THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

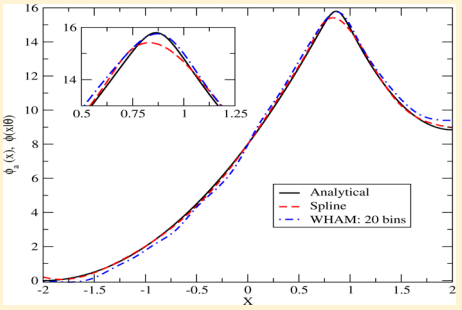
pubs.acs.org/JPCB

Optimization and Automation of the Construction of Smooth Free Energy Profiles

Jeremy Schofield*

Chemical Physics Theory Group, Department of Chemistry, University of Toronto, 80 Saint George Street, Toronto, Ontario M5S 3H6, Canada

ABSTRACT: An adaptive procedure is introduced to construct smooth analytical profiles of the free energy along a reaction coordinate using sampled data from multiple biased simulations. The procedure is based upon identifying problematic regions encountered in maximum likelihood estimators of the profile where there are statistically relevant discrepancies between the empirical and parametrized cumulative distribution functions and preferentially improving the construction of the parametric profile in these regions. The method is designed to produce continuous and smooth analytical fits that satisfy statistical goodness-of-fit tests with a minimum number of parameters. The accuracy of the profile obtained from the adaptive construction is compared by numerical computation to that of smooth interpolations based on an optimally chosen weighted histogram method for a solvated ion pair system and for an activated process for which the analytical form of the potential of mean force is available. In the model where the exact profile is known, the adaptive procedure is shown to reduce the integrated error relative to the optimal histogram construction by a factor of 3 or more in the typical case where the sampling is not extensive. It is demonstrated that the adaptive procedure can be used to produce statistically accurate smooth analytical representations of the free energy profile that can be evaluated with little computational effort and require little user input.



INTRODUCTION

The calculation of the free energy profile or the potential of mean force along a reaction coordinate from sampled data is a crucial task in the estimation of chemical rate constants using transition state theory[1,2] and figures prominently in the general theory of rate processes in the computation of the mean first passage time out of a metastable well.[3,4] The potential of mean force is also of importance in the context of the Fokker−Planck equation[5−7] and in the construction of coarse grain models where the potential describes the effective interactions between components of the configurational space of the coarse grain model. In this context, the potential of mean force is equal to an equilibrium average over a "bath" of presumably rapid motions of the full system. In the coarse grain models, the potential of mean force is frequently constructed from sampled data using the reverse Monte Carlo method,[8,9] iterative Boltzmann inversion,[10,11] or the force matching method.[12,13] These techniques rely on parametric fits of the potential of mean force (or the force itself) to construct smooth potentials of mean force to use as force fields suitable for mesoscale simulations.

Recently, much attention has been devoted to adaptive importance sampling procedures,[14−22] metadynamics,[23−25] and its extensions[26] in which estimates of the potential of mean force or the probability density are iteratively improved upon on the basis of prior estimates constructed from data drawn from a biased sampling procedure. A critical step in the adaptive procedure is the accurate construction of the estimates of the potential of mean force based on all available data sampled from biased densities.

Most approaches of constructing the free energy profile rely on defining a variable set of parameters either in terms of a set of k explicit parameters $\boldsymbol{\theta}_k$ or implicitly through choices of bin size and/or number. The choice of the optimal number of parameters used to represent the profile is frequently unclear and strongly depends on the choice of representation. The quality of the fit of the global free energy profile to the data typically improves with an increase in the number of parameters at the expense of the smoothness of the constructed profile. In essence, for small sample sizes of noisy data, the use of a large number of parameters overfits the data set and overemphasizes the statistical importance of the noise. Another important consideration is the computational demand of calculating values of the potential energy and its gradient if the potential of mean force is to be used as a force field in coarse grain simulations. For such purposes, nonparametric kernel density[27,28] and Gaussian process regression methods[29] that place a Gaussian basis set function at every sampling (or training) point may not be practical.

Within the context of a parametric fit of the potential of mean force, it is therefore desirable to devise a method of deciding when increasing the size of the parameter set results in little improvement in statistical accuracy. The purpose of this

Received: May 19, 2017
Revised: June 22, 2017
Published: June 22, 2017

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

communication is to present a method that adaptively optimizes the parametrization of the free energy profile and determines when a given number of parameters k represents the profile with statistical accuracy, results in a smooth curve that can be evaluated with little computational effort, and requires little user input.

The paper is structured as follows: In the first section on methodology, a brief summary of how various methods of forming smooth fits of the potential of mean force from data from biased simulations depend on parametrization is given. This is followed by a review of the asymptotic properties of maximum likelihood estimators of a parameter set, with particular attention devoted to the convergence of empirical distribution functions to the exact distribution function. Next, methods of assessing the quality of the construction of the free energy profile from the data available based on the maximum deviation observed between suitably defined empirical and parametric cumulative distribution functions are presented. In the Results section, a model of an activated process for which the analytical form of the free energy profile is available is analyzed, and the performance of an adaptive algorithm to determine parameters constructing the profile is compared to those obtained from weighted histogram methods. This is followed by a discussion of the free energy profile of a solvated ion-pair system in which solvation shells lead to an oscillatory potential of mean force. Finally, a brief discussion of the generalization of the adaptive approach to higher dimension and general conclusions are presented in the Discussion and Conclusions section.

METHODOLOGY

The Potential of Mean Force or Free Energy Profile. The central task is to use data generated from biased sampling procedures to construct a smooth and statistically optimal estimate of the dimensionless potential of mean force defined by $\phi(x) = -\ln p(x)$, where $p(x)$ is the probability density of value x of a specified general "reaction" coordinate ξ that depends on the configuration of the system. In this work, we assume that the density is $p(x) > 0$ so that $\phi(x)$ is finite at all points in the domain of interest. Physically, the potential of mean force is the effective potential giving rise to the average force at a specified value of the reaction coordinate, and therefore, its differences can also be interpreted as the reversible work, or change in free energy, required to move the system along the path defined by the reaction coordinate. For this reason, the potential of mean force is also known as the free energy profile, particularly in the context of chemical reaction dynamics. Subsequently, we will use the two terms interchangeably.

In typical applications, $p(x)$ is the reduced density obtained from a multivariate density $P(y_1, ..., y_N)$ of N degrees of freedom

$$p(x) = \int \mathrm{d}\mathbf{y}\, \delta(\xi(\mathbf{y}) - x) P(\mathbf{y})$$

where $\mathbf{y}$ denotes the vector $(y_1, ..., y_N)$ and $\xi(\mathbf{y})$ is the reaction coordinate. In the most common setting in statistical physics, the density $P(\mathbf{y})$ is specified by a choice of an ensemble such as the canonical ensemble in which the density is determined by a microscopic interaction potential $U(\mathbf{y})$ that governs how different degrees of freedom are correlated with one another. In the canonical ensemble, spatial positions are distributed according to the density $P(\mathbf{y}) = \mathrm{e}^{-\beta U(\mathbf{y})}/Z(\beta)$, where $Z(\beta) = \int \mathrm{d}\mathbf{y}\ \mathrm{e}^{-\beta U}$ is a normalization constant and β is a parameter depending on the physical temperature. In the following, we consider an arbitrary choice of reaction coordinate $\xi(\mathbf{y})$ that is determined by the physical problem of interest.

Computing the Free Energy Profile. Given a set $\{\mathbf{y}_1, ..., \mathbf{y}_n\}$ of n independent and identically distributed data drawn from a probability density $P(\mathbf{y})$ from independent Monte Carlo or molecular dynamics simulations, a common strategy used to construct the potential of mean force $\phi(x)$ over a finite range $[x_-, x_+]$ from the data is to model $\phi(x) = \phi(x|\boldsymbol{\theta}_{k+1})$ in terms of a set of parameters $\boldsymbol{\theta}_{k+1} = \{\theta_0, ..., \theta_k\}$ that are optimized in some fashion using the data set $\mathbf{x} = \{x_1, ..., x_n\}$ of values of the reaction coordinate $x_i = \xi(\mathbf{y}_i)$. In situations where the density of the reaction coordinate $p(x)$ is never small and $\phi(x)$ is never large, many methods of constructing smooth densities[30] from sampled data will work well. Approaches such as kernel density estimation[27,28] where the bandwidth is a parameter or fits of the empirical cumulative distribution[31,32] are well-suited for such cases.

However, in most situations of interest in chemical physics, one would like to compute the profile of $\phi(x)$ over a domain $[x_-, x_+]$ in which $\phi(x)$ takes on large values (such as barrier regions in chemical reactions). In regions where the density $p(x)$ is extremely small, estimates of the free energy profile based on direct samples from the multivariate density P converge poorly. To improve the convergence properties of the estimated free energy, biased sampling methods are helpful. One frequently utilized method to enhance the sampling over the domain $[x_-, x_+]$ is umbrella sampling[33] in which a series of continuous biasing weights $w_\alpha(x)$ with $\alpha = 1, ..., S$ are utilized in independent simulations to force sampling of improbable regions where the potential of mean force is large. In each of the umbrella simulations, a set $\mathbf{y}^\alpha = \{\mathbf{y}_1^\alpha, ..., \mathbf{y}_{n_\alpha}^\alpha\}$ of n_α points is independently drawn from the biased density $w_\alpha(\xi(\mathbf{y}))P(\mathbf{y})$. From the sampled points, the data set of reaction coordinates $\{\mathbf{x}^1, ..., \mathbf{x}^S\}$ is used to construct a global free energy profile over the entire domain $[x_-, x_+]$, where the points in the $\mathbf{x}^\alpha = \{\xi(\mathbf{y}_1^\alpha), ..., \xi(\mathbf{y}_{n_\alpha}^\alpha)\}$ data set are distributed according to density $q_\alpha(x) = z_\alpha^{-1} w_\alpha(x) p(x)$, where $z_\alpha = \int \mathrm{d}x\ w_\alpha(x) p(x)$.

