



Published in final edited form as:

*J Chem Theory Comput.* 2014 January 14; 10(1): 24–34. doi:10.1021/ct400691f.

## Roadmaps through free energy landscapes calculated using the multi-dimensional vFEP approach

Tai-Sung Lee<sup>†</sup>, Brian K. Radak<sup>†,‡</sup>, Ming Huang<sup>†,¶</sup>, Kin-Yiu Wong<sup>§</sup>, and Darrin M. York<sup>†</sup>

Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA, Department of Chemistry, University of Minnesota, Minneapolis, MN 55455, USA, Scientific Computation Program, University of Minnesota, Minneapolis, MN 55455, USA, and Department of Physics, High Performance Cluster Computing Centre, and Institute of Computational and Theoretical Studies, Hong Kong Baptist University, Kowloon Tong, Hong Kong

Tai-Sung Lee: taisung@biomaps.rutgers.edu; Darrin M. York: york@biomaps.rutgers.edu

### Abstract

The variational free energy profile (vFEP) method is extended to two dimensions and tested with molecular simulation applications. The proposed 2D-vFEP approach effectively addresses the two major obstacles to constructing free energy profiles from simulation data using traditional methods: the need for overlap in the re-weighting procedure and the problem of data representation. This is especially evident as these problems are shown to be more severe in two dimensions. The vFEP method is demonstrated to be highly robust and able to provide stable, analytic free energy profiles with only a paucity of sampled data. The analytic profiles can be analyzed with conventional search methods to easily identify stationary points (*e.g.* minima and first-order saddle points) as well as the pathways that connect these points. These “roadmaps” through the free energy surface are useful not only as a post-processing tool to characterize mechanisms, but can also serve as a basis from which to direct more focused “on-the-fly” sampling or adaptive force biasing. Test cases demonstrate that 2D-vFEP outperforms other methods in terms of the amount and sparsity of the data needed to construct stable, converged analytic free energy profiles. In a classic test case, the two dimensional free energy profile of the backbone torsion angles of alanine dipeptide, 2D-vFEP needs less than 1% of the original data set to reach a sampling accuracy of 0.5 kcal/mol in free energy shifts between windows. A new software tool for performing one and two dimensional vFEP calculations is herein described and made publicly available.

### Introduction

Free energy is a key concept in modern physical science and offers a wealth of insights into complex molecular problems.<sup>1</sup> One dimensional free energy profiles are routinely employed to study various molecular systems with respect to a certain variable/coordinate and many

Correspondence to: Tai-Sung Lee, taisung@biomaps.rutgers.edu; Darrin M. York, york@biomaps.rutgers.edu.

<sup>†</sup>Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA

<sup>‡</sup>Department of Chemistry, University of Minnesota, Minneapolis, MN 55455, USA

<sup>¶</sup>Scientific Computation Program, University of Minnesota, Minneapolis, MN 55455, USA

<sup>§</sup>Department of Physics, High Performance Cluster Computing Centre, and Institute of Computational and Theoretical Studies, Hong Kong Baptist University, Kowloon Tong, Hong Kong

#### Software Availability

The current vFEP program is freely available at <http://theory.rutgers.edu/vFEP> and <http://sites.google.com/site/cancersimulation/software>

enhanced sampling methods for accurately accomplishing this task have been developed in the past decades.<sup>2</sup> Some of the most widespread include multistage/stratified sampling,<sup>3</sup> statically<sup>4-6</sup> and adaptively<sup>7-9</sup> biased sampling, self-guided dynamics,<sup>10</sup> constrained dynamics,<sup>11,12</sup> as well as multicanonical<sup>13,14</sup> and replica exchange<sup>15</sup> algorithms. In addition, a number of simulation protocols based on non-equilibrium sampling<sup>16-19</sup> have also been recently proposed as well as hybrid algorithms.<sup>20,21</sup>

One of the most widely used methods for determining free energy surfaces for chemical reactions, where often there are geometric coordinates that are known to be aligned with the overall reaction coordinate, is the “umbrella sampling” technique.<sup>5,22</sup> Combining stratification with equilibrium and statically biased sampling, umbrella sampling is particularly amenable to parallel execution, especially in high performance distributed environments,<sup>23-25</sup> as well as extension or combination with replica exchange<sup>26,27</sup> and alchemical simulation techniques.<sup>28</sup>

There are numerous well-developed and widely used methods for constructing one dimensional free energy profiles from umbrella sampling simulations, such as the weighted histogram analysis method,<sup>29,30</sup> umbrella integration (UI),<sup>31</sup> the multistate Bennett acceptance ratio method (MBAR,<sup>32,33</sup> which can be seen as a binless extension of WHAM<sup>34</sup>), and others.<sup>35-38</sup> However, publicly available implementations/programs for two dimensional cases are still very limited. Most notably, both WHAM<sup>39</sup> and MBAR<sup>33</sup> implementations have been made available, while two dimensional UI<sup>40</sup> and GAMUS<sup>37,41</sup> implementations have been reported but are not publicly available. Nevertheless, due to the two key difficulties in umbrella sampling methods, the problem of “data re-weighting” and of “data representation”, the cost and complexity of such calculations can still be quite prohibitive. Here we briefly review these two major problems.

### The need of overlap in data re-weighting

Re-weighting from one sampled distribution to another can, in principle, be solved exactly by the free energy perturbation/Zwanzig relation and the related expression for mechanical observables.<sup>5,42,43</sup> However, WHAM and MBAR based methods to construct free energy profiles were developed with the understanding that this is not true in practice, since the overlap between distributions must be high in order to obtain reliable results.<sup>29,33</sup> An alternative approach is to assume smoothness of the free energy profile between nearby windows. Kästner’s UI approach uses a Gaussian distribution to model the un-weighted probability density for each umbrella window (or, equivalently, quadratic functions for the free energy profile), from which the analytic derivatives are calculated and integrated in order to recover the global probability density.<sup>40,44,45</sup> Hence, no explicit re-weighting is necessary. This approach is equivalent to assuming continuous first derivatives of the free energy profile between windows. Instead of quadratic functions centered on data-based parameters, the 1D-vFEP method<sup>46</sup> utilizes cubic spline functions to model the free energy profile, equivalent to assuming continuous first and second derivatives of the free energy profile between windows. It has been demonstrated in one dimensional cases that UI and 1D-vFEP require less of a degree of overlap between windows compared to WHAM and a histogram-based MBAR estimator.<sup>46</sup>

### Data representation

In order to extract a distribution function from a set of data, it is often necessary to employ a certain type of representation of that function. This is commonly formulated as the density estimation problem.<sup>47,48</sup> Perhaps the simplest method of data representation is to use a histogram estimator of the probability density.<sup>28,29,49</sup> However, this approach is frequently not numerically stable, especially when the data is too sparse such that the width of

histogram bins cannot be made sufficiently small. A useful alternative approach can be to apply a more robust kernel density estimator, but this too will fail with extremely sparse data sets. A completely different type of approach is to fit the overall density distribution through a pre-defined model<sup>37,44,50</sup> by optimizing the model parameters according to a merit function. Maragakis, *et al.* suggested a maximum likelihood approach utilizing the Gaussian-Mixture Model on umbrella sampling (GAMUS) for the global probability density based on the re-weighted data.<sup>37,51</sup> Similarly, Basner and Jarzynski proposed a binless estimator based upon the optimal correction to an arbitrary reference distribution.<sup>52</sup> UI<sup>40,44,45</sup> uses Gaussian models for the un-weighted probability densities and has also recently been extended to higher order densities (*i.e.* skewed Gaussians).<sup>53</sup> It is well known that such parametric approaches lead to a significant reduction in the number of data points needed to obtain a converged result. However this often comes at the expense of increased bias depending on the inherent accuracy of the parameteric form. For example, the approximations/assumptions of UI require near-quadratic (or near quartic) behavior of the local free energy surface for individual windows and this has been demonstrated to be inaccurate in simulations with weak biasing potentials.<sup>46</sup> This problem may be reduced by imposing stronger harmonic biasing potentials but this often leads to lower overlap between windows and hence the same kind of failures associated with sparsely populated histogram estimators.<sup>22</sup> GAMUS has also been shown not to be ideal for quantitatively describing details of the free energy surface.<sup>37</sup>

The vFEP implementation<sup>46</sup> for one-dimensional free energy profiles was demonstrated to be able to effectively address the above difficulties and outperform other methods in terms of the amount and sparsity of the data needed to construct the overall free energy profiles. The vFEP method uses higher-order (beyond quadratic) locally-compact functions to accurately model the free energy profile within each window. At the same time, it exploits the presumed smoothness of the profile in the interstitial regions that connect nearby windows. Nevertheless, its advantage is not greatly impressive since, in most cases, significant computational resources are not required when constructing one dimensional free energy profiles, even with large numbers of windows and data points. On the other hand, many more windows are needed to construct free energy profiles in two dimensions than in one. Furthermore, for each window, many more data points are required to represent the local two dimensional density distribution. This problem can effectively limit the practical applications of two dimensional umbrella sampling simulations.

