\hat{Z}_\theta/\hat{Z}_{\theta_1}$ jointly can be consistently estimated without using any generalized inverse such as the Moore-Penrose inverse.¹⁰ This approach differs from that of Kong *et al.*⁹ and Shirts and Chodera¹¹ based on the asymptotic variance matrix of $(\hat{Z}_{\theta_1}, \hat{Z}_{\theta_2}, \dots, \hat{Z}_{\theta_m})$, which necessarily involves use of generalized inverses. Similarly, the asymptotic variance of the estimator of $\langle h \rangle_\theta$ can be consistently estimated. The resulting variance formula is appropriate even when $h(u)$ is not always non-negative, in contrast with Shirts and Chodera (Sec. IV of Ref. 11). See Appendix B for details.

Fourth, when $m = 2$, the estimator $\hat{Z}_{\theta_2}/\hat{Z}_{\theta_1}$ is equivalent to Bennett's optimal acceptance ratio method (BAR),¹² which attains the smallest asymptotic variance among bridge sampling estimators of the form^{8,13}

$$\frac{n_1^{-1} \sum_{i=1}^{n_1} \alpha(u_{1i}) e^{-(\theta_2 - \theta_1)^T u_{1i}}}{n_2^{-1} \sum_{i=1}^{n_2} \alpha(u_{2i})}, \quad (36)$$

where $\alpha(\cdot)$ is an arbitrary function, for example, $\alpha(u) = \min(e^{-(\theta_1 - \theta_2)^T u}, 1)$. In general, the estimators $(\hat{Z}_{\theta_2}/\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m}/\hat{Z}_{\theta_1})$ and $\hat{Z}_\theta/\hat{Z}_{\theta_1}$ jointly attain the smallest asymptotic variance matrix in the order on positive-definite matrices among a class of extended bridge sampling estimators based on Eq. (36).^{10,11} Similarly, the estimator of $\langle h \rangle_\theta$ attains the smallest variance among corresponding extended bridge sampling estimators. For this reason, the

binless method was called the multi-state Bennett acceptance ratio method (MBAR) by Shirts and Chodera.¹¹

III. APPLICATION: ESTIMATION OF BINDING FREE ENERGIES

This section illustrates the application of the binless method using both the MBAR software (as developed by Shirts and Chodera¹¹) and our computational implementation based on Sec. II D (referred to as unbinned WHAM or UWHAM), to the estimation of protein-ligand binding free energies. As we will show, due to the wide range of values of the binding energies involved, it is difficult to apply the conventional WHAM binning method to this problem unless soft-core potentials are employed. In contrast, the binless approach yields consistent results in all cases.

The binding free energy measures the propensity of a receptor R to be associated in solution with a ligand L . The binding free energy is by definition the difference between the free energy of the receptor-ligand complex and the free energy of the dissociated receptor and ligand. In this work, binding free energies are estimated by simulation in the context of the BEDAM,²⁹ which, in the present formalism can be summarized as follows.

Working within the implicit solvent representation, the potential energy of a conformation x of the complex can be written as^{18,29}

$$U_{RL}(x) = U_R(x) + U_L(x) + b(x), \quad (37)$$

where U_R and U_L are the potential energies of the dissociated receptor and ligand in solution and $b(x)$ is the binding energy of conformation x of the complex, defined as the change in potential energy for bringing into contact the receptor and ligand from infinite separation without intramolecular conformational rearrangements. Based on the notation developed in Sec. II A, we recognize that the coupled (ligand and receptor fully interacting) and decoupled (non interacting ligand and receptor) ensembles can be cast in the form of the generalized ensemble representation of Eqs. (1)–(3) with a two-dimensional potential energy function vector $u = (U_0, b)$ where

$$U_0(x) = U_R(x) + U_L(x) \quad (38)$$

is the reference potential energy function corresponding to the uncoupled state and b is the binding energy function. Using Eq. (37), the potential energy of the decoupled state corresponds to the coefficient vector $\theta_{\text{dcpld}} = (\beta, 0)$ and the one for the coupled ensemble is $\theta_{\text{cpld}} = (\beta, \beta)$. The binding free energy is then given by the ratio of the corresponding partition functions $Z_{\theta_{\text{cpld}}}$ and $Z_{\theta_{\text{dcpld}}}$:

$$\Delta G_b = -kT \log \frac{Z_{\theta_{\text{cpld}}}}{Z_{\theta_{\text{dcpld}}}}. \quad (39)$$

Note that the observable standard binding free energy also includes a standard state concentration-dependent term^{18,19,29} which, being constant among the systems investigated, is included in the results³⁰ but not further discussed in this work.

A series of intermediate states $k = 1, \dots, m$ are introduced with potential energies

$$U_k(x) = U_0(x) + \lambda_k b(x), \quad (40)$$

where $\lambda_1 = 0$ corresponds to the decoupled state and $\lambda_m = 1$ corresponds to the coupled state. The intermediate states with λ_i between 0 and 1 serve as interpolating states in which receptor and ligand partially interact to connect, in a free energy sense, the two end states.²⁵ In general, as stated in Sec. II A, a $(m + 1)$ -dimensional potential energy vector $u = (U_0(x), \omega_1(x), \dots, \omega_m(x))$, with $\omega_k(x) = \lambda_k b(x)$, and corresponding $(m + 1)$ -dimensional θ vectors are necessary to describe this collection of ensembles. However in this case, taking advantage of the particular linear expression of $\omega_k(x)$, it is convenient to collapse the λ_k dependence on the coefficient vector θ so as to lower the dimensionality of the generalized energy vector. By doing so, each of the states corresponds to a two-dimensional θ vector of the form $\theta_k = (\beta, \beta\lambda_k)$ which multiplies the potential energy vector $u = (U_0, b)$ introduced above to yield, by means of Eq. (3) the potential energy functions in Eq. (40).