There are a multitude of ways of computing the free energy profile with respect to a specified reaction coordinate. Many of the methods, such as the commonly used WHAM procedure,[34−36] evaluate either the potential of mean force or the average force[37] at a set of specified values or grid points of the reaction coordinate. For example, the method of steered molecular dynamics[38] can be considered to be in this category. Smooth profiles of the reaction coordinate may then be generated from the values ϕ_i of the potential of mean force at the grid points x_i using interpolation schemes. The grid points may be viewed as parameters upon which the smooth representation of the potential of mean force $\phi(x|\boldsymbol{\theta})$ is based.

In other methods, the parametrization of the free energy profile is of a different form. For example, in the umbrella integration method of Kästner and Thiel,[39−41] a linear combination of the mean force computed in each of the biased simulations is used to construct the global free energy profile. In each of the local computations of the mean force, the free energy profile is represented in terms of a low order expansion of the free energy around some point of the reaction coordinate, and the coefficients in this expansion are effectively parameters determined by averages obtained empirically from the data. Another approach taken recently[42,43] is to expand the potential of mean force in a finite basis set of Gaussian densities

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

and then determine the expansion coefficients either by minimizing the residual of a global least-squares estimator or by using Gaussian process regression.[29,44] Similarly, in the method of weighted residuals,[45] the potential of mean force is expanded in a set of basis functions $\phi(x|\mathbf{A}) = \sum_j A_j \phi_j(x)$, and the set of expansion coefficients A_j are parameters determined from a linear equation that follows from the fact that the residual for the differential equation for the mean force is orthogonal to the functional space formed by the basis set or other test functions. In fact, several other methods that have been introduced for evaluating the potential of mean force such as thermodynamic integration[46] can be derived from the weighted residual framework by making specific choices of test basis functions.[45]

Another simple approach similar in spirit to the method of umbrella integration is to express the estimated density $p(x|\boldsymbol{\theta})$ in terms of a weighted linear combinations of parametrized fits to the biased densities $q_\alpha(x|\boldsymbol{\theta})$, where the weights are determined by minimizing the variance in $p(x|\boldsymbol{\theta})$. This approach forms the basis of the multistate Bennet acceptance ratio method (MBAR).[47]

Maximum Likelihood Estimator of the Free Energy Profile. Many of the methods above can be derived by maximizing the likelihood that the data came from the parametrized density $p(x|\boldsymbol{\theta})$. The particular procedure used to select the set of parameters depends on how the density or equivalently the potential of mean force is constructed.

In this work, we introduce an adaptive procedure using statistical information to select and modify parametrizations of the potential of mean force using data from a set of biased simulations. To illustrate the method of truncating and adapting the parameter set, the parameter set $\boldsymbol{\theta}$ and hence the free energy profile will be determined using the principle of maximum likelihood applied to all available data from possibly multiple biased simulations. In this procedure, the average log-likelihood function of the composite data set of simulation points based on the model $\phi(x) = \phi(x|\boldsymbol{\theta}_{k+1})$ is defined as[48]

$$\hat{l}(\boldsymbol{\theta}_{k+1}) = \frac{1}{n} \log p(\mathbf{x}^1, ..., \mathbf{x}^S|\boldsymbol{\theta}_{k+1})$$

where $n = \sum_\alpha n_\alpha$ is the total number of data points available from the set of simulations. Noting that, for independent sampling points,

$$p(\mathbf{x}^1, ..., \mathbf{x}^S|\boldsymbol{\theta}_{k+1}) = \prod_{\alpha=1}^{S} \prod_{i=1}^{n_\alpha} \frac{w_\alpha(x_i^\alpha)}{z_\alpha(\boldsymbol{\theta}_{k+1})} e^{-\phi(x_i^\alpha|\boldsymbol{\theta}_{k+1})}$$

where $z_\alpha(\boldsymbol{\theta}_{k+1}) = \int dx\, w_\alpha(x) e^{-\phi(x|\boldsymbol{\theta}_{k+1})}$, the average log-likelihood function can be written as

$$\hat{l}(\boldsymbol{\theta}_{k+1}|\mathbf{x}^1, ..., \mathbf{x}^S) = -\sum_{\alpha=1}^{S} \lambda_\alpha \left(\ln z_\alpha(\boldsymbol{\theta}_{k+1}) + \frac{1}{n_\alpha} \sum_{i=1}^{n_\alpha} \phi(x_i^\alpha|\boldsymbol{\theta}_{k+1}) - \frac{1}{n_\alpha} \sum_{i=1}^{n_\alpha} \ln w_\alpha(x_i^\alpha) \right) \quad (1)$$

where x_i^α is a value of the reaction coordinate sampled in the α simulation and $\lambda_\alpha = n_\alpha/n$ is the fraction of points drawn from simulation α. The last term in eq 1 is independent of the parameters $\boldsymbol{\theta}_{k+1}$ and may be dropped from the likelihood function. Note that the average log-likelihood for the combined simulations can be viewed as a linear combination of the average log-likelihood functions for each of the S simulations weighted by the simulation number fractions λ_α.

The optimal parameter set $\boldsymbol{\theta}_{k+1}^*$ is determined by maximizing the average log-likelihood, or equivalently minimizing the potential of mean force $\phi(x|\boldsymbol{\theta})$, with respect to the set of $k+1$ parameters $\boldsymbol{\theta}_{k+1}$. Efficient methods for finding the minimum of convex functions, such as the Broyden−Fletcher−Goldfarb−Shanno algorithm,[49,50] make use of gradients to facilitate convergence. Such gradients can often be evaluated explicitly. For a fixed number $k+1$ of parameters, the free energy profile $\phi(x)$ and density of the reaction coordinate $p(x)$ may be estimated by $\phi(x|\boldsymbol{\theta}_{k+1}^*)$ and $p(x|\boldsymbol{\theta}_{k+1}^*)$, respectively.

In the limit of large numbers $n \to \infty$, the average log-likelihood function is locally quadratic in the parameter space in the vicinity of the maximum $\boldsymbol{\theta}_{k+1}^*$ in all directions except for that corresponding to a constant shift in the free energy. Such constant shift terms can be eliminated by choosing a reduced parametrization with k independent parameters $\boldsymbol{\theta}_k$ defined such that the free energy is fixed to a constant value via a constraint equation $\phi(x_c|\boldsymbol{\theta}_{k+1}) = c$ at some point x_c in the domain $[x_-, x_+]$ or by imposing some other constraint on the parameter space. Henceforth, we assume the free energy is parametrized such that its value is fixed at some point on the interval so that the parameter set that maximizes the average log-likelihood is uniquely determined.

In the limit of large sample sizes (i.e., large n), the set of parameters $\boldsymbol{\theta}_k^*$ are asymptotically normally distributed with mean $\boldsymbol{\theta}_k^0$ and (nonsingular) $k \times k$ covariance matrix $n^{-1}\mathbf{I}(\boldsymbol{\theta})_{lm}^{-1}$, where

$$\mathbf{I}(\boldsymbol{\theta}_k^*)_{lm} = -\int_{x_-}^{x_+} dx\, p(x|\boldsymbol{\theta}_k^*) \left[\frac{\partial^2 \hat{l}}{\partial\theta_l \partial\theta_m} \right]_{\boldsymbol{\theta}_k^*} = \int_{x_-}^{x_+} dx\, p(x|\boldsymbol{\theta}_k^*) \left[\frac{\partial \hat{l}}{\partial\theta_l} \frac{\partial \hat{l}}{\partial\theta_m} \right]_{\boldsymbol{\theta}_k^*}$$

is the positive-definite Fisher information matrix which can be used to form confidence intervals for $\boldsymbol{\theta}$-dependent observables. For example, the 95% confidence intervals for the parameter θ_l are estimated to be $\theta_l^* \pm 1.96\sqrt{(I^{-1})_{ll}/n}$. This procedure lies at the heart of kernel and mixture density estimation as well as many other methods.[27,28,51]

Parameterizations of the Free Energy Profile. A smooth free energy profile can be represented with either global or local parametrizations which differ qualitatively in the manner in which the profile is constructed. In a global parametrization, $\phi(x)$ is expanded in a (truncated) orthogonal series of polynomials, such as the Legendre polynomials $P_n(x)$,

$$\phi(x|\boldsymbol{\theta}) = \sum_{n=1}^{k} \theta_n P_n(x_s(x)) + \theta_0(\boldsymbol{\theta})$$
$$x_s(x) = \frac{2x}{x_+ - x_-} - \frac{x_- + x_+}{x_+ - x_-} \quad (2)$$

where the θ_0 is defined such that $\phi(x_c|\boldsymbol{\theta}) = c$. For example, one can choose $x_c = x_+$ where $P_n(x_s(x_+)) = 1$ so that $\theta_0 = c - \sum_{i=1}^{k} \theta_i$. The expansion in eq 2 is global in the sense that increasing the maximum number of terms k in the expansion affects the whole free energy profile, since the Legendre polynomials are not localized to specific regions of the interval $[x_-, x_+]$. Nonetheless, the representation of the free energy

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

profile is guaranteed to converge to the exact function as the number of terms in the complete basis set goes to infinity.

In contrast to global parametrizations of the potential of mean force, an alternative linear expansion could use functions such as Gaussian component densities (kernels) that are centered around specific points in the domain.[20,21] The advantage of this type of parametrization is in the local control that results from the ability to modify the interpolation in a limited part of the domain. Another parametrization scheme that allows local modifications of the potential of mean force is to represent the profile in terms of natural cubic splines based on a set of knot positions $\{(x_i, \theta_i)\}$, where θ_i are optimized for the data set at a choice of uniformly distributed[48] or adaptively selected reaction coordinate points x_i. In the cubic spline expansion of the free energy, $\phi(x)$ is written for $x_i \le x < x_{i+1}$ as

$$\phi(x) = \theta_i + (x - x_i)c_i + (x - x_i)^2 b_i + (x - x_i)^3 c_i \qquad (3)$$

where the coefficients a_i, b_i, and c_i depend linearly on the choice of grid points (x_i, θ_i). In addition, the variations of $\phi(x|\boldsymbol{\theta})$ with respect to each of the parameters θ_i are easily obtained from linear relations.