In the present work we report the two dimensional extension of the vFEP method. Test cases demonstrate that vFEP is able to produce converged two dimensional free energy profiles in a more robust fashion (fewer and more sparse data points) than several other methods in common use. In addition, the resulting analytic form of free energy profiles allow the facile identification of stationary points and pathways that can be used for analysis, focused sampling, or adaptive biasing.

## Theory

The theory of the maximum likelihood method utilized in vFEP has been already described in detail elsewhere.<sup>46</sup> Since the formulation is virtually identical in any number of dimensions, the key points are outlined here specifically for the two dimensional case. The maximum likelihood method, or maximum likelihood estimation (MLE),<sup>54,55</sup> is a procedure for finding the optimal set of model parameters that maximize the likelihood of a probability distribution function used to represent a given set of observed data.

Here we consider the two dimensional probability density,  $p(x, y)$ , for observing a molecular system near a particular point with a generalized coordinates  $(x, y)$ . This probability density is given by

$$p(x, y) = \frac{e^{-F(x, y)}}{\iint e^{-F(x', y')} dx' dy'}, \quad (1)$$

where  $F(x, y) \equiv F(x, y)/(k_B T)$  is the unitless scaled free energy profile,  $F(x, y)$  is the free energy profile,  $k_B$  is the Boltzmann constant and  $T$  is the absolute temperature. Considering a set of umbrella sampling simulations in which a biasing potential  $W^\alpha(x, y)$  is applied in the  $\alpha$ th window, the probability of finding the system near a certain coordinate value  $(x, y)$  is:

$$p^\alpha(x, y) = \frac{1}{Z^\alpha} \exp\{-[F(x, y) + W^\alpha(x, y)]\}$$

$$Z^\alpha \equiv \iint \exp\{-[F(x, y) + W^\alpha(x, y)]\} dx dy \quad (2)$$

Suppose that, for the simulation of the  $\alpha$ th window, there are  $N^\alpha$  points observed with coordinate values  $\{(x_i^\alpha, y_i^\alpha)\}$ . Since they are observed points, the probability of each point is equal with value  $1/N^\alpha$ . The likelihood of the whole system with an overall free energy profile  $F(x, y)$  can be expressed as the combination of the likelihood of individual windows as:<sup>46</sup>

$$\hat{\ell}(F) \equiv - \sum_{\alpha}^{\text{windows}} \left\{ \ln Z^\alpha + \frac{1}{N^\alpha} \sum_{i=1}^{N^\alpha} [F(x_i^\alpha, y_i^\alpha) + W^\alpha(x_i^\alpha, y_i^\alpha)] \right\} \quad (3)$$

Note that in the above equation the term  $W^\alpha(x_i^\alpha, y_i^\alpha)$  is constant with respect to  $F$  and does not need to be evaluated if the goal is to maximize  $\hat{\ell}(F)$ . It is also worth noting that the term  $-\ln Z^\alpha$  is equivalent to the relative free energies between windows in other re-weighting schemes. In the present vFEP approach, the “re-weighting” procedure is implicitly accomplished through the normalization against the global trial function  $F$ .

The search for the optimal trial  $F$  can be implemented in different ways. The present work utilizes two dimensional cubic spline functions to model  $F$  and searches for the optimal spline coefficients that maximize the likelihood function  $\hat{\ell}(F)$ . Using two dimensional cubic splines is equivalent to assuming that the free energy profile varies slower than a cubic polynomial between windows or that the first and second derivatives of the free energy profile are continuous between windows for both of the coordinates.

## Results

### Implementation

The vFEP program has been modified and extended to two dimensions. However, due to the constraints of spline functions in the AlgLib package (v3.7, <http://www.alglib.net>), only two dimensional regular cubic spline (bicubic spline) functions are employed. Consequently, in addition to the interpolation algorithms originally utilized for 1D-vFEP,<sup>56,57</sup> one and two dimensional regular cubic spline algorithms were added. The vFEP program was written completely in C++ and compiled and tested with Microsoft Visual C++ 2008 on Window 7 and CMake 2.8.10.2 with Gnu C/C++ 4.7.2 on Fedora 16 and 18.

The main computationally intensive work is the numerical evaluation of Equation (3), which consists of two parts: 1) the integration of the exponential of the trial free energy profile and the biasing potential for each window; 2) the calculation of the trial free energy profile on all sampling points. The Gauss-Hermite quadrature rules are utilized to perform the integration of Equation (3) with the exponential of the biasing potential, a Gaussian function, as the background weighting function. For each window, a set of Gauss-Hermite quadrature rules ( $N=48$ ) are set up for each dimension and those rules with positions outside the sampling data range are ignored, resulting, on average, in 15–20 rules remaining in each dimension for each window.

For large data sets, the evaluation of the trial free energy function on all sample points would be extremely time consuming if done directly in each optimization iteration. Since the free energy profile is modeled by a bicubic spline, in a rectangular region enclosed by four spline nodes the free energy profile can be described as products of third order polynomials of two dimensions. Hence, the second part of Equation (3), the summation of all function values over all sampling points, can be re-arranged and calculated as follows: In a region enclosed by four nearby spline nodes, summations of all 16 possible polynomial products over all sampling points can be pre-calculated when sampling data are read-in. The second term of Equation (3) can thus be calculated from the summation of all products of the spline coefficients and the pre-calculated terms from each regions. Since there is no need to do any calculations involving the sample data points except in the initial stage, the cost to evaluate Equation (3) is independent of the sample data size.

Four different methods are compared in analyzing free energy profiles: 1) the WHAM-2d approach, 2) the MBAR approach with a Gaussian kernel density estimator (gKDE), 3) the MBAR approach with a histogram estimator, and 4) the present 2D-vFEP approach. The results of WHAM were calculated by the wham-2d program from Grossfield<sup>39</sup> (v2.0.6, <http://membrane.urmc.rochester.edu/content/wham>). The results of MBAR were calculated using the pymbar library of Shirts and Chodera<sup>33</sup> (v2.0b, <http://simtk.org/home/pymbar>) with an in-house implementation of multi-dimensional kernel density estimation. Histogram calculations employed a fixed bin width and boundaries inferred from the data range, while gKDE calculations used bandwidths obtained with the multivariate form of the method of Silverman,<sup>58</sup> with the covariance matrices taken from each window. Where indicated, these bandwidths were uniformly scaled by a multiplicative constant. The MBAR library does not currently provide results when any histogram bins contain no data points. Due to this limitation, the MBAR approach with histogram analysis failed in all cases reported here, except in the case of alanine dipeptide with 576 windows, 10,000 data points per window, and 180 bins per dimension. In our implementation, this MBAR/histogram case accepted by the MBAR library requires more memory than any local workstations possess and cannot be performed. Hence no results from MBAR/histogram could be reported and all MBAR results in the subsequent sections refer to MBAR/gKDE.

All calculations were performed on a Linux (Fedora 16) workstation with an Intel i7 970 3.2GHz CPU (6-core) and 24 GB memory. WHAM-2d and 2D-vFEP did not utilize any multiple-core ability while the MBAR implementation utilized some multiple-thread functionality in the MBAR library. The results from WHAM-2d, MBAR/gKDE, and 2D-vFEP are reported in the following sections.

### Alanine Dipeptide

The classic test case of a two dimensional free energy profile, the  $(\phi, \psi)$  torsion free energy map of terminally blocked alanine peptide (sequence Ace-Ala-Nme) in explicit solvent, is treated as the benchmark system. The system is relatively simple and has been extensively studied.<sup>59–64</sup> A model Ace-Ala-Nme sequence peptide was generated by the tleap utility in

the Amber12 package<sup>65</sup> and subsequently put in a 30 Å cube of TIP3P<sup>66</sup> water (540 molecules). A biasing potential of 0.02 kcal/mol-deg.<sup>2</sup> was applied to both dimensions for each window. The NAMD package (v 2.9)<sup>67</sup> was used with the Amber FF99SB-IDLN force field<sup>68</sup> under periodic boundary conditions in the NpT ensemble at 300 K and 1 atm (NAMD uses a modified Nosé-Hoover method<sup>69,70</sup> in which Langevin dynamics is used to control fluctuations in the barostat). Each window was simulated for 0.5 ns equilibration and 1 ns of data collection with sampling frequency of 100 fs (10,000 data points per window).

