



ELSEVIER

Computer Physics Communications 91 (1995) 275–282

Computer Physics  
Communications

# The calculation of the potential of mean force using computer simulations

Benoît Roux

*Université de Montréal, Physics Department, C.P. 6128, succ. Centre-Ville, Montréal, Canada H3C 3J7*

Received 10 October 1994

## Abstract

The problem of unbiasing and combining the results of umbrella sampling calculations is reviewed. The weighted histogram analysis method (WHAM) of S. Kumar et al. (J. Comp. Chem. 13 (1992) 1011) is described and compared with other approaches. The method is illustrated with molecular dynamics simulations of the alanine dipeptide for one- and two-dimensional free energy surfaces. The results show that the WHAM approach simplifies considerably the task of recombining the various windows in complex systems.

*Keywords:* Computer simulations; Umbrella sampling; Free energy surface; Potential of mean force; Alanine dipeptide

## 1. Introduction

The potential of mean force (PMF)  $\mathcal{W}(\xi)$  along some coordinate  $\xi$ , first introduced by Kirkwood in 1935 [1], is a key concept in modern statistical mechanical theories of liquids and of complex molecular systems. It is defined from the average distribution function  $\langle \rho(\xi) \rangle$ ,

$$\mathcal{W}(\xi) = \mathcal{W}(\xi^*) - k_B T \ln \left[ \frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi^*) \rangle} \right], \quad (1)$$

where  $\xi^*$  and  $\mathcal{W}(\xi^*)$  are arbitrary constants. The average distribution function along the coordinate  $\xi$  is obtained from a Boltzmann weighted average,

$$\langle \rho(\xi) \rangle = \frac{\int d\mathbf{R} \delta(\xi'[\mathbf{R}] - \xi) e^{-U(\mathbf{R})/k_B T}}{\int d\mathbf{R} e^{-U(\mathbf{R})/k_B T}}, \quad (2)$$

where  $U(\mathbf{R})$  represents the total energy of the system as a function of the coordinates  $\mathbf{R}$  and  $\xi'[\mathbf{R}]$  is a function depending on a few or several degrees of

freedom in the dynamical system (e.g.,  $\xi'[\mathbf{R}]$  may be an angle, a distance, or a more complicated function of the Cartesian coordinates of the system). In particular, conformational equilibrium properties or the transition rate of dynamical activated processes can be expressed conveniently in terms of the function  $\mathcal{W}(\xi)$  [2]. For these reasons, the PMF is a central quantity in computational studies of macromolecular systems.

It is often impractical to compute the PMF  $\mathcal{W}(\xi)$  or the distribution function  $\langle \rho(\xi) \rangle$  directly from a straight molecular dynamics simulation. For example, the presence of large energy barriers along  $\xi$  may prevent an accurate sampling of the configurational space within the available computer time. To avoid such difficulties, special sampling techniques have been designed to calculate the PMF from a molecular dynamics trajectory effectively. One of these approaches is the umbrella sampling technique of Torrie and Valleau [3,4]. In this method, the microscopic system of interest is simulated in the presence of an artificial bi-

asing window potential,  $w(\xi)$ , introduced to enhance the sampling in the neighborhood of a chosen value  $\xi$ . The biased simulations are generated using the potential energy  $[U(\mathbf{R}) + w(\xi)]$ . Typically, the biasing potential serves to confine the variations of the coordinate  $\xi$  within a small interval around some prescribed value, helping to achieve a more efficient configurational sampling in this region (this is the reason why the biasing potential is called a window potential). For example, a reasonable choice to produce the biased ensembles, though not the unique one, is to use harmonic functions of the form  $w_i(\xi) = \frac{1}{2}K(\xi - \xi_i)^2$ , centered on successive values of  $\xi_i$ . Because the sampling is confined to a small region during a given biased simulation, only a small piece of the estimated PMF is sufficiently accurate to be useful. To obtain the PMF over the whole range of interest of  $\xi$ , it is necessary to perform a number of biased window simulations, each biasing the configurational sampling around a different region of  $\xi$ . Ultimately, the results of the various windows are unbiased and then recombined together to obtain the final estimate  $\mathcal{W}(\xi)$ .