Other choices of spline fitting procedures using either other boundary conditions, such as clamped cubic splines, or other interpolation methods, such as the Akima splines,[52] may be preferable depending on the shape of the free energy profile. In fact, for the activated chemical process considered in the next section, one finds that the interpolation error decreases with the number of parameters k roughly as $k^{-1/2}$ for the Akima spline and $k^{-1/4}$ for the natural cubic spline, indicating that the Akima spline may be a better choice of parametrization of this particular free energy profile.

WHAM Estimator of the Free Energy Profile. As an example of the maximum likelihood procedure applied to a particular parametrization scheme, consider a simple parametrization of the density in which it is expressed in terms of a set of M discrete constant probabilities $\{p_j\}$ at points $\overline{x}_j$ on the interval $[x_-, x_+]$. The discrete probabilities correspond to defining $p_j = \int_{-\Delta_j/2}^{\Delta_j/2} dx\, p(\overline{x}_j + x)$ to be the probability of finding a reaction coordinate in a region $[\overline{x}_j - \Delta_j/2, \overline{x}_j + \Delta_j/2]$. Defining the binned versions of the normalization constants $z_\alpha = \sum_{j=1}^{M} w_\alpha(\overline{x}_j) p_j$ and noting that the binned data are distributed according to the multinomial distribution, the average log-likelihood for the set of simulation points may be written as[53,54]

$$\hat{l}(f) = \sum_{j=1}^{M} \nu_j \ln p_j - \sum_{\alpha=1}^{S} \lambda_\alpha \ln\Big(\sum_j p_j w_\alpha(\overline{x}_j)\Big)$$

plus terms that are independent of p_j. Here ν_j is the fraction of the total number of points in the S simulations in the range $[\overline{x}_j - \Delta_j/2, \overline{x}_j + \Delta_j/2]$. Maximizing the likelihood with respect to the set of probabilities $\{p_j\}$, or, equivalently, with respect to the normalizations z_α, leads to S nonlinear equations for the parameters z_α of the form[34,35,53−55]

$$p_j = \frac{\nu_j}{\sum_{\alpha=1}^{S} \lambda_\alpha z_\alpha^{-1} w_\alpha(\overline{x}_j)} = \nu_j h_n(\overline{x}_j)$$
$$z_\alpha = \sum_j w_\alpha(\overline{x}_j) p_j \qquad (4)$$

Note that the factor h_n in eq 4 serves to reweight the biased bin probability ν_j to correct for the biasing densities w_α. The value of the dimensionless free energy at point $\overline{x}_j$ in the interval is then estimated to be $\phi_j = -\log(p_j/\Delta_j)$. The procedure in eq 4 is generally known in the literature as the weighted histogram analysis method (WHAM).[34−36]

Although the WHAM procedure results in estimates of the potential of mean force ϕ_j at a finite set of points $\{\overline{x}_j\}$, smooth approximations of the free energy profile over the entire range $[x_-, x_+]$ can be constructed using interpolation schemes based on the set of points $\{(\overline{x}_j, \phi_j)\}$. In this work, we use natural cubic splines with "knot" positions $(\overline{x}_j, \phi_j)$ to construct a profile $\phi(x)$ that minimizes the approximate strain energy $\epsilon_s \approx \int_{x_-}^{x_+} dx\, (\phi''(x))^2$ while passing through the knot positions.[56] This procedure will be used as a standard for comparison with other approaches of constructing smooth potentials of mean force.

In a similar manner, the normalization constants z_α can be obtained without using binned data from the nonlinear relation

$$z_\alpha = \sum_{\gamma=1}^{S} \sum_{i=1}^{n_\gamma} \frac{e^{-\beta u_\alpha(x_i^\gamma)}}{\sum_{\delta=1}^{S} n_\delta e^{-\beta u_\delta(x_i^\gamma)} z_\delta^{-1}} \qquad (5)$$

which can be derived using extended bridge sampling methods.[47] The potential of mean force at a position x_j can then be estimated using eq 4.

Adaptive Improvements of the Constructed Potential of Mean Force. The necessary and sufficient conditions for the existence of a unique solution to the maximum likelihood equation in eq 1 have been established by Vardi[53] as well as the asymptotic behavior of the estimator of the probability density $p(x)$ using the theory of empirical processes.[57,58] In the nomenclature of the theory of empirical processes, the empirical measure places weight $1/n_\alpha$ at each point x_i^α in the α sample, so the empirical average of an arbitrary integrable function f constructed from the data in simulation α is written as

$$\overline{f}^\alpha = \frac{1}{n_\alpha} \sum_{i=1}^{n_\alpha} f(x_i^\alpha)$$

In the limit of large numbers, the central limit theorem implies that the empirical average $\overline{f}^\alpha$ approaches the true average $\langle f\rangle_\alpha$ of f, where $\langle f\rangle_\alpha = \int dx\, f(x) q_\alpha(x)$ is the average of f with respect to the biased density $q_\alpha(x) = z_\alpha^{-1} w_\alpha(x) p(x)$. Furthermore, the empirical average of f is asymptotically normally distributed with mean $\langle f\rangle_\alpha$ and variance $\sigma_\alpha^2(f) = (\langle f^2\rangle_\alpha - \langle f\rangle_\alpha^2)/n_\alpha$, i.e.,

$$\overline{f}^\alpha \xrightarrow{\mathcal{D}} N(\langle f\rangle_\alpha, \sigma_\alpha^2(f)) \qquad (6)$$

Since each of the S simulations is independent, the empirical average over the data points in the entire set of simulations is given by $\overline{f} = \sum_\alpha \lambda_\alpha \overline{f}^\alpha$, which is normally distributed in the large number limit with mean $\langle f\rangle$ and variance $\sigma^2(f) = \sum_\alpha \lambda_\alpha \sigma_\alpha^2(f)$, where the mean $\langle f\rangle = \int dx\, f(x) q(x)$ is defined with respect to the *mixture* density $q(x) = \sum_\alpha \lambda_\alpha q_\alpha(x) = p(x) \sum_\alpha \lambda_\alpha z_\alpha^{-1} w_\alpha(x)$. From this perspective, the composite set of points from the S simulations can be viewed as being drawn from the mixture density $q(x)$. One important difference between the targeted density $p(x)$ and the mixture density $q(x)$ is that, unlike $p(x)$, the mixture density is never very small over the entire domain $[x_-, x_+]$ due to the biasing factor $\sum_\alpha \lambda_\alpha z_\alpha^{-1} w_\alpha(x)$.

The quality of a statistical model of the density $p(x|\boldsymbol{\theta})$ and hence the potential of mean force $\phi(x|\boldsymbol{\theta})$ may be assessed and adaptively improved using goodness-of-fit tests that identify points of the domain in which the deviations between empirical and parametrized cumulative distribution functions are maximal. The goodness-of-fit tests are based on the observation

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

that, for any integrable function f, the central limit theorem (see eq 6) implies that $\overline{f}^{\alpha}$ approaches $\langle f\rangle_{\alpha}$ in the limit of large numbers. Noting that the empirical average of the Heaviside function

$$H(x-y)=\begin{cases}1 & \text{if } x\le y\\ 0 & \text{otherwise}\end{cases}$$

over the *ordered* data set $\{x_1^{\alpha}, ..., x_{n_\alpha}^{\alpha}\}$ with $x_{i+1}^{\alpha}\ge x_i^{\alpha}$ from simulation α yields the monotonic, stepped empirical cumulative distribution function (ECDF)

$$\overline{H(x-y)}^{\alpha}=F_{n\alpha}(y)=\frac{i}{n_\alpha}\quad \text{for}\quad x_i^{\alpha}\le x<x_{i+1}^{\alpha} \tag{7}$$

the central limit theorem implies the empirical cumulative distribution function converges to the true cumulative distribution function (CDF)

$$\langle H(x-y)\rangle_{\alpha}=F_{\alpha}(y)=\int_{x_-}^{y}\mathrm{d}x\; q_{\alpha}(x) \tag{8}$$

for each simulation α. Note that, unlike the empirical density $q_{n\alpha}(x)=\frac{1}{n_\alpha}\sum_{i=1}^{n_\alpha}\delta(x-x_i^{\alpha})$, the empirical cumulative distribution function is finite over the domain and therefore is suitable for comparing the agreement between empirical and true distributions. Similarly, the empirical distribution function $\overline{H(x-y)}=F_n(y)$ for the *composite* process with ordered data points $\{\mathbf{x}^1, ..., \mathbf{x}^S\}$ converges to the true cumulative distribution function $F(y)=\sum_{\alpha}\lambda_{\alpha}F_{\alpha}(y)=\int_{x_-}^{y}\mathrm{d}x\; q(x)$ of the mixture density $q(x)$.

To examine the quality of the parametrization of the potential of mean force, one defines the statistic

$$d_{\alpha}=\sup_{x\in\mathcal{R}}\sqrt{n_\alpha}\sqrt{\psi(F_{\alpha}(x))}\,|F_{n\alpha}(x)-F_{\alpha}(x)| \tag{9}$$

where $\psi(x)$ is a positive weight function. This statistic quantifies the maximum observed (weighted) deviation between the ECDF and CDF and identifies a location x^* where the deviation occurs in the domain. If $\psi=1$, eq 9 corresponds to the usual Kolmogorov−Smirnov statistic for the simulation set α, which is asymptotically distributed according to

$$K(d)=\sum_{k=-\infty}^{\infty}(-1)^k\mathrm{e}^{-2k^2d^2} \tag{10}$$

independent of the form of F. A number of weight functions have been proposed recently with known asymptotic distributions of the maximum deviation.[59−61] One interesting choice[62,63] is to weight the deviations by the inverse of the local variance $\psi(F_{\alpha}(x))=1.0/(F_{\alpha}(x)(1-F_{\alpha}(x)))$ to remove the lack of sensitivity of the unbiased statistic to tails of the distribution. In this work, we use the weight function

$$\psi(F)=\begin{cases}\dfrac{1}{F(1-F)} & 0.15\le F\le 0.85\\ 0 & \text{otherwise}\end{cases}$$

for which the asymptotic distribution function of maximum deviations is known analytically when the tested distribution is independent of the data set.[63]