The peptide backbone torsion angles,  $\phi$  and  $\psi$ , are defined as the relevant coordinates. Umbrella sampling simulations were performed by placing biasing potentials at evenly-spaced coordinates. Both  $\phi$  and  $\psi$  were scanned from  $-180^\circ$  to  $180^\circ$  with a step size of  $15^\circ$ , resulting in a set of  $24 \times 24$  (576) umbrella sampling windows. Sets of  $4 \times 4$  window data and  $8 \times 8$  window data were extracted from the original  $24 \times 24$  window data set using larger spacing ( $90^\circ$  spacing for the  $4 \times 4$  window set and  $45^\circ$  spacing for the  $8 \times 8$  window set).

**Variation in number of umbrella windows**—The ( $\phi$ ,  $\psi$ ) free energy profiles of alanine dipeptide from different numbers of windows ( $4 \times 4$  (16),  $8 \times 8$  (64), and  $24 \times 24$  (576)) are first calculated and compared. Figure 1 shows the results using the same data sets with different methods: WHAM (upper row), MBAR/gKDE (middle row), and the vFEP (bottom row). The figure suggests that WHAM and MBAR/gKDE need 576 windows to deliver reasonable results while vFEP can give a rough estimation of the free energy profile with only 16 windows. In addition, the fact that, with 16 or 64 windows, WHAM cannot converge in a large number of areas (the white space/spots) and MBAR/gKDE cannot produce qualitatively correct results, demonstrates the strong requirement of overlapping between windows in the traditional approaches such as WHAM and MBAR.

**Variation in the number of sampling points per window**—As previously mentioned, parametric approaches such as Gaussian kernel density estimation methods, compared to the histogram approaches, much reduce the number of data points required to construct the free energy profile. Figure 2 clearly shows such advantage. The original data sets (10,000 points per window) were stripped randomly to 100, 200, 1000 points per window. The results from those reduced data sets using WHAM are not able to deliver meaningful results with only 64 windows. Nevertheless better results are shown with more data points. On the other hand, although MBAR/gKDE cannot deliver good results with 64 windows, Figure 2 demonstrates the advantage of using smooth functions to represent sampling data distribution over the histogram approach utilized in WHAM: only 200 points per window are sufficient to deliver correct results compared to results from 10,000 points per window (Figure 1). 2D-vFEP not only is able to give good results with 64 windows but also only needs 200 points per window to reach more or less the same accuracy.

With large numbers of windows, WHAM and MBAR do not suffer from the overlap problem. Figure 3 shows the simulation results with 576 windows but reduced data sets similar to Figure 2. Results obtained from different numbers of data points per window are shown and suggest that, with 576 windows, MBAR/gKDE is able to deliver good results while WHAM still has difficulty to converge in every region even with 1,000 points per window. 2D-vFEP only needs 100 to 200 points per window to deliver qualitatively correct results.

**Identification of stationary points and pathways through the analytic free energy surface**—In the current vFEP implementation bicubic spline functions are used to represent the model free energy profile. Such a representation provides a simple and easy way to locate stationary points, since, given a certain value of one dimension, the zero gradient points of the other dimension can be calculated analytically along with the analytic

first and second gradients of both dimensions at the zero gradient points of the second dimension. Hence, by searching all zero gradient points of each dimension and calculating the analytic first and second gradients (*i.e.* the Hessian) of the other dimension, the extrema and saddle points can be located.

The minimum and saddle points identified by vFEP and Nudged Elastic Band<sup>71,72</sup> (NEB) paths are shown in the bottom rows of Figures 1 to 3. NEB calculations were performed with the DL-FIND library.<sup>73</sup> The results agree well and all of the NEB paths pass through minima and saddle points identified by vFEP, even though the initial points for the NEB calculations were not from vFEP results. In the cases with very few data points and windows, where WHAM and MBAR cannot deliver appropriate free energy profiles (Figures 1 and 2, left columns), vFEP is still able to identify reasonable minimum and saddle points, although more false positive points are found due to a higher level of noise.

### A model QM/MM phosphoryl transfer reaction

Using the AMBER12 simulation package,<sup>65</sup> hybrid quantum mechanical/molecular mechanical (QM/MM) umbrella sampling simulations were performed for 2-hydroxy ethyl ethyl phosphate (HEEP, see Figure 4), a model compound for RNA 2'-*O*-transesterification. The system was modeled in the same way as in a recent study of this system,<sup>74</sup> except that the periodic box was a 35.7 Å cube of TIP3P rigid water molecules<sup>66</sup> and the QM region Lennard-Jones parameters were taken from the AMBER FF10 force field.<sup>75</sup>

Two dimensional umbrella sampling was performed using the phosphorus to leaving group (bond breaking) distance ( $R_1$ ) and nucleophile to phosphorus (bond forming) distance ( $R_2$ ) (Figure 4). A total of 192 short simulations (400 ps with 50 ps equilibration) were performed with a harmonic biasing potential,

$$U_{\text{bias}}(R_1, R_2) = \frac{1}{2}k[(R_1 - R_{1,0})^2 + (R_2 - R_{2,0})^2] \quad (4)$$

where  $k = 100 \text{ kcal/mol-Å}^2$  and the values of the bias centers,  $R_{1,0}$  and  $R_{2,0}$ , spanned a grid of values at 0.25 Å intervals from 1.50 to 4.25 and 1.50 to 5.25, respectively.

The free energy profile results obtained from WHAM, MBAR/gKDE, and vFEP, are shown in Figure 5. Due to the relatively thin set of data, histogramming does not produce a complete surface, even when the number of bins along each axis is reduced considerably from 200 to 50. In the latter case the largest bin width is 0.075 Å, approaching a point where the results are dubiously quantitative. This problem can be somewhat mitigated by using a kernel density estimate to smooth out the regions between areas of low sampling. Using a simple, data-based choice of smoothing parameter in each window,<sup>58</sup> a complete surface can be generated (Figure 5, upper middle panel), but at the expense of spurious features in the reactant basin (top left of the surface). This problem is exaggerated further by scaling the smoothing parameters by a factor of two, a practice previously recommended in one dimensional cases in order to mitigate the appearance of spurious minima and maxima<sup>74</sup> (Figure 5, upper right panel). Unfortunately, decreasing the extent of smoothing leads to the same kinds of empty regions observed in the WHAM result (Figure 5, dark red spots in the upper left panel). Although one might reasonably expect that *some* combination of smoothing parameters will lead to a more balanced result, it is not at all clear what this combination will be or what kind of automatic and deterministic procedure could be followed to obtain it. Once again, however, vFEP provides a single, unique solution for the surface that maintains all of the features, both coarse and detailed, observed with the other methods.

As with alanine dipeptide above, the vFEP minimum and saddle points and the NEB paths were calculated (Figure 5, bottom right). The results again suggest that the vFEP minima and saddle points agree well with the NEB paths. The “extra” path found moving right from the reactant basin is especially interesting as it suggests the possibility of an alternative, albeit extremely high energy, pathway towards the products (note that the products here are two solvated anions and thus no proper product basin will be resolved without introducing periodic boundary artifacts). It is not clear whether or not this extra basin is an artifact of vFEP, although WHAM and MBAR indicate a similar region (Figure 5). Nonetheless, the automatic search for zero-gradient points made possible by vFEP provides a simple mechanism to broadly survey the free energy profile, at least making the user aware of the additional minima and thus able to assess it on the grounds of chemical intuition.

### QM/MM simulations of HHR

We have previously reported a set of large scale (921 windows) QM/MM umbrella sampling simulations to explore metal-assisted phosphoryl transfer and general acid catalysis in the extended hammerhead ribozyme (HHR).<sup>76</sup> Free energy profiles were constructed with two reaction coordinates:  $Z_1$ , the reaction coordinate of the phosphoryl transfer, and  $Z_4$ , the  $Mg^{2+}$  binding distance. For each window, data was collected every 1 fs for a total of 10 ps, resulting in 10,000 data points per window, in order to provide sufficient data points for the WHAM analysis.