The partition function of each state is computed from Eq. (22) setting $Z_{\theta_{\text{dcpld}}} = Z_{\theta_1} = 1$. Using Eq. (3) and the above, it is easy to see that the term $(\theta_k - \theta_r)^T u_{ji}$ in Eq. (20) in this case simplifies to

$$(\theta_k - \theta_r)^T u_{ji} = \beta(\lambda_k - \lambda_r) b_{ji}, \quad (41)$$

which does not include the total reference potential energy U_0 and depends only on the binding energy b_{ji} of the i th sampled conformation x_{ji} from a simulation at $\lambda = \lambda_j$. Analogously, it is straightforward to show that Eq. (31) simplifies to

$$\begin{aligned} \kappa(\log z_1, \dots, \log z_m) &= c + \frac{1}{n} \sum_{j=1}^m \sum_{i=1}^{n_j} \log \left\{ \sum_{r=1}^m \frac{n_r}{n} z_r^{-1} e^{-\beta\lambda_r b_{ji}} \right\} \\ &+ \sum_{r=1}^m \frac{n_r}{n} \log z_r, \end{aligned} \quad (42)$$

where c is a constant that depends only on the observations of U_0 and does not affect the position of the minimum.

Similarly, in the denominator of the WHAM equation (Eq. (8)), the $(\theta_k - \theta_r)^T u$ term reduces to $\beta(\lambda_k - \lambda_r)b$, which depends only on the binned value b of the binding energy. Furthermore expressing Eq. (8) as

$$\begin{aligned} \hat{Z}_{\lambda_k} &= \sum_{U_0} \sum_b \frac{\sum_{r=1}^m N_r(U_0, b)}{\sum_{r=1}^m n_r \hat{Z}_{\lambda_r}^{-1} e^{\beta(\lambda_k - \lambda_r)b}} \\ &= \sum_b \frac{\sum_{r=1}^m N_r(b)}{\sum_{r=1}^m n_r \hat{Z}_{\lambda_r}^{-1} e^{\beta(\lambda_k - \lambda_r)b}}, \end{aligned} \quad (43)$$

we see that the two-dimensional histogram $N_r(u) = N_r(U_0, b)$ can be replaced by the one-dimensional marginal histogram $N_r(b) = \sum_{U_0} N_r(U_0, b)$ of the binding energy. Consequently, in both the WHAM and MBAR calculations that follow it has been sufficient to collect only the binding energy samples from the molecular simulations.

Binding energies are collected from Hamiltonian replica exchange all-atom molecular dynamics simulations of the protein complexes as described^{29,30,32} for a series of λ values from 0 (decoupled state) to 1 (coupled state). The binding

energy data is then fed into Eq. (8), using binning, or Eq. (42), without binning, to compute the ratios of partition functions and ultimately the binding free energy from Eq. (39). See below for a description of the biological systems and simulation settings.

A. WHAM estimates with binning

The distributions of binding energies depend critically on the λ value at which they are obtained. At $\lambda = 1$, when the ligand and the receptor fully interact, binding energies are typically centered around favorable (negative) values (see Fig. 1). In contrast at $\lambda = 0$, in the absence of receptor-ligand interactions, the ligand is likely to sample conformations with unfavorable clashes between receptor and ligand atoms, corresponding to large unfavorable (positive) values of the binding energy (see Fig. 2). In principle, because the Lennard-Jones and Coulomb interatomic potentials tend to infinity at zero interatomic separation, there is no finite upper limit to the range of binding energies that can be observed. As shown here, this causes major difficulties for the binning of binding energy data to be used in conjunction with WHAM (Eq. (8)), since in this case the binding energy samples are spread out very sparsely in a region spanning many orders of magnitude which is impossible to bin reliably without using very wide bins leading to large integration errors.

Conventional wisdom dictates that the number of bins should be small enough so that each bin contains more than a few samples so as to minimize statistical noise in the resulting histograms. On the other hand, the binning resolution should be sufficiently fine so as to avoid significant integration errors when replacing the integral in Eq. (6) with the summation over bins in Eq. (8). It is not always clear how to balance these opposing requirements especially when, as in this case, the range of values to be binned is unbounded. Of course, as shown above, we now know that it is justifi-

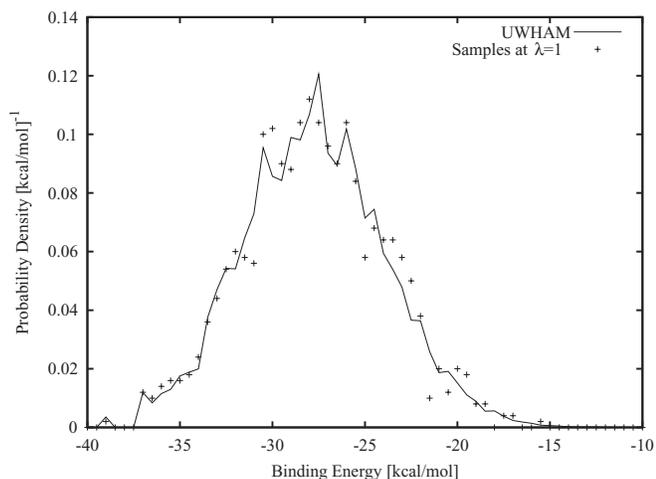


FIG. 1. Computed probability density at $\lambda = 1$, $p_1(b)$, for the complex with ligand 6 with the unmodified potential.³⁰ The line represents the UWHAM estimate from the data collected from all λ -replicas. The crosses correspond to the probability density computed from the histogram of the binding energy data at only $\lambda = 1$. Good correspondence between the two densities is observed.