The goodness-of-fit test for simulation α consists of testing the "null" hypothesis that the n_α ordered data points $\{x_1^{\alpha}, x_2^{\alpha}, ..., x_{n_\alpha}^{\alpha}\}$ in simulation α, where $x_i^{\alpha}>x_{i-1}^{\alpha}$, are drawn from the biased density $q_{\alpha}(x)=z_{\alpha}^{-1}w_{\alpha}(x)p(x)$, which has a cumulative distribution function $F_{\alpha}(x)=\int_{x_-}^{x}\mathrm{d}z\; q_{\alpha}(z)$. The decision of whether to accept or reject the current set of parameters $\boldsymbol{\theta}_k$ is based on the p-value for the test statistic, which is the probability of observing a value of the test statistic at least as large as the maximum observed deviation d_{α} assuming the current model is true (i.e., the data are drawn from $q_{\alpha}(x)$). If the p-value is less than a significance level p_{cut}, typically chosen to be $p_{\mathrm{cut}}=0.05$, it suggests the data are not consistent with the parameter set $\boldsymbol{\theta}_k$, and suggests the parameter set must be enlarged. Similar tests have been used recently in the context of analyzing the distribution of transition times in metadynamics against a predicted Poisson distribution.[64]

The value of the maximum deviation d_{α} for a data set is determined using eq 9, which amounts to computing the absolute value of the difference between the (numerically) calculated $F_{\alpha}(x_i^{\alpha})$, a monotonic function, and the stepped, monotonic empirical cumulative distribution function $F_{n\alpha}$ in eq 7 before and after the step discontinuity at each of the data points x_i^{α} in the simulation. As a result, the maximum deviation d_{α}^* occurs at one of the data points $x_{\alpha}^*=x_j^{\alpha}$ in a set of simulation points. The p-value for the data set, given by $K(d_{\alpha}^*)$ for the Kolmogorov−Smirnov statistic, is then compared to the threshold value to determine whether or not the data are consistent with having been drawn from $q_{\alpha}(x)$. Since the actual density $q_{\alpha}(x)$ is not known, the test consists of replacing q_{α} with the maximum likelihood density $q_{\alpha}(x|\boldsymbol{\theta})=z_{\alpha}^{-1}(\boldsymbol{\theta})w_{\alpha}(x)p(x|\boldsymbol{\theta})$ and hence $F_{\alpha}(x_i^{\alpha})$ by $F_{\alpha}(x_i^{\alpha}|\boldsymbol{\theta})=\int_{x_-}^{x_i^{\alpha}}\mathrm{d}x\; z_{\alpha}^{-1}(\boldsymbol{\theta})w_{\alpha}(x)p(x|\boldsymbol{\theta})$, where the parameter set $\boldsymbol{\theta}$ has been optimized for the data set. In general, this biases the distribution of the statistic d_{α} relative to $K(x)$, which is generally underestimated, leading to an overestimation of the p-value. In such cases, the actual distribution of d_{α} is unknown and dependent on the parameter set. Nonetheless, the correct p-value may be estimated using parametric bootstrap methods[65,66] in which a number N_s of synthetic samples, each of n_α points, is drawn from the parametrized density $q_{\alpha}(x|\boldsymbol{\theta})$. The maximum deviation d_s for each of these samples is computed after reoptimizing the parameters $\boldsymbol{\theta}$ and the p-value is estimated as

$$p=\frac{\text{number of simulations with } d_s\ge d_{\alpha}^*}{N_s} \tag{11}$$

which has a standard deviation of $\sigma_p=\sqrt{p(1-p)/N_s}$. Since the bootstrap procedure can be computationally intensive because it requires the recalculation of the $\boldsymbol{\theta}$ parameters that maximize the likelihood for the synthetic data, it is useful to first determine if the p-value using $K(x)$ is below the threshold value p_{cut} and the fit is poor, since $K(d_{\alpha})$ overestimates the actual p-value.

The procedure described above provides a test of the quality of fit of the parametrized density using only the data from a single simulation. Since typically the biased simulations intentionally have points distributed in a small subset of the interval $\mathcal{R}$ of interest, the test does not examine the quality of fit outside this region and says little about the global fit of $p(x|\boldsymbol{\theta})$. To adapt this procedure for the composite data set from all of the biased simulations, the maximum deviations can be computed for each of the S simulations to compile a set $\{d_1, ..., d_S\}$, each of which is independently distributed according to some distribution $K(x)$. The overall quality of fit of the density

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

can be assessed by taking the maximum d_{max} of the d_α deviations, which is then distributed according to $K(d_{max})^S$. Once again, if $p(x)$ is parametrized and the distribution $K(x)$ is unknown, bootstrap methods must be used to estimate the p-value of d_{max}, which will typically be smaller than $K(d_{max})^S$.

To test the quality of the construction of the free energy profile over the entire range, we introduce a new goodness-of-fit test. The test makes use of the global data set $\{\mathbf{x}^1, ..., \mathbf{x}^S\}$ and defines the maximum deviation in terms of the empirical distribution for the entire data set, $F_n(x)$, which converges to the cumulative distribution function $F(x)$ of the mixture density $q(x)$. The procedure is analogous to that of the individual simulations with a maximum deviation d defined with unit weight function by

$$d = \sup_{x\in\mathcal{R}} \sqrt{n}|F_n(x) - F(x)| \tag{12}$$

where $F_n(x)$ is the empirical distribution function for the global data set. Unlike the case of the individual simulations, the distribution of the statistic d depends on the form of the weight functions w_α and is more narrowly distributed than the Kolmogorov−Smirnov distribution in eq 10 when $\psi = 1$, since the variance $\sigma^2(F) = \sum_\alpha \lambda_\alpha F_\alpha(y)(1 - F_\alpha(y))$ is smaller than that for points drawn from a mixture density. In the limit where there is negligible overlap between adjacent simulations, one expects $(S)^{1/2}d$ to be approximately distributed by $K(d)$ under the null hypothesis. Nonetheless, as above, the p-value must be determined by bootstrap estimation, since $F(x)$ is unknown and is approximated by $F(x|\boldsymbol{\theta})$, which depends parametrically on the data.

At first glance, it seems reasonable to develop goodness-of-fit tests for the unbiased probability density $p(x)$ directly, since this probability density defines the potential of mean force. This would be profitable if the quality of the fit of the unbiased density in regions where the density is large is particularly important. Such tests can be devised using weighted empirical distribution functions, since $\overline{h_n(x)H(x - y)} = \tilde{F}_n(y)$ converges in distribution to the cumulative distribution function $\langle h_n(x) H(x - y)\rangle = \tilde{F}(y) = \int_{x_-}^{y} dx\ p(x)$ of the unbiased density. However, in chemical applications, one is interested not in the quality of the fit of the peaks of multimodal densities but rather in the quality of fit of the negative logarithm of the density, the potential of mean force, in regions where the density itself is very small. In fact, since the slope of the cumulative distribution function of $p(x)$ is the density itself, the cumulative distribution function has a large plateau region at intermediate points in the domain R in the barrier region, and deviations between cumulative and empirical cumulative distribution functions are not particularly sensitive to fits in these regions. The cumulative distribution function of the mixed density $F(x)$ does not exhibit large plateaus, since the mixture density $q(x)$ is never small in the domain $[x_-, x_+]$.

Failure of the goodness-of-fit tests indicates that the current set of parameters used to construct the potential of mean force is inadequate. The statistic based on the maximum weighted deviations of the ECDF and CDF also provides useful information about which region of the domain is poorly characterized. The deviations will only be reduced if improvements in the construction of the free energy profile are made in the vicinity of the problematic region. When local parametrizations of the free energy profile are used, the quality of the fit in the problematic region near the point x^* can be enhanced without significant changes to the fit outside this region by inserting additional fitting parameters such as additional grid points or localized basis functions centered at x^*. The central idea of the adaptive approach is to first identify problematic points in the interval in which large deviations between the cumulative and empirical cumulative distribution functions are observed. These points are then used to expand the parameter set of the potential of mean force locally in the vicinity of the problematic points until p-value convergence is obtained for the relevant goodness-of-fit test. In this work, we define an adaptive parametrization procedure based on the goodness-of-fit test described in eq 12.

The adaptive approach is summarized by the following algorithm:

Algorithm 1 Adaptive knot positioning

1: Initialize number of grid points $k = S$
2: Initialize $p = 0$, $k_{start} = S$, $d_c = \infty$
3: Select k grid points $\{x_i\}$ uniformly on interval
4: **while** $p < p_{cut}$ and $k_{start} < k_{max}$ **do**
5: Minimize average log-likelihood to find optimal ϕ_i^*
6: Compute d and x^* using Eq. (12)
7: Set $p = 0$
8: **if** $d < d_c$ **then**
9: Draw n_b parametric bootstrap samples using ϕ_i^*
10: Optimize new ϕ_i for each bootstrap sample
11: Find d for each bootstrap sample
12: Order set of n_b deviations, $\{d_1, \ldots d_{n_b}\}$
13: Use Eq. (11) to compute p
14: Update $d_c = d_i$, with $i = 5n_b/6$
15: **end if**
16: **if** x^* is not in set of x_i **then**
17: Set $k = k + 1$ and augment set $\{x_i\}$ with x^*.
18: **else**
19: Set $k_{start} = k_{start} + 1$ and $k = k_{start}$.
20: Select new set of k uniform grid points.
21: Reset $d_c = \infty$
22: **end if**
23: **end while**

One caveat to using an automated adaptive procedure with parameter updates based on the location of maximum deviations is the possibility of finding the same point x^* of maximum deviation between iterations and hence an insertion point that is already in the set of parameter grid points $\{x_i\}$. Such scenarios can be handled in a variety of ways, such as redistributing the grid points in the vicinity of x^* or starting the procedure afresh from a uniform grid of larger size (as in lines 19−21 in Algorithm 1 above). Note that the test in line 8 is included only to increase the efficiency of the adaptive procedure, since the estimation of p-values from the bootstrapped synthetic samples can be computationally demanding albeit simple to evaluate in parallel. Other useful strategies include initially setting d_c to a value based on the asymptotic distribution of deviations valid when the cumulative distribution function is independent of the data to avoid the computational cost of estimating d_c when the p-value is very small.