Here, selected data from this set of simulations were used as a realistic test of the performance of the proposed 2D-vFEP method. WHAM, MBAR/gKDE, and vFEP were applied to either the full 921 windows or else 100 or 400 windows selected randomly. Tests were performed with 50, 500, and 5000 data points per window. Since WHAM cannot converge with fewer than 5000 data points per window and both MBAR and vFEP deliver similar results with different numbers of data points per window, only the case of 50 data points per window is shown here (Figure 6). WHAM and MBAR/gKDE need 400 windows to deliver reasonable results, but some regions are still poorly converged. Conversely, vFEP can give a rough, but smooth, estimate of the free energy profile with only 100 windows and 50 data points per window.

With both 400 and 921 windows, the vFEP minima and saddle points agree well with the NEB paths (Figure 6, bottom row). This is not as clearly true with 100 windows, where the points and path are qualitatively different. Nonetheless, the basin locations from the two calculations are in rough agreement and mostly qualitatively accurate in comparison to the 400 and 921 window cases.

Lastly, in the original work,<sup>76</sup> WHAM was employed and hence 921 windows along with a large number of data points in each window were necessary in order to construct a reasonable free energy profile useful for mechanistic inference. Use of vFEP not only improves the calculation with the original data set, but can get essentially identical results with less than half the windows, providing a significant reduction in the necessary computational resources.

### Sampling errors in free energy shifts

While the 2D-vFEP method involves no explicit re-weighting procedure, the  $-\ln Z^\alpha$  terms are equivalent to the relative free energy shifts,  $\Delta f^\alpha$ , defined in MBAR and WHAM. In vFEP they are obtained implicitly through global optimization of the free energy profile, while in the MBAR and WHAM approaches they are calculated as the results of the re-weighting procedure. Simple bootstrapping schemes<sup>77</sup> were utilized to estimate the statistical sampling



errors in free energy shifts. The errors were estimated by calculating the standard deviation between randomly chosen data sets with the same data size.

The calculated bootstrapping errors of the free energy shifts with different numbers of windows and data points are listed in Table 1. With  $8 \times 8$  windows and 200 data points, 2D-vFEP is able to deliver an accuracy of 0.5 kcal/mol when only 0.22% of the original data is used ( $24 \times 24$  windows; 10,000 points per window).

## Timing

To assess the practical usability of the current 2D-vFEP implementation, the time needed to perform 2D-vFEP of the alanine dipeptide system in a TIP3P water box was measured with different settings. Measurements were done with a set of  $24 \times 24$  (576)-window umbrella sampling simulation and a set of  $8 \times 8$  (64)-window umbrella sampling data created by strapping from the 576 window data set, the same procedure as mentioned in the previous sections. Results with WHAM and MBAR/gKDE were also measured. These programs were written in different languages (vFEP is written in C++, WHAM in C, MBAR in mixed C and Python), with different levels of optimization, and different implementations of underlining theories/algorithms; hence the timing results should not be compared directly. Rather, our purpose here is to demonstrate the usability of the 2D-vFEP program for realistic two dimensional free energy profile problems.

Table 2 lists the time needed to complete free energy profiles for alanine dipeptide system. The setting notations ( $8 \times 8$ ,  $24 \times 24$  windows; 100 pt/w, 200 pt/w, etc.) are the same as used in the previous sections. The MBAR run with 576 windows and  $10^4$  data points failed to allocate enough memory on a 24GB workstation. The results confirm that the computational cost of 2D-vFEP is more or less independent of the number of data points and is proportional to the number of windows.

## Discussion

There are various strategies to generate free energy profiles from umbrella sampling simulation data. WHAM and MBAR first model the free energy profiles of individual windows and then merge them into a global profile through re-weighting. Such an approach inevitably needs significant overlap between windows so that the combined free energy profile is sensible.

The approach taken by UI, GAMUS, and the present work is to assume smoothness of the free energy profile between windows and construct the global free energy profile by finding the most likely trial profile. However these approaches are different in the way they define the “best” global free energy profile. The original UI approach assumed that the free energy profile can be approximated as a combination of quadratic functions centered at umbrella sampling window centers<sup>40,44,45</sup> (this has now been extended to cubic and quartic functions<sup>53</sup>). GAMUS utilizes a Gaussian Mixture Model to model the density distribution with Gaussians with optimal centers and widths, while vFEP approximates the global free energy profile with local cubic spline functions. Such approaches do not strictly require significant overlap of data between nearby windows and hence are capable of modeling free energy profiles with many fewer umbrella simulation windows.

Certain data fitting procedures need to be performed when constructing free energy profiles from umbrella sampling data. It is well-known that using parametric analytic functions to represent the observed data leads to faster and stable convergence at the expense of potential bias. Hence one would expect more stable results if smooth functions are used to model the local density distributions or free energy profiles, instead of deriving from non-parametric

functions such as histograms. This is clearly demonstrated in the test cases in this work. However, there is no well-accepted way to choose suitable functions to represent the local density distributions or free energy profiles. Using quadratic functions for free energy profiles or Gaussians for density distributions requires a theoretically sound procedure to define the centers and widths of the corresponding Gaussians. Furthermore, these second order approximations may not be sufficiently accurate to describe the local windows well. Using third or higher order functions to model the local free energy profiles may encounter numerical instability before combining them into global free energy profiles, as well as increased bias in the parameters. On the other hand, in vFEP, the continuous and piece-wise third-order approximation of the global free energy profiles in vFEP achieves both reduction of data points needed and numerical stability. The fact that this mathematical assumption is also physically sound means that it should lead to a minimal bias. That is, it is greatly expected that the surface and its first two derivatives are piece-wise smooth and therefore it should not introduce considerable error to *impose* this from the start.

Kernel density smoothing is a popular way to reduce the noise in non-parametric density estimation.<sup>58</sup> However, the results from kernel density smoothing are critically dependent on the selection of the optimal smoothing bandwidth. While a complete discussion of this topic is beyond the scope of this paper, the QM/MM results shown in Figure 5 clearly demonstrate that kernel density smoothing methods need to be used with caution. Reasonable results may only be obtained when different bandwidths are used in different regions and there are, to our knowledge, no obvious methods for accomplishing this in a simple and general fashion.

The proposed vFEP approach, in both one and two dimensional cases, requires less overlap between windows compared to WHAM and MBAR and is more accurate compared to UI and GAMUS. Hence the vFEP approach successfully addresses the two major problems in constructing free energy profiles from umbrella sampling simulations and significantly reduces the number of data points needed. The following potential advantages of vFEP could significantly advance current free energy estimation techniques:

### Fast estimate of rough free energy profiles and biasing potentials

In recent years there has been much effort to enhance sampling in MD simulations to more efficiently construct free energy profiles and/or to extend the effective time scale of simulations.<sup>6-9,78-90</sup> The vFEP method may be integrated with these methods to potentially enhance their performance in following ways:

To adaptively and iteratively refine free energy profiles, often it is necessary to have a reasonable initial guess of free energy profile as the starting point. For two or higher dimensions, it is frequently prohibitively expensive to obtain even an initial rough guess of a free energy profile using methods such as WHAM, especially when the overall shape of the surface is unknown. In contrast, vFEP only requires very small amount and sparse data to deliver a semi-quantitative free energy profile that can serve as an initial rough guess to direct further simulations to collect data in the most important regions of the free energy landscape and refine the profile.

As vFEP only requires distribution data derived from simulations that may use any type of well-behaved biasing potential, or even without biasing potential (Equations (2) and (3)), it can also be utilized as a robust free energy profile estimator for distribution data from various types of MD simulations, with or without biasing potentials. Such a strategy would involve using the unbiased distribution data from accelerated MD simulations as input to vFEP in order to create an analytic free energy profile for the coordinates of interest. This

strategy could then, in principle, be integrated into subsequent MD simulations as an added biasing potential to further flatten the free energy landscape and enhance sampling.

### Paths, minimum points, and saddle points in 2D free energy profiles

When using any analytic form of a free energy profile, it is trivial to calculate first and second derivatives with respect to the relevant reaction coordinates. Although this advantage is not outstanding compared to other methods for the one dimensional case, the results for two dimensions demonstrate that vFEP could be significantly beneficial. To our knowledge, there is no other method capable of calculating the first and second derivatives analytically and easily. This advantage quickly leads to the efficient identification of the local minima and saddle points, as well as all zero gradient points, which can serve as candidates for possible reaction paths.