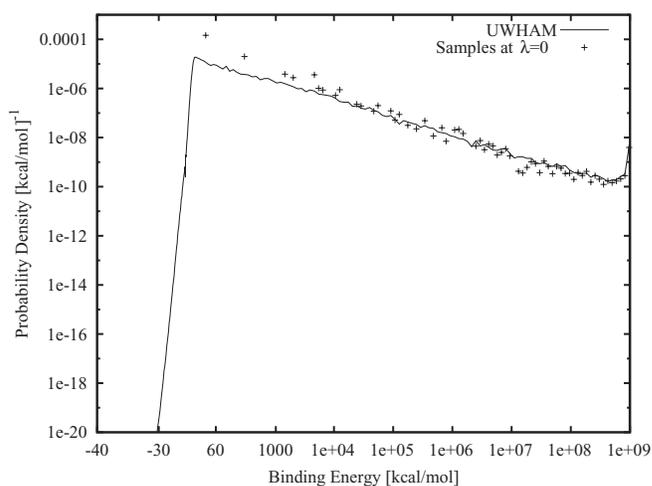


FIG. 2. Computed probability density at $\lambda = 0$, $p_0(b)$, for the complex with ligand 6 with the unmodified potential.³⁰ The line represents the UWHAM estimate from the data collected from all λ -replicas. The crosses correspond to the probability density computed from the histogram of the binding energy data at only $\lambda = 0$. There is good correspondence between the two densities in the range explored by the $\lambda = 0$ replica. The binding energy grid used for this plot has 200 bins, equally spaced (0.5 kcal/mol bin sizes) for negative binding energies and exponentially increasing spacing for positive values to up to 10^9 kcal/mol. Even though $p_0(u)$ is predicted to be maximal at approximately $u = 20$ kcal/mol, it is rare to observe binding energies in that range because of the small integrated cumulative probability at low binding energies (note the logarithmic layout of the binding energy axis). The UWHAM estimate instead extends to as low as -40 kcal/mol (the lowest observed sample at all λ 's) with an estimated probability density on the order 10^{-28} kcal/mol⁻¹ (not shown for clarity).

able to increase the number of bins indefinitely, reaching the limit where the WHAM formula is indistinguishable from the MBAR formula, which is based on the sampled values directly without binning.

In Table I we report WHAM binding free energy estimates for the complex with ligand 2 (see below for a description of protein-ligand complexes) varying the number of bins. In these calculations a uniform grid spacing has been used in the favorable binding energy range and an exponentially increasing bin spacing for unfavorable binding energies. We see that the results change significantly as the grid resolution is increased. With fewer bins and coarser bin widths WHAM under-predicts binding affinities. As the number of bins is increased the estimate of the binding free energy gets closer to the limiting value of $\Delta G_b \simeq -2.2$ kcal/mol obtained

TABLE I. WHAM results for ligand 2 with the unmodified potential varying the number of bins.

N_{bins}	ΔG_b^a
100	3.50
120	7.62
150	0.81
200	-0.07
250	-0.51
1000	-1.45
∞^b	-2.21

^aIn kcal/mol.

^bMBAR/UWHAM result from Table III.

TABLE II. WHAM results for ligand 2 with the unmodified potential changing the energy limit of the last bin with $N_{\text{bins}} = 250$.

b_c^a	ΔG_b^a
20	-5.92
80	-5.23
200	-4.88
10^7	-1.46
10^9	-0.51

^aIn kcal/mol.

with, effectively, an unlimited number of bins (see results in Table III). These results indicate that binning the unfavorable range of binding energies, even with a thousand bins, leads to large errors.

1. Data censoring bias

One simple way to circumvent the need for binning a very large range of binding energies is to terminate the binning at a large but finite grid value b_c and assign all of the samples with values larger than the maximum to this last bin (data censoring).²⁹ This approach intuitively appears valid based on the argument that unfavorable binding energies much larger than thermal energy are equally unlikely to be sampled by the complex regardless of their specific value. However, as shown in Table II, this leads to significant bias in the binding free energy estimates. The estimates in Table II are obtained for ligand 2 with 250 bins and the binning limit, b_c , indicated. The results show that using a small b_c (but still much larger than binding energy values achievable at $\lambda = 1$ at standard temperature) leads to overestimation of binding affinities, and that the bias progressively shifts to less negative values as b_c is increased – but overshoots the correct value because, with a fixed bin size, the bins became too coarse as the the energy limit of the last bin increases.

The origin of the data censoring bias can be understood in general terms by recognizing that it amounts to assuming that the potential energy of the system is bounded although no limit is actually present. In other words, the data is being analyzed with a statistical model inconsistent with the system that generated the data. To understand the effect in numerical terms consider the denominator in Eq. (43) for $b = b_c \gg kT$. When $\lambda_k = 1$, the quantities $\exp[\beta(1 - \lambda_r)b_c]$ are all positive and some are very large. It follows that the sum in the denominator is large and the contribution to $Z_{\lambda=1}$ from the bin at b_c is negligible regardless of the specific value of b_c . A similar conclusion can be reached for any large value of λ_k . However for small values of λ_k such that $\beta\lambda_k b_c \simeq 1$ the values of the quantities $\exp[\beta(\lambda_k - \lambda_r)b_c]$ can vary significantly depending on the specific value of b_c . This leads to incorrect estimates of the free energy profile at small λ 's and, in turn, of the total free energy change.

2. Soft-core binding energy function

Another approach that we have explored in this work is to, in effect, prevent the generation of large binding energies

by adopting a soft core potential in the simulations. Soft core potentials are commonly used to attempt to improve the convergence of free energy calculations.^{26,27,33} In this work, a soft core potential is introduced in terms of a modified binding energy function $b'(x)$ of the form^{30,34}

$$b'(x) = \begin{cases} b_{\text{max}} \tanh[b(x)/b_{\text{max}}] & b(x) > 0 \\ b(x) & b(x) \leq 0 \end{cases}, \quad (44)$$

where b_{max} is some large positive value, set in this work to either 10^3 kcal/mol (soft core) or 10^9 kcal/mol (referred to below as the “unmodified” binding energy function). The modified binding energy function $b'(x)$ serves the purpose of capping the maximum value of the binding energy while leaving unchanged the values of favorable binding energies. Here it is used throughout in the molecular simulations and the statistical analysis in place of the actual binding energy function. The potential energy of $\lambda = 0$ state is equal to $u_0(x)$ (see Eq. (40)) and is unaffected by the binding energy function. Furthermore, the $\lambda = 1$ state with the soft core binding energy function is virtually indistinguishable from the original one as large positive values of the binding energy are never sampled during the simulation. We conclude therefore that the free energy difference between the $\lambda = 0$ and $\lambda = 1$ states (that is the binding free energy) is not significantly affected by the introduction of the modified binding energy function.²⁸ Indeed, as shown below, we obtain statistically indistinguishable binding free energy estimates with the two binding energy functions, with any small difference possibly attributable to other factors, such as insufficient equilibration and convergence.