Algorithm 1 can be modified to include multiple goodness-of-fit tests simultaneously such as those in eq 9 and in eq 12 by calculating the p-value of different tests as well as the locations of their maximum deviations and requiring that all p-values exceed some threshold. Combining different tests may be beneficial if the tests weight deviations differently in different regions of the domain, and may help to avoid repeatedly finding the same insertion point.

Many other tests of the goodness-of-fit can be implemented and their utility compared to the method of maximum

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

deviations. Perhaps the simplest and most obvious test for convergence of the potential of mean force with respect to a set of parameters is to monitor when the average weighted integrated squared difference between two models that differ in the number of parameters drops below some threshold value. We define the L^2-difference between two curves $f_1(x)$ and $f_2(x)$ by

$$\hat{f}(x) = f(x) - \int_{x_-}^{x_+} dx\, f(x)\rho(x)$$

$$L^2(f_1, f_2) = \int_{x_-}^{x^+} dx\, (\hat{f}_1(x) - \hat{f}_2(x))^2 \rho(x) \tag{13}$$

where the density $\rho(x)$ may be defined to selectively weight different regions of the interval $[x_-, x_+]$. Here, we use a uniform density $\rho(x) = 1/(x_+ - x_-)$. The subtractions are necessary in eq 13, since the free energy profiles are only determined up to a constant. Although in most cases the maximum likelihood procedure applied to determine a given parameter set $\boldsymbol{\theta}_k$ has a unique solution in the large n limit, a small L^2-difference between two sets of parameters of differing size does not necessarily mean that quality of the fit of the constructed potential of mean force to the data is high—two parameter sets may produce similar free energy profiles and yet yield poor fits.

In many instances, the models of the free energy profile are nested, meaning that models of the profile with a smaller number of parameters can be obtained from those with a larger number by imposing a set of constraint conditions on a subset of the parameters, such as restricting the values of some parameters to a particular set of values (for example, zero). For such nested models, a way to select and optimize models from a set of parameters is to use the log-ratio test[67] or the Akaike information criterion[68,69] to define weights w_k for the model with k parameters. The optimized model is then written as a weighted sum of each of the models, $\phi(x) = \sum_k w_k \phi(x|\boldsymbol{\theta}_k)$. The Akaike information criterion is defined for a model with k parameters and log-likelihood $\hat{l}(\boldsymbol{\theta}_k)$ by computing $a_k = 2k - 2\hat{l}(\boldsymbol{\theta}_k)$ for a range of the maximum number of parameters $k = 1, ..., k_m$ and identifying the minimum value $a_{\min}$. The Akaike weights w_i, which provide the likelihood of model i, are then defined as[70]

$$w_i = \frac{e^{-\Delta_i/2}}{\sum_k e^{-\Delta_k/2}} \tag{14}$$

where $\Delta_i = a_i - a_{\min}$. The relative performance of this convergence test on global parametrizations of the free energy profile will be compared with the adaptive scheme as well as that obtained from an optimally selected weighted histogram approach.

In the following section, a numerical study is conducted on simple model systems to evaluate the performance of the adaptive algorithm relative to other methods of evaluating the free energy profile and to assess the use of other convergence criteria for various parametric methods of constructing the free energy profile. The results presented make use of the simultaneous application of the goodness-of-fit tests presented above, with the critical p-value set at a relatively high level of $p_{\text{cut}} = 0.15$ to discourage early convergence at the cost of potentially assigning too much significance to noise.

RESULTS

Case Study of an Activated Chemical Processs. In order to assess the properties of statistical convergence tests on the accuracy of the constructed smooth potential of mean force, it is essential to consider systems for which the actual potential of mean force is either known analytically, or with great statistical precision. We first consider the convergence tests outlined above applied to constructing smooth free energy profiles on a two-dimensional system designed to mimic a typical activated process between a stable and a metastable state with a substantial activation energy between them. The activated system has an (un-normalized) bimodal density given by the sum of two normal (Gaussian) densities

$$\rho(x, y) = a_1 e^{-b_1(x-x_1)^2} e^{-b_1(y-y_1)^2} + a_2 e^{-b_2(x-x_2)^2} e^{-b_2(y-y_2)^2} \tag{15}$$

centered at $x_1 = y_1 = -2$, $x_2 = y_2 = 2$ with height and width parameters $a_1 = 1 \times 10^5$, $a_2 = 2.5$, $b_1 = 2$, and $b_2 = 6$. The reduced density for the reaction coordinate x, $p_a(x) = \int_{-\infty}^{\infty} dy\, \rho(x, y)$, is therefore given by a bimodal density corresponding to an asymmetric double well potential of mean force

$$p_a(x) = \frac{b_1^2}{\sqrt{\pi}(a_1 b_2 + a_2 b_1)} \left(\frac{a_1}{\sqrt{b_1}} e^{-b_1(x-x_1)^2} + \frac{a_2}{\sqrt{b_2}} e^{-b_2(x-x_2)^2} \right) \tag{16}$$

The two maxima of $p_a(x)$, and hence the two local minima of $\phi_a(x) = -\ln p_a(x)$, are approximately located at $x_- = x_1 = -2$ and $x_+ = x_2 = 2$. Here the activation energy is roughly 16, in dimensionless units.

Umbrella sampling simulations with an unnormalized biasing weight

$$w_\alpha(x) = e^{-(K/2)(x-x_\alpha)^2} \tag{17}$$

with $K = 25$ were used to facilitate sampling data to reconstruct $\phi(x)$ over the entire domain $\mathcal{R} = [x_-, x_+]$. The form of the biasing density used here is not important in the data analysis, and in principle, any bias can be used, including a single global bias such as those frequently used in adaptive importance sampling procedures[14−22] and metadynamics.[23−25] In the simulation studies here, the central positions of the umbrella densities are chosen by centering densities at the limits of the interval x_- and x_+ as well as near the maximum of the potential of mean force (i.e., the barrier) at a position $x_b = 0.8$. For simulations with more than three umbrella windows, the remaining densities are evenly spaced over the intervals $[x_-, x_b]$ and $[x_b, x_+]$.

As is evident from the plots of the biased densities $q_\alpha(x)$ in Figure 1, the regions where the $q_\alpha(x)$ overlap are small for the smallest number of umbrella windows ($S = 5$), and the data from different umbrella simulations are not likely to overlap unless a large number of data points per window are generated. Indeed, one finds that the probability of overlap of adjacent windows for the first few windows in a five-window system is less than 0.01. As the overlap of data samples between biased simulations is essential[53] to have a unique solution of eq 4, convergence problems can arise due to inadequate sampling when the number of umbrella windows S is small, and otherwise lead to poor fits. Clearly, better choices for the biasing umbrella densities using variable width parameters or

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

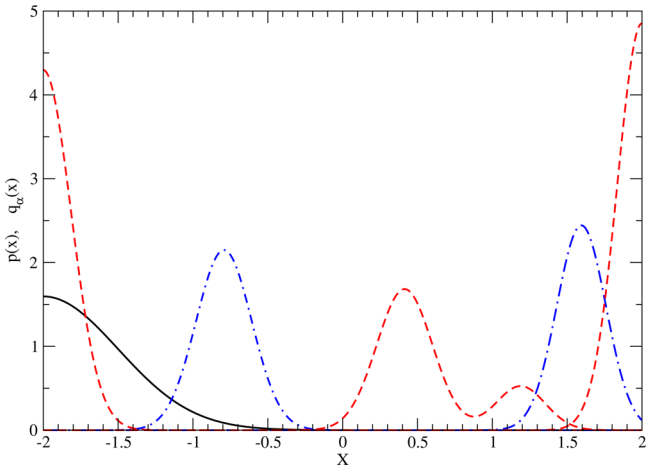


Figure 1. Analytical target density $p_a(x)$ (black, solid line) and the biased densities $q_\alpha(x)$ for a five-umbrella system. The red curves correspond to the anchoring umbrella densities centered at x_-, x_b, and x_+.

other functional forms for the biasing densities could be constructed, since the analytical density is known.

A plot of the dimensionless analytical potential of mean force is shown in Figure 2 showing a barrier with a maximum value of

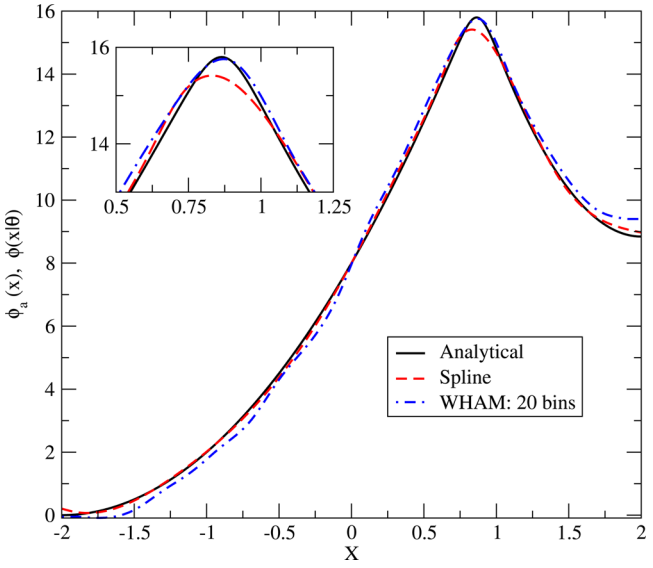


Figure 2. Analytical potential of mean force $\phi_a(x)$ (black, solid line) and example fits based on data from seven windows with 1000 points per window using spline knots as parameters (red curve) and an interpolation of the 20-bin WHAM solution using natural cubic splines (blue curve).

15.9 at $x = 0.865$ separating the minima at $x = -2$ and $x = 2$. Note that the density in the region near $x = -2$ is roughly $e^9 \approx 8100$ times larger than in the region near $x = 2$ and roughly $e^{16} \approx 8.9 \times 10^7$ times larger than in the barrier region. As is clear from Figure 3, the analytical cumulative distribution function of the unbiased density $\tilde{F}(x)$ has a long plateau region where the density of the bimodal system is small, whereas the cumulative distribution function $F(x)$ of the mixture density $q(x)$ does not exhibit such behavior and is therefore better suited for global goodness-of-fit tests when the quality of the fit of the potential of mean force over the entire interval is of interest.