### Conclusion

Here we report the extension of the variational free energy profile (vFEP) method to two dimensions. vFEP utilizes the maximum likelihood principle and bicubic spline functions to construct global free energy profiles. Test cases show that the vFEP approach efficiently addresses the major obstacles to constructing free energy profiles encountered by traditional methods: the need for overlap in the re-weighting procedure and the problem of data representation. As a result, 2D-vFEP outperforms other methods in terms of the amount and sparsity of the data needed to construct the overall free energy profile. In the case of alanine dipeptide, 2D-vFEP only needs  $8 \times 8$  (64) windows and 200 data points per window, roughly 0.22% of the original data set (576 windows and 10,000 data points per window), to reach a sampling accuracy of 0.5 kcal/mol in free energy shifts between windows. Additionally, free energy stationary points, including minimum and saddle points, can be identified analytically and easily from the 2D-vFEP approach. Overall, the vFEP method is demonstrated to be highly robust, even with a paucity of sampled windows and data points, and offers a promising new tool for the construction and analysis of free energy profiles that can be used to interpret mechanisms and potentially as the foundation for new methods for adaptive biasing and enhanced sampling.

### Acknowledgments

The project is supported by the CDI type-II grant #1125332 fund by the National Science Foundation (to DY). The authors are grateful for financial support provided by the National Institutes of Health (GM62248 to DY). Computational resources from The Minnesota Supercomputing Institute for Advanced Computational Research (MSI) were utilized in this work. This research was partially supported in part by the National Science Foundation through TeraGrid resources provided by Ranger at TACC and Kraken at NICS under grant numbers TG-MCB100054 to TL and TG-MCB110101 to DY.

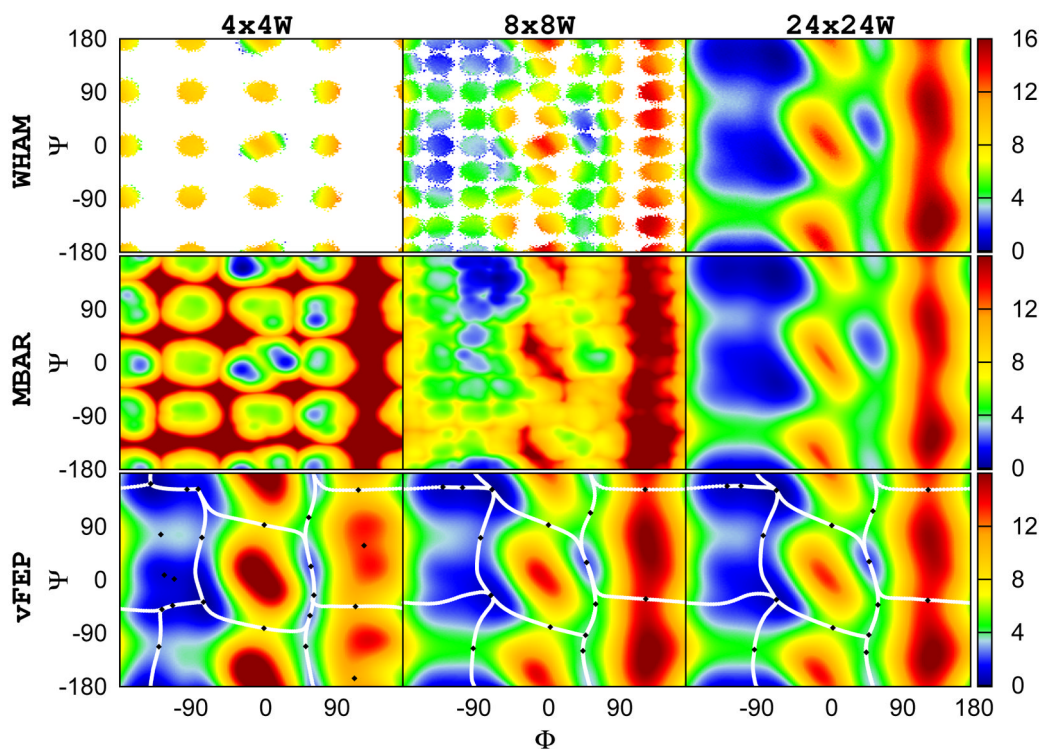
### References

1. Pohorille, A.; Chipot, C., editors. Springer Series in Chemical Physics. Springer, Berlin: 2007. Free Energy Calculations.
2. Zuckerman DM. *Annu Rev Biophys.* 2011; 40:41–62. [PubMed: 21370970]
3. Valleau JP, Card DN. *J Chem Phys.* 1972; 57:5457–5462.
4. Torrie GM, Valleau JP. *Chem Phys Lett.* 1974; 28:578–581.
5. Torrie GM, Valleau JP. *J Comput Phys.* 1977; 23:187–199.
6. Hamelberg D, Mongan J, McCammon JA. *J Chem Phys.* 2004; 120:11919–11929. [PubMed: 15268227]
7. Darve E, Rodríguez-Gómez D, Pohorille A. *J Chem Phys.* 2008; 128:144120. [PubMed: 18412436]
8. Laio A, Parrinello M. *Proc Natl Acad Sci USA.* 2002; 99:12562–12566. [PubMed: 12271136]

9. Babin V, Roland C, Sagui C. *J Chem Phys.* 2008; 128:134101. [PubMed: 18397047]
10. Wu X, Brooks BR. *Adv Chem Phys.* 2012; 150:255–326. [PubMed: 23913991]
11. den Otter WK. *J Chem Phys.* 2000; 112:7283–7292.
12. Darve E, Pohorille A. *J Chem Phys.* 2001; 115:9169–9183.
13. Berg BA, Neuhaus T. *Phys Rev Lett.* 1992; 68:9–12. [PubMed: 10045099]
14. Nakajima N, Nakamura H, Kidera A. *J Phys Chem B.* 1997; 101:817–824.
15. Sugita Y, Kitao A, Okamoto Y. *J Chem Phys.* 2000; 113:6042–6051.
16. Jarzynski C. *Phys Rev Lett.* 1997; 78:2690–2693.
17. Crooks GE. *J Stat Phys.* 1998; 90:1481–1487.
18. Hummer G, Szabo A. *Proc Natl Acad Sci USA.* 2001; 98:3658–3661. [PubMed: 11274384]
19. Minh DD, Chodera JD. *J Chem Phys.* 2009; 131:134110–134110. [PubMed: 19814546]
20. Nilmeier JE, Crooks GE, Minh DDL, Chodera JD. *Proc Natl Acad Sci USA.* 2011:108.
21. Ballard AJ, Jarzynski C. *J Chem Phys.* 2012; 136:194101. [PubMed: 22612074]
22. Kästner J. *WIREs Comput Mol Sci.* 2011; 1:932–942.
23. Luckow, A.; Lacinski, L.; Jha, S. The 10th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. ACM; 2010. p. 135-144. Chapter SAGA BigJob: An Extensible and Interoperable Pilot-Job Abstraction for Distributed Applications and Systems
24. Luckow, A.; Santcroos, M.; Weidner, O.; Merzky, A.; Maddineni, S.; Jha, S. Proceedings of the 21st International Symposium on High-Performance Parallel and Distributed Computing, HPDC'12; ACM; 2012. Chapter Towards a common model for pilot-jobs
25. Radak, BK.; Romanus, M.; Gallicchio, E.; Lee, T-S.; Weidner, O.; Deng, N-J.; He, P.; Dai, W.; York, DM.; Levy, RM.; Jha, S. Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery. XSEDE '13; 2013. p. 26:1-26:8.
26. Jiang W, Roux B. *J Chem Theory Comput.* 2010; 6:2559–2565. [PubMed: 21857813]
27. Gallicchio E, Levy RM. *Curr Opin Struct Biol.* 2011; 21:161–166. [PubMed: 21339062]
28. Souaille M, Roux B. *Comput Phys Commun.* 2001; 135:40–57.
29. Kumar S, Bouzida D, Swendsen R, Kollman P, Rosenberg J. *J Comput Chem.* 1992; 13:1011–1021.
30. Roux B. *Comput Phys Commun.* 1995; 91:275–282.
31. Kästner J, Thiel W. *J Chem Phys.* 2005; 123:144104. [PubMed: 16238371]
32. Shirts MR, Bair E, Hooker G, Pande VS. *Phys Rev Lett.* 2003; 91:140601. [PubMed: 14611511]
33. Shirts MR, Chodera JD. *J Chem Phys.* 2008; 129:124105. [PubMed: 19045004]
34. Tan Z, Gallicchio E, Lapelosa M, Levy RM. *J Chem Phys.* 2012; 136:144102. [PubMed: 22502496]
35. Sprik M, Ciccotti G. *J Chem Phys.* 1998; 109:7737–7744.
36. Bartels C. *Chem Phys Lett.* 2000; 331:446–454.
37. Maragakis P, van der Vaart A, Karplus M. *J Phys Chem B.* 2009; 113:4664–4673. [PubMed: 19284746]
38. Hansen N, Dolenc J, Knecht M, Riniker S, van Gunsteren WF. *J Comput Chem.* 2012; 33:640–651. [PubMed: 22228455]
39. Grossfield, A. WHAM: the weighted histogram analysis method, version 2.0.4. 2005.
40. Kästner J. *J Chem Phys.* 2009; 131:034109. [PubMed: 19624183]
41. Spiriti J, Binder JK, Levitus M, van der Vaart A. *Biophys J.* 2011; 100:1049–1057. [PubMed: 21320450]
42. Zwanzig RW. *J Chem Phys.* 1954; 22:1420–1426.
43. Allen, MP.; Tildesley, DJ. *Computer Simulation of Liquids.* Oxford Science Publications; New York: 1987.
44. Kästner J, Thiel W. *J Chem Phys.* 2005; 123:144104. [PubMed: 16238371]
45. Kästner J, Thiel W. *J Chem Phys.* 2006; 124:234106. [PubMed: 16821906]
46. Lee TS, Radak BK, Pabis A, York DM. *J Chem Theory Comput.* 2013; 9:153–164. [PubMed: 23457427]