These last steps are the most important in an umbrella sampling calculation. According to Eq. (2), the biased distribution function obtained from the  $i$ th biased ensemble is

$$\langle \rho(\xi) \rangle_{(i)} = e^{-w_i(\xi)/k_B T} \langle \rho(\xi) \rangle \langle e^{-w_i(\xi)/k_B T} \rangle^{-1}. \quad (3)$$

The unbiased PMF from the  $i$ th window is

$$\mathcal{W}_i(\xi) = \mathcal{W}(\xi^*) - k_B T \ln \left[ \frac{\langle \rho(\xi) \rangle_{(i)}}{\langle \rho(\xi^*) \rangle} \right] - w_i(\xi) + F_i, \quad (4)$$

where the undetermined constant  $F_i$  defined from

$$e^{-F_i/k_B T} = \langle e^{-w_i(\xi)/k_B T} \rangle, \quad (5)$$

represents the free energy associated with introducing the window potential. There have been numerous efforts at addressing the problems of unbiasing and recombining the information from umbrella sampling calculations. Traditionally, the unknown free energy constants  $F_i$  are obtained by adjusting the various  $\mathcal{W}_i(\xi)$  of adjacent windows in the region in which they overlap until they match [3,4]; the matching can be done manually or automatically, using a least-squares procedure [5]. After the free energy constants have

been obtained, the PMF for the whole range of interest is generated by connecting the various  $\mathcal{W}_i(\xi)$  together and discarding the superfluous data in the region in which they overlap [3,4]. Although this is a valid procedure to construct the PMF, it is very limited in practical applications. For example, a significant overlap between the adjacent windows is necessary to overcome the statistical errors in each individual estimates. Thus a large quantity of the simulation data is not used. Furthermore, because the process of matching the adjacent windows is somewhat arbitrary, the uncertainty involved in the process results in a global error that grows with the number of windows. Additional problems arise in the case of free energy surfaces in a space of higher dimensionality [6], e.g., the value of a free energy constant  $F_i$  allowing a best match along one coordinate may differ from that allowing a best matching along other coordinates due to statistical fluctuations.

Different approaches have been proposed to construct a PMF function valid over the range of interest and avoid discarding any information in combining the available data from the umbrella sampling simulations. To avoid adjusting the unknown free energy constants  $F_i$  on the basis of a matching of the adjacent windows, Haydock et al. [7] proposed to calculate the relative value of the free energy constants  $F_i$  using free energy perturbation [8]. The central idea of the approach is to compute the free energy difference  $\Delta F_{i,j} = [F_i - F_j]$  between the adjacent windows. Since the absolute value of the constants  $F_i$  is not required, only the relative value of each of the constants  $F_i$  needs to be determined according to the normalization of the distribution function. One arbitrary constant,  $F_1$ , is set from  $\xi^*$  and  $\mathcal{W}(\xi^*)$  in Eq. (1). Woolf and Roux obtained the constants  $F_i$  with free energy perturbation and expressed the best estimate of the PMF as a weighted sum over the unbiased PMF's of the windows [9]. In a calculation of the three-dimensional free energy surface of superoxide inside the active site of superoxide dismutase, Shen and McCammon used a least-squares fitting procedure to obtain the free energy constants  $F_i$  and expressed the best estimate of the distribution function as a weighted sum over the unbiased distributions of the individual windows [5]. Nevertheless, most approaches require a significant amount of overlap between adjacent windows and may not be optimal computationally.

The weighted histogram analysis method (WHAM) proposed by Kumar et al. [10] aims at using all the information present in the umbrella sampling simulations and avoids the problems mentioned above. One of the main advantages of WHAM is that it can be easily extended to treat the case of a PMF depending on more than one variable [10,11]. The approach represents a generalization of the histogram method developed by Ferrenberg and Swendsen [12]. The central idea, which goes back to the maximum overlap method developed by Bennet to estimate free energy differences [13], consists in constructing an optimal estimate of the unbiased distribution function as a weighted sum over the data extracted from all the simulations and determining the functional form of the weight factors that minimizes the statistical error. The WHAM approach is now routinely used to calculate the PMF along a single coordinate [10,11,14]. However, no applications of the approach to the case of a multidimensional free energy surface could be found in the literature when this paper was written despite the fact that this represents a straightforward extension of the methodology [11].

In this paper, we describe and compare different methods to unbiased and recombine the results from umbrella sampling calculations. Three different weighting methods, the weighted histogram analysis method (WHAM) [10], a weighted distribution function (WDF) [5], a weighted potential of mean force (WPMF) [9,15] as well as an approach based on free energy perturbation (FEP) [7], are considered. The alanine dipeptide was chosen for a model system and PMF's depending on one and two coordinates were examined.