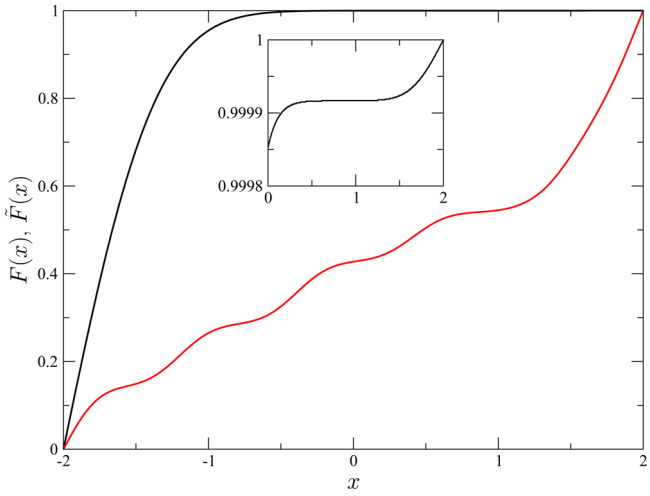


Figure 3. Analytical cumulative distribution $F(x)$ of the mixture density $q(x)$ (red) and the cumulative distribution $\tilde{F}(x)$ of the unbiased target density $p(x)$ (black). The inset shows the detail of $\tilde{F}(x)$ in the barrier region where the density $p(x)$ is very small.

A total of 5000 data sets consisting of $S = 5$ to $S = 11$ umbrellas with $n_\alpha = 50-1000$ data points per window was carried out, and the free energy profiles were analyzed using the fitting methods described in the Methodology section. Estimates of the errors arising through the approximation or interpolation schemes used to represent $\phi(x|\boldsymbol{\theta})$ and through finite sampling were computed using the integrated squared difference defined in eq 13 between the fitted free energies $\phi(x|\boldsymbol{\theta}^*)$ and the analytical profile $\phi_a(x) = -\ln p_a(x)$, i.e., $\epsilon(\boldsymbol{\theta}^*) = L^2(\phi(x|\boldsymbol{\theta}^*), \phi_a(x))$. The integrated squared error $\epsilon(\boldsymbol{\theta}^*)$ is nonzero due to systematic deviations that arise from two sources: finite sampling where $\boldsymbol{\theta}_k^* \neq \boldsymbol{\theta}_k^0$ and histogram approximations and interpolation errors. In the limit of large samples, the integrated error should approach zero as the number of parameters used to represent the free energy profile increases.

For the spline-fitted WHAM procedure with the number of bins ranging from 5 to 30, one finds that for the bimodal model the optimal choice of the number of bins depends rather strongly and unpredictably on the number of umbrella simulations S (windows) and the number n_α of points sampled per window. As is evident in Table 1, for a fixed number of windows, the optimal number of bins increases with the number of points per window and the integrated error squared, L^2-error (denoted as ϵ_i for a simulation with $S = i$ windows), decreases roughly as n_α^{-1}. The statistical uncertainties in the error estimate are on the order of 1%. As expected, for fixed n_α, the absolute errors diminish at a faster rate than that predicted due to an increased number of sampling points (i.e., $\epsilon_i/\epsilon_{i+2} > (i + 2)/i$), indicating an improvement in the sampling over the interval $[x_-, x_+]$.

In Figure 4, the relative performance of the adaptive convergence test based on algorithm 1 with a cubic spline representation of the potential of mean force is shown. In the left panel, the performance is measured relative to the *optimal* number of bins listed in Table 1 selected for all values of the number of windows and number of points per window, which are of course unknown *a priori* but represent the best possible choice of the number of bins for the simulation conditions. Note that the adaptive scheme which places spline knot positions at points of maximal deviation d^* outperforms the

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 1. Wham Errors

		n_α				
		50	100	200	500	1000
5 windows	optimal bin number	7	9	9	10	14
	error ϵ_5	2.402	1.370	1.064	0.623	0.340
7 windows	optimal bin number	11	15	15	23	23
	error ϵ_7	0.580	0.271	0.136	0.064	0.033
9 windows	optimal bin number	14	14	20	23	23
	error ϵ_9	0.310	0.144	0.075	0.031	0.017
11 windows	optimal bin number	14	17	19	23	23
	error ϵ_{11}	0.226	0.111	0.054	0.023	0.012

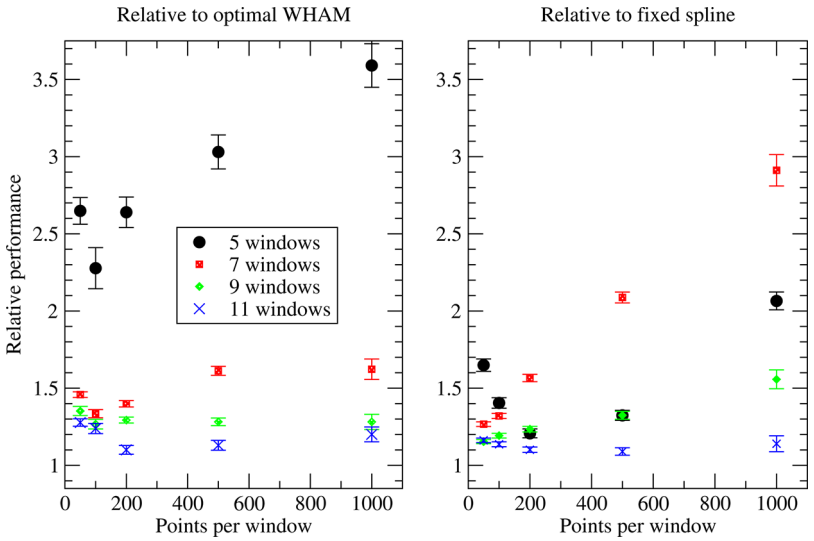


Figure 4. Performance of the adaptive procedure with cubic spline parametrization of potential of mean force. The points in the left panel represent the ratio of the integrated squared error of the adaptive scheme to that of the *optimal* WHAM procedure (left panel) and to that of a fixed number of uniformly spaced knots (right panel) for a given number of windows and points per window. The error bars are 95% confidence intervals.

optimal WHAM procedure by a factor of 2.3−3.6 when the overall quality of the sampling is poor (i.e., only five windows in the umbrella simulation), which is the typical case when sampling an unknown system with high barriers in the potential of mean force. For moderate sampling with seven or nine windows, the performance relative to the optimal interpolated WHAM fits is near 1.5 for all values of the points per window. On the other hand, when the sampling is extensive and the statistical representation of the free energy profile is excellent, the adaptive procedure still outperforms the optimal WHAM sampling but only by a relatively modest 10−20%. Nonetheless, the adaptive procedure does not require any parameters to be chosen or adjusted. In the right panel of Figure 4, the performance of the adaptive procedure relative to that in which the number of knots is fixed and the values of the reaction coordinate are uniformly distributed is plotted. Since the optimal number of knots n is not known, following Lee et al.,[48] n is set equal to one less than twice the number of biasing windows so that knots are positioned near the center and between each of the maxima of the biasing densities. Once again, the most impressive performance gains, ranging from factors of 1.3 to 3, occur when the data come from relatively few biasing simulation windows.

Global parametrizations of the potential of mean force generally do not converge with respect to the number of parameters required as quickly as local parametrizations using the goodness-of-fit test and have larger L^2-errors. For nested global parametrizations, the log-ratio test and the weighted Akaike information criterion perform comparably with less impressive relative performance values than the adaptive scheme based on the goodness-of-fit test for all sampling conditions investigated here, as can be seen in Figure 5. However, the convergence tests do not require a choice of the number and sizes of bins yet produce superior results to the optimal interpolated WHAM construction of the potential of mean force.

In spite of the inferior performance of the global parametrization scheme under the nested convergence tests compared to the adaptive algorithm, the log-ratio and Akaike information criterion are attractive, since they do not require additional bootstrap sampling and are therefore much less computationally demanding, which may be an important consideration when data sets are large. They constitute very simple convergence tests that require no adjustable parameters yet provide more reliable statistical fits of the potential of mean force than cubic spline interpolations of the optimal WHAM procedure. In addition, the global parametrizations and nested convergence tests are simple to extend to higher dimensions.

Another simple test for convergence of the maximum likelihood procedures is to compute the L^2-difference (average integrated squared difference) between two successive global parametrizations of the potential of mean force. One finds that this convergence test (with a choice of convergence threshold of 5×10^{-4}) underperforms relative to the log-ratio test and

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

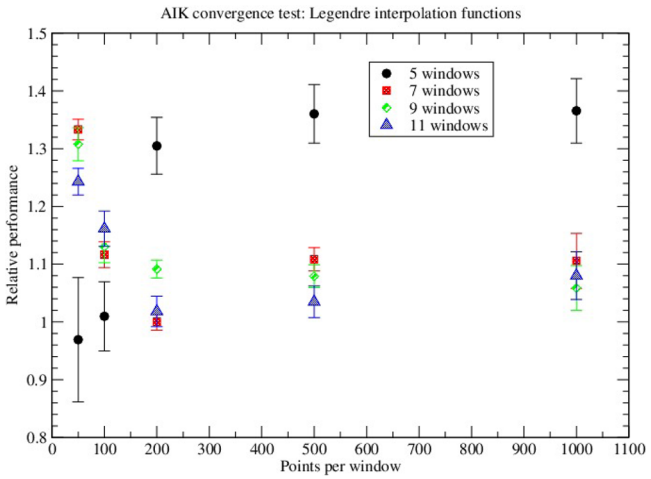


Figure 5. Performance of the Akaike information convergence procedure for a Legendre polynomial parametrization of the potential of mean force. The points represent the ratio of the L^2-error of the information criterion convergence test to the *optimal* WHAM procedure for a given number of windows and points per window. The error bars represent 95% confidence intervals.

Akaike information procedure, with typical relative peformance numbers near unity when there are more than 200 points per window (data not shown).