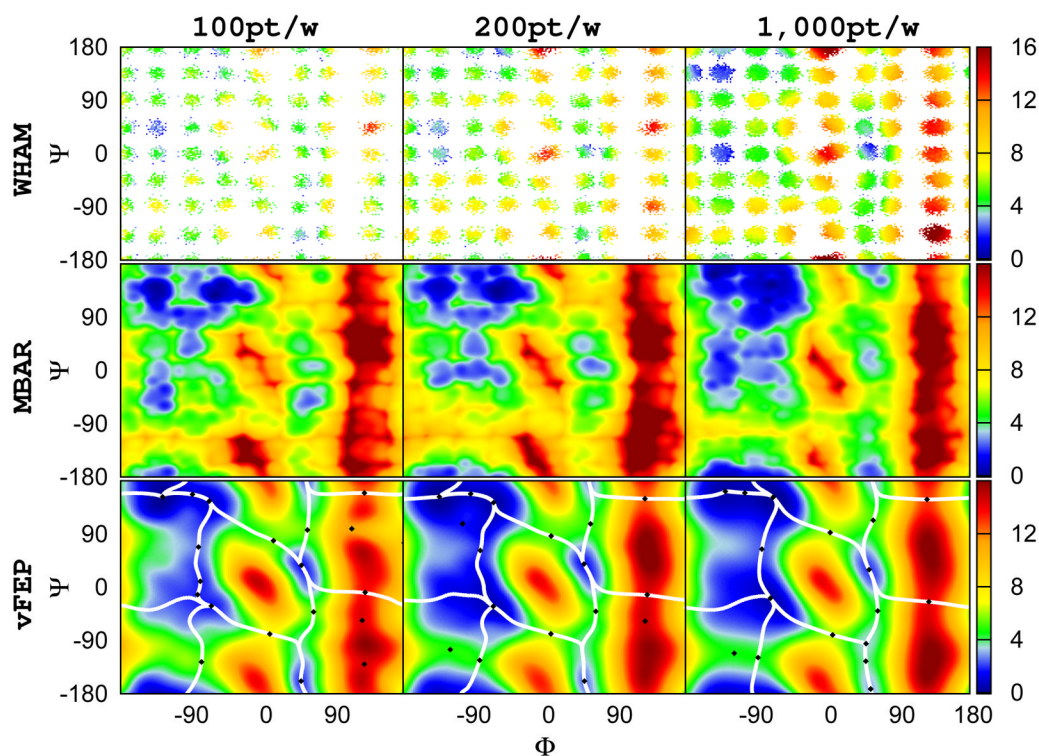
47. Silverman BW. *Ann Stat.* 1982; 10:795–810.
48. Sheather SJ. *Statist Sci.* 2004; 19:588–597.
49. Chodera JD, Swope WC, Pitera JW, Seok C, Dill KA. *J Chem Theory Comput.* 2007; 3:26–41.
50. Chakravorty DK, Kumarasiri M, Soudackov AV, Hammes-Schiffer S. *J Chem Theory Comput.* 2008; 4:1974–1980. [PubMed: 19319209]
51. Spiriti J, Kamberaj H, Van Der Vaart A. *Int J Quantum Chem.* 2012; 112:33–43.
52. Basner JE, Jarzynski C. *J Phys Chem B.* 2008; 112:12722–12729. [PubMed: 18793024]
53. Kästner J. *J Chem Phys.* 2012; 136:234102. [PubMed: 22779576]
54. Fisher RA. *Phil Trans R Soc Lond A.* 1922; 222:309–368.
55. Aldrich J. *Statist Sci.* 1997; 12:162–176.
56. Akima H. *J ACM.* 1970; 17:589–602.
57. Floater MS, Hormann K. *Numerische Mathematik.* 2007; 107:315–331.
58. Silverman, B. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall; New York: 1986.
59. Tobias DJ, Brooks CL. *J Phys Chem.* 1992; 96:3864–3870.
60. Apostolakis J, Ferrara P, Caflisch A. *J Chem Phys.* 1999; 110:2099–2108.
61. Rosso L, Abrams JB, Tuckerman ME. *J Phys Chem B.* 2005; 109:4162–4167. [PubMed: 16851477]
62. Ensing B, De Vivo M, Liu Z, Moore P, Klein ML. *Acc Chem Res.* 2006; 39:73–81. [PubMed: 16489726]
63. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC. *J Chem Phys.* 2007; 126:155101. [PubMed: 17461665]
64. Hermans J. *Proc Natl Acad Sci USA.* 2011; 108:3095–3096. [PubMed: 21310970]
65. Case, DA., et al. *AMBER 12.* University of California, San Francisco; San Francisco, CA: 2012.
66. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J Chem Phys.* 1983; 79:926–935.
67. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kaleč L, Schulten K. *J Comput Chem.* 2005; 26:1781–1802. [PubMed: 16222654]
68. Lindorff-Larsen K, Stefano P, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. *Proteins.* 2010; 78:1950–1958. [PubMed: 20408171]
69. Nosé S, Klein ML. *Mol Phys.* 1983; 50:1055–1076.
70. Hoover WG, Ree FH. *J Chem Phys.* 1967; 47:4873–4878.
71. Henkelman G, Uberuaga BP, Jónsson Hannes. *J Chem Phys.* 2000; 113:9901–9904.
72. Henkelman G, Jónsson H. *J Chem Phys.* 2000; 113:9978–9985.
73. Kästner J, Carr JM, Keal TW, Thiel W, Wander A, Sherwood P. *J Phys Chem A.* 2009; 113:11856–11865. [PubMed: 19639948]
74. Radak BK, Harris ME, York DM. *J Phys Chem B.* 2013; 117:94–103. [PubMed: 23214417]
75. Cornell WD, Cieplak P, Bayly CI, Gould IR, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. *J Am Chem Soc.* 1995; 117:5179–5197.
76. Wong KY, Lee TS, York DM. *J Chem Theory Comput.* 2011; 7:1–3. [PubMed: 21379373]
77. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap.* Chapman & Hall; New York: 1993.
78. Wu X, Wang S. *J Phys Chem B.* 1998; 102:7238–7250.
79. Lahiri A, Nilsson L, Laaksonen A. *J Chem Phys.* 2001; 114:5993–5999.
80. Iannuzzi M, Laio A, Parrinello M. *Phys Rev Lett.* 2003; 90:238302. [PubMed: 12857293]
81. Wang J, Gu Y, Liu H. *J Chem Phys.* 2006; 125:094907–094907. [PubMed: 16965119]
82. Babin V, Roland C, Darden TA, Sagui C. *J Chem Phys.* 2006; 125:204909–204909. [PubMed: 17144742]
83. Lelievre T, Rousset M, Stoltz G. *J Chem Phys.* 2007; 126:134111–134111. [PubMed: 17430020]
84. Li H, Min D, Liu Y, Yang W. *J Chem Phys.* 2007; 127:094101–094101. [PubMed: 17824726]
85. Zheng H, Zhang Y. *J Chem Phys.* 2008; 128:204106. [PubMed: 18513009]