Table III reports WHAM binding energy estimates obtained with the soft core binding energy function (Eq. (44)) with $b_{\text{max}} = 10^3$ kcal/mol. These calculations employed a grid with 250 bins similar to the one used above for the unmodified binding energy function (Table I) but extending only up to $b = b_{\text{max}}$ since no samples are present beyond this value. The limited extent of the range of binding energies makes it possible to select a sufficiently fine binning grid with a reasonable number of bins. Because the binding free energy estimates so obtained are in agreement with the MBAR/UWHAM estimates (see below) obtained with the unmodified potential function, we conclude that the soft core WHAM results indeed reflect the correct binding free energies for this system. Conversely, based on the results above, we conclude that application of WHAM to the data with the unmodified potential leads to incorrect results with any reasonable binning choice we attempted.

B. MBAR/UWHAM estimates without binning

As discussed in Sec. II, binless free energy estimation methods make it unnecessary to bin the data in order to compute free energies. Binding energy samples b_{ji} , $i = 1, \dots, n_j$, from each simulation at $\lambda = \lambda_j$ are simply fed into Eq. (20), which is solved for the Z_{λ_j} 's by self-consistency¹¹ (referred to as the MBAR implementation) or by minimization of Eq. (42) (referred here as the UWHAM implementation); see below for details on the numerical implementation. The resulting binding free energy estimates for

TABLE III. Comparison of MBAR/UWHAM and WHAM computed binding free energies for the six complexes of FKBP with (“soft-core”) and without (“unmodified”) the soft-core binding energy function.

Ligand	Expt ^a	MBAR/UWHAM ^{a,c} (sub-sampled)				
		WHAM ^{a,b}	MBAR/UWHAM ^{a,b}		MBAR/UWHAM ^{a,c} (sub-sampled)	
		Soft-core	Unmodified	Soft-core	Unmodified	Soft-core
2	-7.80 ± 0.1	-2.46 ± 0.18	-2.21 ± 0.12	-2.56 ± 0.19	-2.22 ± 0.45	-2.10 ± 0.51
3	-8.40 ± 0.1	-3.90 ± 0.24	-3.86 ± 0.20	-4.01 ± 0.23	-4.63 ± 0.47	-3.85 ± 0.48
5	-9.50 ± 0.1	-3.85 ± 0.37	-4.13 ± 0.23	-3.98 ± 0.32	-4.03 ± 0.50	-4.30 ± 0.50
6	-10.80 ± 0.3	-3.74 ± 0.36	-3.74 ± 0.21	-3.86 ± 0.31	-3.77 ± 0.49	-3.79 ± 0.49
8	-10.90 ± 0.1	-4.45 ± 0.28	-5.42 ± 0.14	-4.59 ± 0.30	-5.81 ± 0.51	-3.61 ± 0.53
9	-11.10 ± 0.2	-6.19 ± 0.35	-6.03 ± 0.19	-6.31 ± 0.32	-6.20 ± 0.55	-6.34 ± 0.55

^aIn kcal/mol.^bUsing all of the data, statistical errors computed by block-bootstrapping.^cUsing 1-in-50 sub-sampled data and statistical errors computed as described in Appendix B.

the six protein-ligand systems are given in Table III. Identical results are obtained with either the MBAR or UWHAM implementations. Also reported in Table III are the results obtained with WHAM on the data with the soft core potential.

We immediately notice that the MBAR/UWHAM results with the unmodified potential agree very closely with those using the soft core potential. The fact that we obtained consistent results from two independent sets of simulations, each providing very different binding energy datasets is a strong indication that both of these results reflect the actual binding free energies for these systems. This is a significant result because it shows that binless methods are capable of treating correctly the distribution at high binding energies even though this extends to extremely large values (10^9 kcal/mol) and it is extremely sparsely sampled. For example, for the complex with ligand 6 there is on average only one observation every 10 000 kcal/mol in the range between 10^6 and 10^7 kcal/mol, a regime in which, clearly, binning is not a feasible option. As discussed above, reliable WHAM results could be obtained only for the soft core data because of challenges with binning the unmodified binding energy data. The agreement between MBAR/UWHAM soft core and unmodified potential results and with the WHAM soft core results confirms the ability of the binless inference methods to handle non soft-core data reliably.

The MBAR results obtained are based on the same equation (Eq. (8)) derived here.¹¹ The only difference is the computational procedure to solve it. UWHAM uses a minimization procedure with the criterion function (Eq. (42)), whereas MBAR employs a self-consistent procedure optionally supplemented by Newton-Raphson iterations. Here we have used the simple self-consistent solution starting with the default initial guess $Z_{\lambda_k} = 1$ (the same initial guess used for UWHAM). In our experience, UWHAM has provided a converged solution in significantly less computational time than MBAR (seconds vs. minutes typically), a feature that has been particularly helpful in block-bootstrapping uncertainty calculations involving 100 independent free energy evaluations per ligand. MBAR and UWHAM yielded virtually identical results thereby validating numerically the new minimization procedure presented here.

The last two columns in Table III are the results based on subsampled data,³⁵ including the point estimates and ana-

lytical errors computed as described in Appendix B. For each system, a subsample of size 20 has been selected, with 1 in every 50 time points, from the original sample of size 1000. The point estimates are reasonably close to those based on the original samples. The analytical errors, assuming uncorrelated data, are approximately 0.50 kcal/mol for all the systems, with or without the soft-core potential. Adjusting for sample sizes, the errors based on uncorrelated data of size 1000 would be about $0.50/\sqrt{50} \approx 0.07$ kcal/mol. Comparison of such adjusted errors with the block-bootstrap errors for the original data then indicates statistical inefficiency³⁶ caused by correlations. For example, for ligand 2, the factor of statistical inefficiency due to correlated data is $(0.12/0.07)^2 = 2.9$ for the unmodified potential and $(0.19/0.07)^2 = 7.4$ for the soft-core potential. It is interesting to note that the statistical inefficiencies with the soft-core potential are consistently larger than those with for the unmodified potential, implying smaller correlations and faster convergence of binding free energies with the latter.