Through extensive numerical study, it is also found that other methods of constructing a smooth analytical fit for the potential of mean force outlined in the Methodology section, such as using linear combinations of local fits to the cumulative distribution function[32] that minimize the variance of the mixture density and the method of weighted residuals,[45] are found to perform worse than the methods based on a maximum likelihood approach, including the WHAM procedure. This is likely due to the fact that the variance of the expansion parameters in the method of weighted residuals is proportional to fluctuations of the generalized force, and these fluctuations can be large in multidimensional systems.

Potential of Mean Force of a Solvated Ion Pair. To examine the utility of using the adaptive convergence procedure to construct other types of free energy profiles, the effective pair potential for a Na^+ and Cl^- ion-pair immersed in TIP3P water at a temperature of T = 300 K and pressure of 1 atm was computed using the simulation data available via the vFEP software package.[71] These data consist of 10000 ion separation distances measured in angstroms obtained from 21 biased simulations using a biasing density of the form in eq 17 with centers x_c ranging from 2.4 to 7.4 Å located at intervals of 0.25 Å, each with a force constant of $K = K'/(k_BT)$, where K' = 20 kcal/(mol Å^2). Since the exact analytical potential of mean force is not known for this system, the baseline potential of mean force (shown in Figure 6) was estimated using all of the simulation data (21 windows with 10000 data points per window) and error estimates were computed using the average integrated squared differences relative to this profile.

Numerical experiments were conducted by choosing subsets of the full simulation data uniformly, where the experiments differed in the number of windows and the number of points per window. It has been shown[71] that a procedure based on a maximum likelihood procedure applied to a natural cubic spline parametrization of the free energy profile with a fixed number of knots generally outperforms the weighted histogram[34,35] and

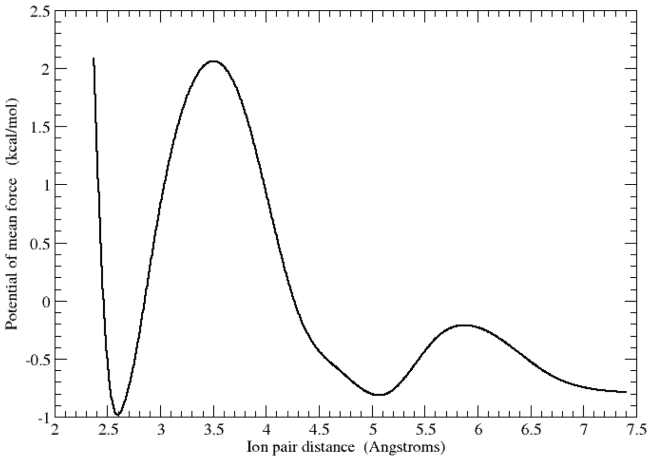


Figure 6. Cubic spline interpolation of the potential of mean force for a solvated Na^+−Cl^- ion pair as a function of distance between the ions at a temperature of 300 K and 1 atm of pressure.

MBAR methods.[47] In Figure 7, the performance of the adaptive maximum likelihood procedure for a cubic spline representa-

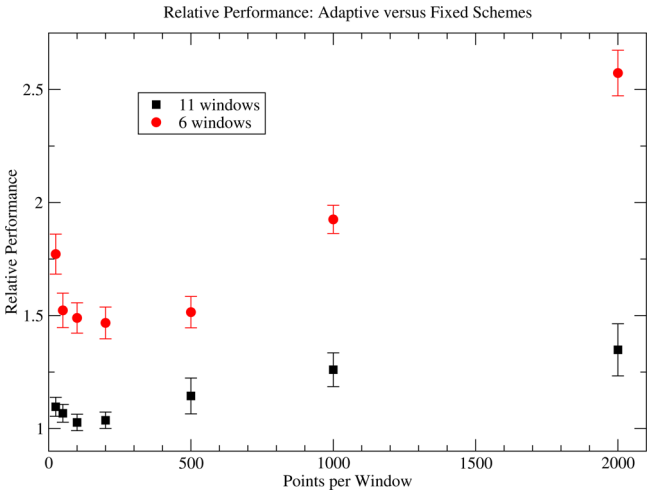


Figure 7. Relative performance of the adaptive convergence algorithm to that using a fixed number of parameters measured as the ratio of the integrated squared error or the ratio of the variance of the free energy in a cubic spline representation of the potential of mean force for the solvated ion pair system. The fixed number n of equally spaced knots was set equal to $n = 2S - 1$, where the number of biasing densities S was either 6 or 11. The error bars represent 95% confidence intervals.

tion of the potential of mean force relative to that using a fixed number n of knot positions as parameters is shown. The results were obtained by averaging over 1000 numerical experiments for each value of the total number of windows and number of points per window. Once again, we set the number of spline points n equal to one less than twice the number of biasing windows so that knots are positioned near the center and between each of the maxima of the biasing densities. It is found that on average the adaptive procedure provides significantly better fits to the data than the alternative with a fixed number of parameters for all cases examined. Similarly to what was observed in the study of the activated system, the largest improvements are observed when the sampling is relatively poor where the statistical errors for the adaptive procedure are roughly half those of the alternative with a fixed number of

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

parameters. Once again, in the atypical case where the sampling is extensive and the statistical representation of the free energy profile is excellent, the adaptive procedure outperforms the fixed alternative but only by a relatively modest 5−10%.

DISCUSSION AND CONCLUSIONS

For parametric fits of the potential of mean force of a single scalar quantity, the optimal approach was found to use the method of maximum likelihood in combination with a local basis set parametrization of the free energy profile, where the local placement of the basis set is determined by statistical considerations such as the locations of maximum deviations between the empirical cumulative and cumulative distribution functions. The resulting free energy profiles represent the data with statistical significance yet are relatively smooth, free of adjustable parameters, and easy to compute.

Goodness-of-fit and other statistical convergence tests can also be utilized with recently proposed methods of constructing the potential of mean force that are not based on the principle of maximum likelihood, such as linear regression.[43] For example, as an alternative to histogram constructions, the biased densities from each of the umbrella simulations could be fit with a parametric maximum-likelihood procedure using a convergence test to form estimates of the response variable at sampled values of the reaction coordinate. In turn, these data could then be used in a linear regression procedure, where the number and *locations* of basis set functions used to represent the global density are controlled using the goodness-of-fit criteria.

The parametric construction of the potential of mean force can also be used to decide if it is desirable to insert additional sampling windows by comparing the overlap of adjacent biased densities $O_\alpha(\boldsymbol{\theta}) = \int_{x_-}^{x_+} \mathrm{d}x\; q_\alpha(x|\boldsymbol{\theta}) q_{\alpha+1}(x|\boldsymbol{\theta})$ to one another. If a particular O_α is small for one of the sampling windows, it suggests that additional sampling should be carried out in the region of the domain between the maxima of the densities q_α and $q_{\alpha+1}$. Another alternative is to use the locations of the maximum deviation in the global goodness-of-fit test to identify regions of the domain where increased sampling would improve the statistical quality of the parametrization.

The construction of smooth fits to multidimensional potentials of mean force is generally more problematic than the one-dimensional case, and typically requires a large number of samples from many biased simulations. Nonetheless, convergence criteria for the number of parameters required for the multidimensional free energy profiles can be defined on the basis of multivariate goodness-of-fit tests. For example, nearest neighbor distances with respect to a norm in the multidimensional parameter space can be used to define a statistic whose distribution is asymptotically uniform and independent of the underlying density.[72−74] For low dimensions, another option is to use the distribution-free multivariate Kolmogorov−Smirnov goodness-of-fit test.[75] The points at which the maximum deviation of the weighted empirical process and the uniform cumulative distribution function occur can be used to refine the parameters used in the model to construct an approximate smooth multidimensional potential of mean force in a similar fashion to the adaptive procedure presented here for a one-dimensional system. For multidimensional systems, multivariate Akima splines[76] or Gaussian density representations of the free energy profile provide tractable ways of parametrizing multivariate densities.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jmschofi@chem.utoronto.ca.

ORCID

Jeremy Schofield: 0000-0001-5366-559X

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Sciences and Engineering Research Council of Canada.

REFERENCES

(1) Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3*, 107.

(2) Chandler, D. Statistical Mechanics of Isomerization Dynamics in Liquids and the Transition State Approximation. *J. Chem. Phys.* **1978**, *68*, 2959−2970.

(3) Szabo, A.; Schulten, K.; Schulten, Z. First Passage Time Approach to Diffusion Controlled Reactions. *J. Chem. Phys.* **1980**, *72*, 4350−4357.

(4) Schulten, K.; Schulten, Z.; Szabo, A. Dynamics of Reactions Involving Diffusive Barrier Crossing. *J. Chem. Phys.* **1981**, *74*, 4426−4432.

(5) Schofield, J.; Oppenheim, I. The Hydrodynamics of Inelastic Granular Systems. *Phys. A* **1993**, *196*, 209−240.

(6) Romero-Rochin, V.; Oppenheim, I. Relaxation Properties of Weakly Coupled Classical Systems. *J. Stat. Phys.* **1988**, *53*, 307.

(7) Oppenheim, I.; McBride, J. Transport Equations for Suspensions of Inelastic Particles. *Phys. A* **1990**, *165*, 279−302.

(8) McGreevy, R.; Pusztai, L. Reverse Monte Carlo Simulation: A New Technique for the Determination of Disordered Structures. *Mol. Simul.* **1988**, *1*, 359−367.

(9) Lyubartsev, A.; Laaksonen, A. Calculation of Effective Interaction Potentials from Radial Distribution Functions: A Reverse Monte Carlo Approach. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1995**, *52*, 3730.

(10) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving Effective Mesoscale Potentials from Atomistic Simulations. *J. Comput. Chem.* **2003**, *24*, 1624−1636.

(11) Moore, T.; Iacovella, C.; McCabe, C. Derivation of Coarse-Grained Potentials via Multistate Iterative Boltzmann Inversion. *J. Chem. Phys.* **2014**, *140*, 224104.

(12) Ercolessi, F.; Adams, J. Interatomic Potentials from 1st-Principles Calculations - The Force Matching Method. *Europhys. Lett.* **1994**, *26*, 583−588.

(13) Izvekov, S.; Voth, G. A Multi-Scale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.