86. Zheng L, Chen M, Yang W. Proc Natl Acad Sci USA. 2008; 105:20227–20232. [PubMed: 19075242]
87. Hansen HS, Hunenberger PH. J Comput Chem. 2010; 31:1–23. [PubMed: 19412904]
88. Min D, Chen M, Zheng L, Jin Y, Schwartz MA, Sang QXA, Yang W. J Phys Chem B. 2011; 115:3924–3935. [PubMed: 21417256]
89. Chodera JD, Shirts MR. J Chem Phys. 2011; 135:194110. [PubMed: 22112069]
90. Arrar M, de Oliveira CAF, Fajer M, Sinko W, McCammon JA. J Chem Theory Comput. 2013; 9:18–23. [PubMed: 23316122]



**Figure 1.**

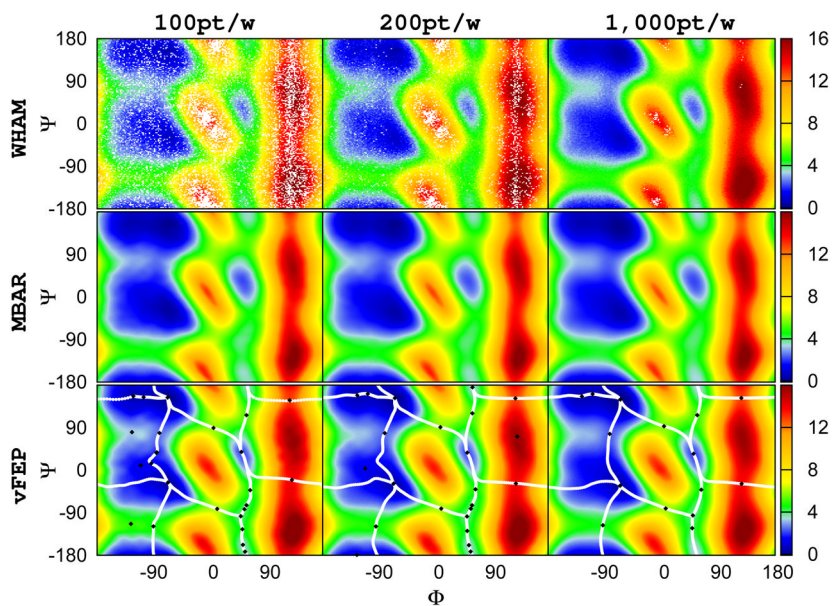
The 2D ( $\phi$ ,  $\psi$ ) free energy profile of alanine dipeptide calculated from different methods using the same data sets: WHAM, MBAR/gKDE, and vFEP results are shown in the upper, middle, and bottom rows respectively. Results obtained from  $4 \times 4$  (16),  $8 \times 8$  (64), and  $24 \times 24$  (576) windows are shown in the left, middle, and right columns respectively. All figures show results with 10,000 points per simulation window, except for the case of MBAR/gKDE with 576 windows where our current implementation only can handle 1,000 points per simulation window. White space in the WHAM results indicate empty histogram bins. WHAM and MBAR/gKDE need 576 windows to deliver reasonable results while vFEP can give a rough estimate of the profile with only 16 windows. All angles are in degrees. The color scheme of the free energy profile is shown at the right side color bar with units in kcal/mol. The black diamonds and the white lines shown in the vFEP contour maps are the saddle/minimum points and the possible paths found by the NEB method, respectively.



**Figure 2.**

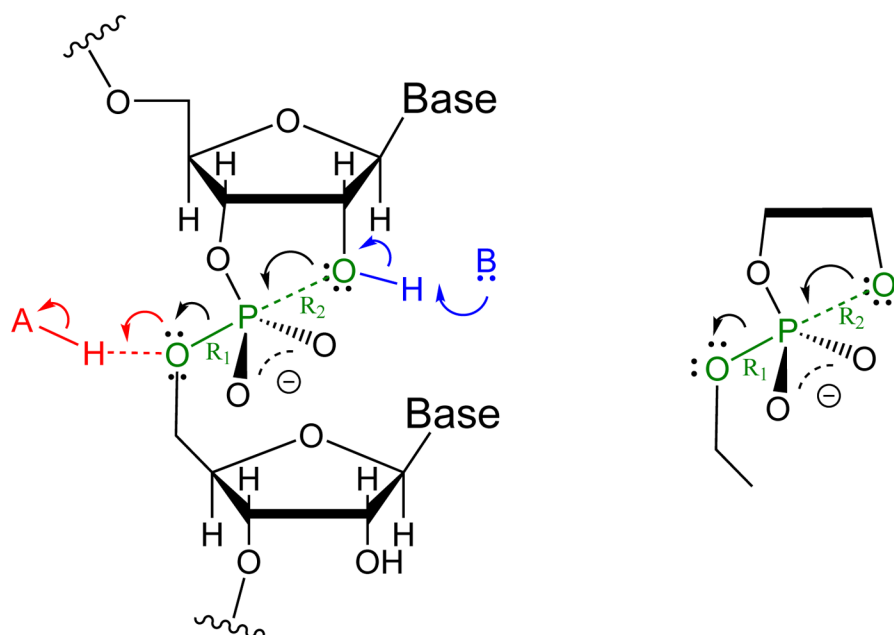
The 2D ( $\phi$ ,  $\psi$ ) free energy profile of alanine dipeptide calculated from different methods using  $8 \times 8$  (64) windows: WHAM, MBAR/gKDE, and vFEP results are shown in the upper, middle, and bottom rows respectively. Results obtained from different 100, 200, and 1000 points per window are shown in the left, middle, and right columns respectively. White space in the WHAM results indicate empty histogram bins. WHAM and MBAR/gKDE do not produce clearly meaningful results with any number of points, although the advantage of gKDE over histograms is evident. Conversely, the 2D-vFEP results with 200 and 1000 points are essentially the same. All angles are in degrees. The color scheme of the free energy profile is shown at the right side color bar with units in kcal/mol. The black diamonds and the white lines shown in the vFEP contour maps are the saddle/minimum points and the possible paths found by the NEB method, respectively.