1. MBAR/UWHAM probability densities

The BEDAM binding free energy theory highlights the fundamental importance of probability densities $p_\lambda(b)$ of the binding free energy as a function of the progress parameter λ . For example we have shown²⁹ that the binding free energy (Eq. (39)) can be written as

$$\Delta G_b = -kT \log \int p_0(b) e^{-\beta b} db, \quad (45)$$

where $p_0(b)$ is the probability density of the binding energies at $\lambda = 0$, that is in absence of ligand-receptor interactions. The probability density of binding energies $p_1(b)$ of the ligand-receptor coupled state is also of special interest. The mean of $p_1(b)$ is the average binding energy $\langle b \rangle_1$ which measures the driving force toward binding provided by favorable ligand-receptor interactions. The difference between the binding free energy and the average binding energy is the binding reorganization free energy that measures energetic strain and entropic factors which oppose binding. In addition to thermodynamic decompositions of this kind, $p_1(b)$ also leads to conformational decompositions of the binding free energy. $p_1(b)$ can be interpreted as the contribution to the binding affinity of

conformations with binding energy b and, consequently, distinct macrostates of the complex contribute to binding affinity proportionally to the integrated intensity of the corresponding components of $p_1(b)$.^{18,29}

It is straightforward to estimate these probability densities by binning and WHAM using Eqs. (4) and (7), which for the present application can be condensed as²

$$p_\lambda(b)\Delta b = \frac{1}{Z_\lambda} \frac{\sum_r N_r(b)}{\sum_r n_r Z_{\lambda_r}^{-1} e^{\beta(\lambda-\lambda_r)b}}, \quad (46)$$

where $p_\lambda(b)$ is the estimate of the probability density in correspondence with a bin centered at b , with bin width Δb , and $N_r(b)$ is the number of observations in that bin from the simulation at $\lambda = \lambda_r$. As presented above (see Eqs. (22) and (23)) the procedure to obtain probability densities and their moments (such as expectation values) is somewhat different when using binless methods. First each binding energy observation b_{ji} is assigned a λ -dependent statistical weight given in this case by

$$F_\lambda(b_{ji}) = \frac{1}{Z_\lambda \sum_r n_r Z_{\lambda_r}^{-1} e^{\beta(\lambda-\lambda_r)b_{ji}}}. \quad (47)$$

The sum of statistical weights over the samples is automatically unitary. Expectation values are computed as weighted averages using the weights in Eq. (47). For example the average binding energy at λ is

$$\langle b \rangle_\lambda = \sum_{ji} b_{ji} F_\lambda(b_{ji}). \quad (48)$$

Averages of other properties can be obtained similarly by replacing b_{ji} in Eq. (48) with any property of the sampled conformation ji . As discussed in Sec. II, this expression can also be used to estimate probability densities, such as the binding energy densities $p_\lambda(b)$. These can be approximated by the relationship

$$p_\lambda(b_k)\Delta b_k \simeq \langle \delta_{b_k}(b) \rangle_\lambda, \quad (49)$$

where $\delta_{b_k}(b)$ is a function defined as 1 if the argument falls within the bin centered at binding energy b_k with width Δb_k and zero otherwise. Then the average in Eq. (49) is computed using the equivalent of Eq. (48)

$$p_\lambda(b_k)\Delta b_k \simeq \sum_{ji} \delta_{b_k}(b_{ji}) F_\lambda(b_{ji}). \quad (50)$$

So the UWHAM calculation of $p_\lambda(b)$ basically consists of binning samples based on their binding energies and then creating a histogram in which the height of each bin is the sum of the weights $F_\lambda(b_{ji})$ of the observations collected in that bin.

Figures 1 and 2 illustrate the $p_1(b)$ and $p_0(b)$ probability densities obtained by UWHAM and Eq. (50) for the complex with ligand 6.³⁰ These are compared with the corresponding probability density estimates from the histograms of the data collected only at $\lambda = 1$ and $\lambda = 0$, respectively. There is good agreement between the two estimates in the region of binding energies well sampled at the respective λ values, further validating the UWHAM results. The tails of the probability densities are estimated much more accurately by UWHAM

than by the direct histograms because these are rarely sampled by the simulations conducted only at a specific λ . The UWHAM probability densities are instead estimated from data obtained from simulations at multiple λ values between 0 and 1 which explore a much wider range of binding energies. Obtaining accurate tails of probability densities is very important in a variety of applications such as for example when employing Eq. (45) to estimate the binding free energy from $p_0(b)$ (see Fig. 2). Due to the exponential term in the integrand of Eq. (45), the $p_0(b)$ density in the range $-40 < b < -10$ kcal/mol dominates the estimate of the binding free energy and the data collected at $\lambda = 0$ constitute a very poor estimate of $p_0(b)$ in this region of binding energies (although this is difficult to see in Fig. 2 because of log-log representation).

C. Simulation setup and numerical analysis

BEDAM calculations²⁹ were performed for six complexes of FKBP with ligands 2, 3, 5, 6, 8, and 9 from Ref. 37, from a ligand set which was the subject of previous binding free energy calculations.^{30,38,39} Complexes were prepared as described³⁰ based on the crystal structures of ligands 8 and 9 (PDB ID's 1FKG and 1FKH, respectively). Two BEDAM calculations were conducted for each complex, both employing Eq. (44) to represent the protein-ligand interaction potential, one with $b_{\max} = 10^9$ kcal/mol (referred to as the unmodified potential) and the other with $b_{\max} = 10^3$ kcal/mol (referred to as the soft-core potential). Soft-core calculations employed 15 BEDAM replicas at $\lambda = 0, 10^{-3}, 2 \times 10^{-3}, 4 \times 10^{-3}, 6 \times 10^{-3}, 8 \times 10^{-3}, 10^{-2}, 2 \times 10^{-2}, 6 \times 10^{-2}, 0.1, 0.25, 0.5, 0.75, 0.9$, and 1. Calculations with the unmodified potential employed 18 replicas at $\lambda = 0, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.15, 0.25, 0.35, 0.5, 0.75, 0.9$, and 1. Hamiltonian replica exchange simulations were conducted for 2 ns per replica (396 ns total simulation time). Binding energies were recorded at 1 ps intervals during the second half of the simulations, yielding 1000 observations per replica.