(14) West, M. *Bayesian Statistics*; Oxford University Press: Oxford, U.K., 1992; pp 503−525.

(15) Oh, M.; Berger, J. Integration of Multimodal Functions by Monte Carlo Importance Sampling. *J. Am. Stat. Assoc.* **1993**, *88*, 450−456.

(16) Ortiz, L.; Kaelbling, L. *Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence*; Morgan Kaufmann Publishers: San Francisco, CA, 2000.

(17) Liu, J.; Liang, F.; Wong, W. A Theory of Dynamic Weighting in Monte Carlo Computation. *J. Am. Stat. Assoc.* **2001**, *96*, 561−573.

(18) Pennanen, T.; Monte, M. *Monte Carlo and Quasi-Monte Carlo methods 2004*; Springer-Verlag: Berlin, 2004; pp 443−455.

(19) Cappé, O.; Guillin, M.; Marin, J.; Robert, C. Population Monte Carlo. *J. Comput. Graph. Statist.* **2004**, *13*, 907−929.

(20) Douc, R.; Guillin, A.; Marin, J.-M.; Robert, C. Convergence of Adaptive Mixtures of Importance Sampling Schemes. *Ann. Statist.* **2007**, *35*, 420−448.

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(21) Douc, R.; Guillin, A.; Marin, J.-M.; Robert, C. Minimum Variance Importance Sampling via Population Monte Carlo. *ESAIM: Probability and Statistics* **2007**, *11*, 427−447.
(22) Cornuet, J.-M.; Marin, J.-M.; Mira, A.; Robert, C. Adaptive Multiple Importance Sampling. *Scandinavian J. Statist.* **2012**, *39*, 798−811.
(23) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(24) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603−020606.
(25) Tiwary, P.; Parrinello, M. A Time-Independent Free Energy Estimator for Metadynamics. *J. Phys. Chem. B* **2015**, *119*, 736−742.
(26) Chen, M.; Cuendet, M. A.; Tuckerman, M. E. Heating and Flooding: A Unified Approach for Rapid Generation of Free Energy Surfaces. *J. Chem. Phys.* **2012**, *137*, 024102.
(27) Rosenblatt, M. Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Stat.* **1956**, *27*, 832−837.
(28) Parzen, E. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* **1962**, *33*, 1065−1076.
(29) Rasmussen, C.; Williams, C. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning series)*; MIT Press: Cambridge, MA, 2005.
(30) Silverman, B. *Density Estimation for Statistics and Data Analysis*, 1st ed.; Chapman & Hall: London, 1986.
(31) Berg, B. A.; Harris, R. C. From Data to Probability Densities Without Histograms. *Comput. Phys. Commun.* **2008**, *179*, 443−448.
(32) van Zon, R.; Schofield, J. Constructing Smooth Potentials of Mean Force, Radial Distribution Functions, and Probability Densities from Sampled Data. *J. Chem. Phys.* **2010**, *132*, 154110.
(33) Torrie, G. M.; Valleau, J. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(34) Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J. The Weighted His- togram Analysis Method for Free-Energy Calculations on Biomolecules: I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(35) Roux, B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(36) Zhu, F.; Hummer, G. Convergence and Error Estimation in Free Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **2012**, *33*, 453−465.
(37) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. Constrained Reaction Coordinate Dynamics For the Simulation of Rare Events. *Chem. Phys. Lett.* **1989**, *156*, 472−477.
(38) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Free Energy Calculation From Steered Molecular Dynamics Simulations Using Jarzynski's Equality. *J. Chem. Phys.* **2003**, *119*, 3559.
(39) Kästner, J.; Thiel, W. Bridging the Gap Between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: Umbrella Integration. *J. Chem. Phys.* **2005**, *123*, 144104.
(40) Kästner, J. Umbrella Integration in Two or More Reaction Coordinates. *J. Chem. Phys.* **2009**, *131*, 034109.
(41) Kästner, J. Umbrella Integration With Higher-Order Correction Terms. *J. Chem. Phys.* **2012**, *136*, 234102.
(42) Buhmann, M. D. Radial Basis Functions. *Acta Numerica* **2000**, *9*, 1−38.
(43) Meng, Y.; Roux, B. Efficient Determination of Free Energy Landscapes in Multiple Dimensions from Biased Umbrella Sampling Simulations using Linear Regression. *J. Chem. Theory Comput.* **2015**, *11*, 3523−3529.
(44) Stecher, T.; Bernstein, N.; Csányi, G. Free Energy Surface Reconstruction From Umbrella Samples Using Gaussian Process Regression. *J. Chem. Theory Comput.* **2014**, *10*, 4079−4097.
(45) Cyr, E. C.; Bond, S. D. Using the Method Of Weighted Residuals To Compute Potentials Of Mean Force. *J. Comput. Phys.* **2007**, *225*, 714−729.
(46) Mezei, M.; Beveridge, D. Free Energy Simulation. *Ann. N. Y. Acad. Sci.* **1986**, *482*, 1−23.
(47) Shirts, M.; Chodera, J. Statistically Optimal Analysis Of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.
(48) Lee, T.-S.; Radak, B.; Pabis, A.; York, D. A New Maximum Likelihood Approach To Free Energy Profile Construction From Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9*, 153−164.
(49) Broyden, C. A Class Of Methods For Solving Nonlinear Simultaneous Equations. *Mathematics of Computation* **1965**, *19*, 577−593.
(50) Fletcher, R. *Practical Methods Of Optimization*, 2nd ed.; John Wiley and Sons: New York, 1987.
(51) Redner, R. A.; Walker, H. Mixture Densities, Maximum Likelihood and the EM Algorithm. *SIAM Rev.* **1984**, *26*, 195−239.
(52) Akima, H. A New Method Of Interpolation and Smooth Curve Fitting Based On Local Procedures. *J. Assoc. Comput. Mach.* **1970**, *17*, 589−602.
(53) Vardi, Y. Empirical Distributions In Selection Bias Models. *Ann. Statist.* **1985**, *13*, 178−203.
(54) Gallicchio, E.; Andrec, M.; Felts, A.; Levy, R. Temperature Weighted Histogram Analysis Method, Replica Exchange, and Transition Paths. *J. Phys. Chem. B* **2005**, *109*, 6722−6731.
(55) Ferrenberg, A.; Swendsen, R. Optimized Monte Carlo Data Analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.
(56) Holladay, J. Smoothest Curve Interpolation. *Math. Tables Aids Computation* **1957**, *11*, 233−243.
(57) Gill, R.; Vardi, Y.; Wellner, J. Large Sample Theory Of Empirical Distributions In Biased Sampling Models. *Ann. Statist.* **1988**, *16*, 1069−1112.
(58) Pollard, D. *Empirical Processes: Theory and Applications* in *NSF-CBMS Regional Conference Series in Probability and Statistics*, Vol. *2*; Institute of Mathematical Sciences: Hayward, California, 1990.
(59) Zhang, J. Powerful Goodness-Of-Fit Tests Based On the Likelihood Ratio. *J. R. Statist. Soc. B* **2002**, *64*, 281−294.
(60) Jager, L.; Wellner, J. Goodness-Of-Fit Tests Via Phi-Divergences. *Ann. Statist.* **2007**, *35*, 2018−2053.
(61) Stepanova, N.; Pavlenko, T. Goodness-Of-Fit Tests Based On Sup-Functionals Of Weighted Empirical Processes. 2014; http://arxiv.org/abs/1406.0526, preprint arXiv:1406.0526 [math.ST].
(62) Anderson, T.; Darling, D. Asymptotic Theory Of Certain "Goodness Of Fit" Criteria Based On Stochastic Processes. *Ann. Math. Stat.* **1952**, *23*, 193−212.
(63) Chicheportiche, R.; Bouchaud, J. P. Weighted Kolmogorov-Smirnov Test: Accounting For the Tails. *Phys. Rev. E* **2012**, *86*, 041115.
(64) Salvalaglio, M.; Tiwary, P.; Parrinello, M. Assessing the Reliability of the Dynamics Reconstructed from Metadynamics. *J. Chem. Theory Comput.* **2014**, *10*, 1420−1425.
(65) Efron, B. Bootstrap Methods: Another Look At the Jackknife. *Annals of Statistics* **1979**, *7*, 1−26.
(66) Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*; Chapman and Hall: New York, 1993.
(67) Wilks, S. The Large-Sample Distribution Of the Likelihood Ratio For Testing Composite Hypotheses. *Ann. Math. Stat.* **1938**, *9*, 60−62.
(68) Akaike, H. *Information Theory as an Extension Of the Maximum Likelihood Principle in Second International Symposium on Information Theory*; Akademiai Kiado: Budapest, Hungary, 1973.
(69) Akaike, H. A New Look At the Statistical Model Identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716−723.
(70) Akaike, H. Likelihood Of a Model and Information Criteria. *J. Econometrics* **1981**, *16*, 3−14.
(71) Lee, T.; York, D. vFEP software package available at http://theory.rutgers.edu/Group/vFep.shtml. 2013; http://theory.rutgers.edu/Group/vFep.shtml.
(72) Bickel, P.; Breiman, L. Sums of Functions of Nearest Neighbor Distances, Moment Bounds, Limit Theorems And a Goodness Of Fit Test. *Ann. Probability* **1983**, *11*, 185−214.
(73) Schilling, M. Goodness Of Fit Testing in R^m Based on the Weighted Empirical Distribution of Certain Nearest Neighbor Statistics. *Ann. Statist.* **1983**, *11*, 1−12.

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(74) Schilling, M. An Infinite-Dimensional Approximation for Nearest Neighbor Goodness of Fit Tests. *Ann. Statist.* **1983**, *11*, 13−24.
(75) Justel, A.; Peña, D.; Zamar, R. A Multi-Variate Kolmogorov-Smirnov Test of Goodness of Fit. *Statistics and Probability Letters* **1997**, *35*, 251−259.
(76) Akima, H. A Method of Bivariate Interpolation and Smooth Surface Fitting For Values Given At Irregularly Distributed Points. *ACM Trans. Mathematical Software* **1978**, *4*, 148−159.

DOI: 10.1021/acs.jpcb.7b04871
J. Phys. Chem. B XXXX, XXX, XXX−XXX