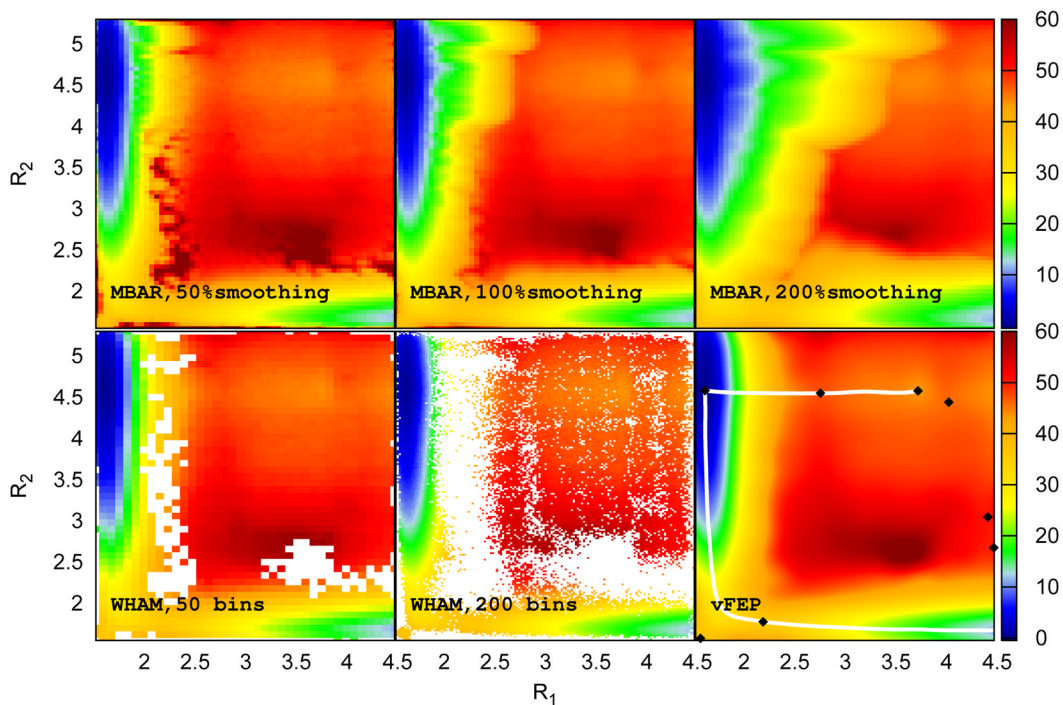


**Figure 3.**

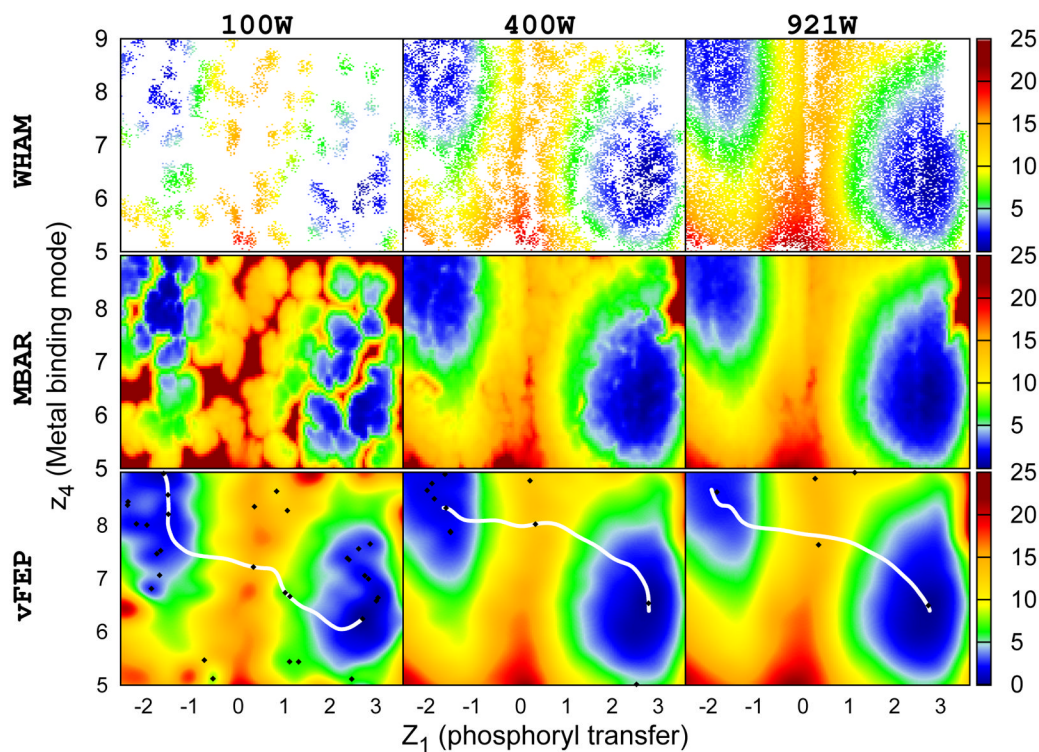
The 2D ( $\phi$ ,  $\psi$ ) free energy profile of alanine dipeptide calculated from different methods using  $24 \times 24$  (576) windows: WHAM, MBAR/gKDE, and vFEP results are shown in the upper, middle, and bottom rows respectively. Results obtained from different 100, 200, and 1000 points per window are shown in the left, middle, and right columns respectively. White space in the WHAM results indicate empty histogram bins. MBAR/gKDE is able to deliver good results while WHAM still cannot converge in all regions, even with 1000 points per window. The 2D-vFEP results with 200 and 1000 points are essentially the same, and the results with 100 points per window extremely similar in qualitative terms. All angles are in degrees. The color scheme of the free energy profile is shown at the right side color bar with units in kcal/mol. The black diamonds and the white lines shown in the vFEP contour maps are the saddle/minimum points and the possible paths found by the NEB method, respectively.



**Figure 4.** Left: Reaction scheme for RNA phosphoryl transfer reaction. Right: The HEPP model system used in the present work.



**Figure 5.** 2D QM/MM free energy profiles of the HEEP system (Figure 4) calculated from different methods. *Upper row:* MBAR results with kernel density smoothing schemes with different scaled Silverman bandwidth settings: *Left:* scaled by 0.5 (under-smoothed), *Middle:* no scaling (normal smoothing), and *Right:* scaled by 2.0 (over-smoothed). *Lower row:* WHAM results with 50 bins (left) and 200 bins along each axis (middle), and the vFEP result (right). Coordinate values are in Å and energies are in kcal/mol. The black diamonds and the white lines shown in the vFEP contour maps are the saddle/minimum points and the possible paths found by the NEB method, respectively.



**Figure 6.**

The 2D free energy profile of HHR ribozyme (see Wong et al. <sup>76</sup> for reaction coordinate definitions). WHAM, MBAR/gKDE, and vFEP results are shown in the upper, middle, and bottom rows respectively. Results obtained from 100, 400, and 921 windows are shown in the left, middle, and right columns respectively. For each window data was collected every 100 fs for a total of 5 ps, resulting in 50 data points per window. Calculations with 100 and 400 windows used randomly selections from the original 921 windows. White space in the WHAM results indicate empty histogram bins. WHAM gives noticeably sparser results compared to the original data set. Conversely, 2D-vFEP gives a rough, but smooth, estimation of the free energy profile with only 100 windows. The two coordinates are in Å. The color scheme of the free energy profile is shown at the right side color bar with units in kcal/mol. The black diamonds and the white lines shown in the vFEP contour maps are the saddle/minimum points and the possible paths found by the NEB method, respectively.

Estimated bootstrap errors (50 and 100 calculations with random data sets) of free energy shifts calculated by the vFEP method. The system is an alanine dipeptide in a TIP3P water box with 8×8 and 24×24 umbrella windows.

**Table 1**

	100pt/w			200pt/w			1,000pt/w		
	N=50	N=100	N=50	N=100	N=50	N=100	N=50	N=100	
8×8 windows	min	0.503	0.481	0.319	0.308	0.095	0.091		
	max	3.822	4.212	0.712	0.662	0.282	0.288		
	ave	0.822	0.819	0.453	0.446	0.194	0.192		
	100pt/w			200pt/w			1,000pt/w		
	N=50	N=100	N=50	N=100	N=50	N=100	N=50	N=100	
24×24 windows	min	0.082	0.088	0.052	0.059	0.021	0.021		
	max	0.185	0.188	0.119	0.114	0.058	0.053		
	ave	0.115	0.115	0.079	0.079	0.032	0.032		

The numbers here are derived from a set of 24×24 (576)-window umbrella sampling simulation on an alanine dipeptide in a TIP3P water box with a biasing potential of 0.02 kcal/mol-deg.<sup>2</sup> for each window, and the same results from a set of 8×8 (64)-window umbrella sampling data created by strapping from the 576 window data set (see Results). The results are estimated by performing bootstrapping type error analysis on the free energy shift term,  $-\ln Z^{\alpha}$ , with the same calculations performed on 50 or 100 randomly chosen data sets of the same size out from the original data set, e.g., for “100pt/w”, 100 random points were selected from the original 10,000 data points for each window. The bootstrapping standard deviation of each window among the data sets was calculated. The minimal (min) and maximal (max) bootstrapping standard deviation values for all windows, along with their averages (ave) are listed. All values are in units of kcal/mol.

Table 2

Results of timing of different methods on an alanine dipeptide in a TIP3P water box with 8×8 and 24×24 umbrella windows and different numbers of data points in each window (in seconds).

	100 pt/w	200 pt/w	1,000 pt/w	10,000 pt/w
WHAM	3,264.3	4,325.1	7,419.2	7,748.9
MBAR	10.5	25.0	631.1	42,832.2
vFEP	36.4	31.2	39.8	47.3
	100 pt/w	200 pt/w	1,000 pt/w	10,000 pt/w
WHAM	16,689.0	16,723.5	16,111.8	16,191.7
MBAR	157.1	249.0	4,682.3	failed
vFEP	378.2	396.4	507.1	674.4

The numbers here are derived from a set of 24×24 (576)-window umbrella sampling simulation on an alanine dipeptide in a TIP3P water box with a biasing potential of 0.02 kcal/mol/degree<sup>2</sup> for each window, and the same results from a set of 8×8 (64)-window umbrella sampling data created by strapping from the 576 window data set (see the Results section). The results were obtained from the UNIX command "time". vFEP is written in C++, WHAM in C, MBAR in mixed C and Python. *The numbers should be considered as for reference only, rather than performance tests on those methods since different implementations of WHAM and MBAR could lead to significantly different results.* The convergence criteria was set to 10<sup>-3</sup> for WHAM (the largest free energy shift change), 10<sup>-9</sup> for MBAR (the maximum relative change in the free energy shifts). Different size of data sets are strapped out from the original data set, e.g., for "100 pt/w", one data point was selected in every 100 points, resulting 100 points out of the original 10,000 data points for each window. The MBAR run with 576 windows and 10<sup>4</sup> data points failed to allocate enough memory. All timing results were performed on a Linux (Fedora 16) workstation with an Intel i7 970 3.2GHz CPU (6-core) and 24 GB memory. WHAM and vFEP did not utilize any multiple-core ability while MBAR utilized some multiple-thread functionality in the MBAR library. All values are in seconds.