WHAM analysis has been performed employing Eqs. (43), (45), and (46) as described² on the collected binding energy data, b_{ji} , using a binning grid starting at -40 kcal/mol (the lowest recorded binding energy value) to a set maximum (see Tables I and II) for the unmodified binding energy function or the to maximum allowed value ($b_{\max} = 10^3$ kcal/mol) for the soft core data. Grid spacing was set to 0.3 kcal/mol in the $(-40, -10)$ binding energy range, increasing exponentially starting from this value at a rate adjusted to reach the variable maximum set value with the given number of bins.

UWHAM analysis was conducted on the same binding energy data to obtain the logarithm of the partition functions $\log Z_{\lambda_k}$ (relative to $\log Z_0$, which is set to 0) by minimization of the function $\kappa(\log z_1, \dots, \log z_m)$ in Eq. (42) with respect to $\log z_k$ setting the free energy of the unbound state to zero ($\log z_1 = \log Z_0 = 0$). For the minimization, we used the trusted region algorithm¹⁵ as implemented in the R statistical package "trust".⁴⁰ A similar procedure has been recently proposed in the context of WHAM.⁶ The code for the UWHAM R module we employed and a set of use examples in R are

available by the authors upon request. MBAR calculations were performed using the code kindly provided by Chodera and Shirts.¹¹

Statistical uncertainties were computed using the block bootstrapping method.¹⁷ Similarly to the traditional block averaging approach,⁴¹ the method consists of dividing the sampled data in N_b time-contiguous blocks ($N_b = 20$ in this work). However, in this context blocks span all of the replicas; that is each block contains the data generated from all of the replicas in the same time window. Each block is then assigned an integer identifier and the list of block identifiers plays the same role of the data samples in the standard bootstrap method. Namely, a new block identifier list of length N_b is created by sampling with repetition from the original list and a corresponding new binding energy dataset is generated by collating the data contained in the blocks of the new list. This is repeated a number of times (100 times in this case) and the statistical uncertainty of the binding free energy is estimated from the standard deviation of the free energy values from each bootstrap sample. The advantage of this bootstrap technique is that it accounts for time correlation of samples originating from each replica as well as cross-correlations between replicas due to λ exchanges.

IV. CONCLUSION

We demonstrate the statistical validity and usefulness of interpreting MBAR as a binless formulation of WHAM. Like WHAM, the binless formulation can be used not only to estimate free energies and equilibrium expectations, but also to estimate equilibrium distributions. This development allows practitioners to easily build on their current applications of WHAM, but without discretizing observations into bins, which may sometimes incur substantial biases. This is illustrated for alchemical absolute binding free energy calculations using the BEDAM technique. While UWHAM and MBAR¹¹ binless implementations yield equivalent results for either the unmodified and soft-core potentials, binning of the unmodified data leads to substantial biases which vary depending on the level of discretization. These results indicate that binless multi-state inference approaches are potentially a straightforward alternative to soft-core potentials for binding free energy alchemical calculations.

ACKNOWLEDGMENTS

This work has been supported in part by research grants from the National Institute of Health (Grant No. GM30580) and the National Science Foundation (CDI type II Grant No. 1125332 and DMS-0749718). The calculations reported in this work have been performed at the BioMaPS High Performance Computing Center at Rutgers University funded in part by the NIH shared instrumentation (Grant Nos. 1 S10 RR022375 and 1 S10 RR027444), and on the Lonestar4 cluster at the Texas Advanced Computing Center under TeraGrid/XSEDE National Science Foundation allocation (Grant No. MCB100145). The authors are grateful to John Chodera and Michael Shirts for providing valuable suggestions and guidance.

APPENDIX A: COMPUTING POINT ESTIMATORS

To compute $\log(\hat{Z}_{\theta_2}/\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m}/\hat{Z}_{\theta_1})$, we minimize $\kappa(0, \zeta_2, \dots, \zeta_m)$ using the trust region algorithm implemented by the R package `trust`.⁴⁰ This algorithm is globally convergent at the second order (Sec. 4.2 of Ref. 15). Below we provide formulas for evaluating κ and its gradient and Hessian, which are required by the trust region algorithm.

Arrange the pooled data into a column vector $(u_1, \dots, u_n)^T$. Let Π_s be the $m \times m$ diagonal matrix with the (j, j) th element n_j/n , and Q_s and W_s be the $n \times m$ matrices, respectively, with (i, j) th element

$$Q_{ij} = e^{-\zeta_j} e^{-\theta_j^T u_i}, \quad W_{ij} = \frac{e^{-\zeta_j} e^{-\theta_j^T u_i}}{\sum_{r=1}^m \frac{n_r}{n} e^{-\zeta_r} e^{-\theta_r^T u_i}}.$$

Write $\zeta = (\zeta_1, \dots, \zeta_m)^T$ and 1_m (or 1_n) as the column vector of m (or n) ones. Then

$$\kappa(\zeta) = \frac{1}{n} \log(Q_s \Pi_s 1_m) + \sum_{r=1}^m \frac{n_r}{n} \zeta_r,$$

$$\frac{\partial \kappa}{\partial \zeta}(\zeta) = -\Pi_s W_s^T \frac{1_n}{n} + \Pi_s 1_m,$$

$$\frac{\partial^2 \kappa}{\partial \zeta \partial \zeta^T}(\zeta) = -\frac{1}{n} \Pi_s W_s^T W_s \Pi_s + \text{diag} \left(\Pi_s W_s^T \frac{1_n}{n} \right),$$

where $\text{diag}(c)$ is the diagonal matrix with (j, j) element c_j for a vector $c = (c_1, \dots, c_m)$. The gradient (or Hessian) of $\kappa(0, \zeta_2, \dots, \zeta_m)$ is formed by deleting the first element (or the first row and column) from that of $\kappa(\zeta)$.

APPENDIX B: COMPUTING VARIANCE MATRICES

Suppose that $\hat{Z}_{\theta}/\hat{Z}_{\theta_1}$ is computed for k values, $\theta_{m+1}, \dots, \theta_{m+k}$, of θ . Let R be the $n \times m$ matrix with (i, j) th element $(n_j/n)^{-1}$ if u_i is sampled from F_{θ_j} and 0 otherwise, and W be the $n \times (m+k)$ matrix with (i, j) th element

$$W_{ij} = \frac{(\hat{Z}_{\theta_j}/\hat{Z}_{\theta_1})^{-1} e^{-\theta_j^T u_i}}{\sum_{r=1}^m \frac{n_r}{n} (\hat{Z}_{\theta_r}/\hat{Z}_{\theta_1})^{-1} e^{-\theta_r^T u_i}}.$$

Let I_{m+k} be the identity matrix of size $m+k$, $0_{(m+k) \times k}$ be the $(m+k) \times k$ matrix of zeros, and

$$O = \frac{1}{n} W^T W, \quad B = \{O_s \Pi_s, 0_{(m+k) \times k}\} - I_{m+k},$$

$$D = \frac{1}{n} W^T R, \quad A = O - D \Pi_s D^T,$$

where O_s is the $(m+k) \times m$ matrix consisting of the first m columns in O . The (j, r) th element of D is the sample average of (i, r) th elements of W for $i = 1, \dots, n$ such that u_i is sampled from F_{θ_j} . The asymptotic variance matrix of $\log(\hat{Z}_{\theta_2}/\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_m}/\hat{Z}_{\theta_1}, \hat{Z}_{\theta_{m+1}}/\hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_{m+k}}/\hat{Z}_{\theta_1})$ can be consistently estimated by

$$\frac{1}{n} B_{(1)}^{-1} A_{(1)} B_{(1)}^{-1}, \quad (\text{B1})$$

where $A_{(1)}$ and $B_{(1)}$ are formed by deleting the first row and column from A and B . Alternatively, formula (B1) can be used with A replaced by $O - O_s \Pi_s O_s^T$. The resulting formula does

not require the use of the information about which observation is sampled from which distribution.

Suppose that $\langle h \rangle_\theta$ is estimated for $\theta = \theta_1, \dots, \theta_m, \theta_{m+1}, \dots, \theta_{m+k}$. Write formula (22) as the ratio $\hat{Z}_\theta^h / \hat{Z}_\theta$, where

$$\hat{Z}_\theta^h = \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{h(u_{ji})}{\sum_{r=1}^m n_r \hat{Z}_{\theta_r}^{-1} e^{(\theta - \theta_r)^T u_{ji}}}.$$

Redefine W as the $n \times (m+k)$ matrix with (i, j) th element

$$W_{ij} = \frac{e^{-\theta_j^T u_i}}{\sum_{r=1}^m \frac{n_r}{n} (\hat{Z}_{\theta_r} / \hat{Z}_{\theta_1})^{-1} e^{-\theta_r^T u_i}}.$$

Let W^h be the $n \times (m+k)$ matrix with (i, j) th element

$$W_{ij}^h = \frac{h(u_i) e^{-\theta_j^T u_i}}{\sum_{r=1}^m \frac{n_r}{n} (\hat{Z}_{\theta_r} / \hat{Z}_{\theta_1})^{-1} e^{-\theta_r^T u_i}}.$$

Now replace W by (W, W^h) throughout and redefine

$$B = \{O_s C^{-1} \Pi_s C^{-1}, 0_{(2m+2k) \times (m+2k)}\} - I_{2m+2k},$$

where O_s is the $(2m+2k) \times m$ matrix consisting of the first m columns in O , and C is the $(m+k) \times (m+k)$ diagonal matrix with (j, j) th element $\hat{Z}_{\theta_j} / \hat{Z}_{\theta_1}$. The asymptotic variance matrix of $(\hat{Z}_{\theta_2} / \hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_{m+k}} / \hat{Z}_{\theta_1}, \hat{Z}_{\theta_1}^h / \hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_{m+k}}^h / \hat{Z}_{\theta_1})$ can be consistently estimated by formula (B1), which is denoted by V . The asymptotic variance matrix of $(\hat{Z}_{\theta_1}^h / \hat{Z}_{\theta_1}, \dots, \hat{Z}_{\theta_{m+k}}^h / \hat{Z}_{\theta_{m+k}})$ can be consistently estimated by

$$C^{-1} (-C^h, I_{m+k}) \begin{pmatrix} 0 & 0_{2m+2k-1}^T \\ 0_{2m+2k-1} & V \end{pmatrix} \begin{pmatrix} -C^h \\ I_{m+k} \end{pmatrix} C^{-1},$$

where $0_{2m+2k-1}$ is the column vector of $2m+2k-1$ zeros and C^h is the $(m+k) \times (m+k)$ diagonal matrix with (j, j) th element $\hat{Z}_{\theta_j}^h / \hat{Z}_{\theta_j}$.

- ¹A. M. Ferrenberg and R. H. Swendsen, "Optimized Monte Carlo data analysis," *Phys. Rev. Lett.* **63**, 1195–1198 (1989).
- ²E. Gallicchio, M. Andreac, A. K. Felts, and R. M. Levy, "Temperature weighted histogram analysis method, replica exchange, and transition paths," *J. Phys. Chem. B* **109**, 6722–6731 (2005).
- ³*Free Energy Calculations. Theory and Applications in Chemistry and Biology*, Springer Series in Chemical Physics, edited by C. Chipot and A. Pohorille (Springer, Berlin/Heidelberg, 2007).
- ⁴M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University Press, New York, 1999).
- ⁵C. Bartels and M. Karplus, "Multidimensional adaptive umbrella sampling: Application to main chain and side chain peptide conformations," *J. Comput. Chem.* **18**, 1450–1462 (1997).
- ⁶F. Zhu and G. Hummer, "Convergence and error estimation in free energy calculations using the weighted histogram analysis method," *J. Comput. Chem.* (in press).
- ⁷C. J. Geyer, "Estimating normalizing constants and reweighting mixtures in Markov chain Monte Carlo," Technical report, University of Minnesota, School of Statistics, 1994.
- ⁸X.-L. Meng and W. H. Wong, "Simulating ratios of normalizing constants via a simple identity: A theoretical explanation," *Stat. Sin.* **6**, 831–860 (1996).
- ⁹A. Kong, P. McCullagh, X.-L. Meng, D. Nicolae, and Z. Tan, "A theory of statistical models for Monte Carlo integration," *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **65**, 585–618 (2003).
- ¹⁰Z. Tan, "On a likelihood approach for Monte Carlo integration," *J. Am. Stat. Assoc.* **99**(468), 1027–1036 (2004).
- ¹¹M. R. Shirts and J. D. Chodera, "Statistically optimal analysis of samples from multiple equilibrium states," *J. Chem. Phys.* **129**(12), 124105 (2008).

- ¹²C. H. Bennett, "Efficient estimation of free energy differences from Monte Carlo data," *J. Comput. Phys.* **22**(2), 245–268 (1976).
- ¹³N. Lu, J. K. Singh, and D. A. Kofke, "Appropriate methods to combine forward and reverse free-energy perturbation averages," *J. Chem. Phys.* **118**(7), 2977–2984 (2003).
- ¹⁴P. Billingsley, *Probability and Measure* (Wiley, New York, 1995).
- ¹⁵J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer-Verlag, New York, 1999).
- ¹⁶R. Gill, Y. Vardi, and J. Wellner, "Large sample theory of empirical distributions in biased sampling models," *Ann. Stat.* **16**, 1069–1112 (1988).
- ¹⁷M. R. Chernick, *Bootstrap Methods: A Guide for Practitioners and Researchers*, 2nd ed. (Wiley, Hoboken, NJ, 2008).
- ¹⁸E. Gallicchio and R. M. Levy, "Recent theoretical and computational advances for modeling protein-ligand binding affinities," in *Advances in Protein Chemistry and Structural Biology* (Academic, 2011), Vol. 85, pp. 27–80.
- ¹⁹M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, "The statistical-thermodynamic basis for computation of binding affinities: A critical review," *Biophys. J.* **72**, 1047–1069 (1997).
- ²⁰D. L. Mobley and K. A. Dill, "Binding of small-molecule ligands to proteins: 'what you see' is not always 'what you get'," *Structure (London)* **17**(4), 489–498 (2009).
- ²¹Y. Deng and B. Roux, "Computations of standard binding free energies with molecular dynamics simulations," *J. Phys. Chem. B* **113**(8), 2234–2246 (2009).
- ²²J. Michel and J. W. Essex, "Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls, and expectations," *J. Comput.-Aided Mol. Des.* **24**(8), 639–658 (2010).
- ²³J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, and V. S. Pande, "Alchemical free energy methods for drug discovery: Progress and challenges," *Curr. Opin. Struct. Biol.* **21**, 150–160 (2011).
- ²⁴M. R. Shirts and V. S. Pande, "Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration," *J. Chem. Phys.* **122**(14), 144107 (2005).
- ²⁵A. Pohorille, C. Jarzynski, and C. Chipot, "Good practices in free-energy calculations," *J. Phys. Chem. B* **114**(32), 10235–10253 (2010).
- ²⁶T. Steinbrecher, D. L. Mobley, and D. A. Case, "Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations," *J. Chem. Phys.* **127**(21), 214108 (2007).
- ²⁷T. Steinbrecher, I. Joung, and D. A. Case, "Soft-core potentials in thermodynamic integration: Comparing one- and two-step transformations," *J. Comput. Chem.* **32**(15), 3253–3263 (2011).
- ²⁸F. P. Buelens and H. Grubmüller, "Linear-scaling soft-core scheme for alchemical free energy calculations," *J. Comput. Chem.* **33**(1), 25–33 (2012).
- ²⁹E. Gallicchio, M. Lapelosa, and R. M. Levy, "Binding energy distribution analysis method (BEDAM) for estimation of protein-ligand binding affinities," *J. Chem. Theory Comput.* **6**(9), 2961–2977 (2010).
- ³⁰M. Lapelosa, E. Gallicchio, and R. M. Levy, "Conformational transitions and convergence of absolute binding free energy calculations," *J. Chem. Theory Comput.* **8**, 47–60 (2012).
- ³¹S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," *J. Comput. Chem.* **13**, 1011–1021 (1992).
- ³²E. Gallicchio and R. M. Levy, "Advances in all atom sampling methods for modeling protein-ligand binding affinities," *Curr. Opin. Struct. Biol.* **21**(2), 161–166 (2011).
- ³³C. D. Christ, A. E. Mark, and W. F. van Gunsteren, "Basic ingredients of free energy calculations: a review," *J. Comput. Chem.* **31**(8), 1569–1582 (2010).
- ³⁴E. Gallicchio and R. M. Levy, "Prediction of sampl3 host-guest affinities with the binding energy distribution analysis method (BEDAM)," *J. Comp.-Aided Mol. Des.* (in press).
- ³⁵H. Paliwal and M. R. Shirts, "A benchmark test set for alchemical free energy transformations and its use to quantify error in common free energy methods," *J. Chem. Theory Comput.* **7**(12), 4115–4134 (2011).
- ³⁶W. Janke, "Statistical analysis of simulations: Data correlations and error estimation," in *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms* (John von Neumann Institute for Computing, Jülich, Germany, 2002), pp. 423–445.

- ³⁷D. A. Holt, J. I. Luengo, D. S. Yamashita, H. J. Oh, A. L. Konialian, H. K. Yen, L. W. Rozamus, M. Brandt, and M. J. Bossard, "Design, synthesis, and kinetic evaluation of high-affinity fkbp ligands and the x-ray crystal structures of their complexes with fkbp12," *J. Am. Chem. Soc.* **115**(22), 9925–9938 (1993).
- ³⁸J. Wang, Y. Deng, and B. Roux, "Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials," *Biophys. J.* **91**(8), 2798–2814 (2006).
- ³⁹H. Fujitani, Y. Tanida, M. Ito, J. Guha, C. D. Snow, M. R. Shirts, E. J. Sorin, and V. S. Pande, "Direct calculation of the binding free energies of fkbp ligands," *J. Chem. Phys.* **123**(8), 084108 (2005).
- ⁴⁰C. J. Geyer, Trust region optimization, R package 0.1-2., 2009, see <http://www.stat.umn.edu/geyer/trust/>.
- ⁴¹M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1993).