## 2. Theory and method

We consider an umbrella sampling calculation involving  $N_w$  biased window simulations. The WHAM equations express the optimal estimate for the unbiased distribution function as a  $\xi$ -dependent weighted sum over the  $N_w$  individual unbiased distribution functions [ $\langle \rho(\xi) \rangle_{(i)}^{\text{unbiased}}$ ] [10,12],

$$\langle \rho(\xi) \rangle = \sum_{i=1}^{N_w} [\langle \rho(\xi) \rangle_{(i)}]^{\text{unbiased}}$$

$$\times \left[ \frac{n_i e^{-[w_i(\xi) - F_i]/k_B T}}{\sum_{j=1}^{N_w} n_j e^{-[w_j(\xi) - F_j]/k_B T}} \right], \quad (6)$$

where  $n_i$  is the number of independent data points used to construct the biased distribution function. Based on Eqs. (3) and (5), the individual unbiased distribution function is

$$[\langle \rho(\xi) \rangle_{(i)}]^{\text{unbiased}} = e^{+w_i(\xi)/k_B T} \langle \rho(\xi) \rangle_{(i)} e^{-F_i/k_B T}, \quad (7)$$

and Eq. (6) can be re-written in the form [10,11]

$$\langle \rho(\xi) \rangle = \sum_{i=1}^{N_w} n_i \langle \rho(\xi) \rangle_{(i)} \times \left[ \sum_{j=1}^{N_w} n_j e^{-[w_j(\xi) - F_j]/k_B T} \right]^{-1}. \quad (8)$$

The free energy constants  $F_i$ , needed in Eq. (8), are determined from Eq. (5) using the optimal estimate for the distribution function,

$$e^{-F_i/k_B T} = \int d\xi e^{-w_i(\xi)/k_B T} \langle \rho(\xi) \rangle. \quad (9)$$

Because the distribution function itself depends on the set of constants  $\{F_j\}$ , the WHAM equations (8) and (9) must be solved self-consistently. In practice this is achieved through an iteration procedure. Starting from an initial guess for the  $N_w$  free energy constants  $F_i$ , an estimate for the unbiased distribution is obtained from Eq. (8). This estimate for  $\langle \rho(\xi) \rangle$  is used in Eq. (9) to generate new estimates for the  $N_w$  free energies constants  $F_i$  and a new unbiased distribution is generated with Eq. (8). The iteration cycle is repeated until both equations are satisfied.

Extensions of the method to multidimensional cases is straightforward. For example, in the case of two variables  $\xi_1$  and  $\xi_2$ , Eqs. (6) and (9) become

$$\langle \rho(\xi_1, \xi_2) \rangle = \sum_{i=1}^{N_w} n_i \langle \rho(\xi_1, \xi_2) \rangle_{(i)} \times \left[ \sum_{j=1}^{N_w} n_j e^{-[w_j(\xi_1, \xi_2) - F_j]/k_B T} \right]^{-1}, \quad (10)$$

and

$$e^{-F_i/k_B T} = \int d\xi_1 \int d\xi_2 e^{-w_j(\xi_1, \xi_2)/k_B T} \langle \rho(\xi_1, \xi_2) \rangle. \quad (11)$$

As in the one-dimensional case, the WHAM equations (10) and (11) are solved self-consistently.

For a comparison with the PMF calculated with WHAM, let us consider the free energy perturbation method (FEP) of Haydock et al. [7]. The central idea of the method is to calculate the free energy difference  $\Delta F_{i,i-1}$  between the window potentials  $w_i$  and  $w_{i-1}$ ,

$$\begin{aligned} \Delta F_{i,i-1} &= F_i - F_{i-1} \\ &= -k_B T \ln \left[ \langle e^{-[w_i(\xi) - w_{i-1}(\xi)]/k_B T} \rangle_{(i-1)} \right]. \end{aligned} \quad (12)$$

An equivalent expression can be written for the backward free energy  $F_{i-1,i}$ . In one dimension, the constants  $F_i$  can be obtained by recurrence, i.e.,

$$F_i = F_1 + \Delta F_{2,1} + \dots + \Delta F_{i,i-1}. \quad (13)$$

Extension of the FEP method to the case of higher dimensions is relatively straightforward, although it requires special care. The relative free energy difference between the adjacent windows can be calculated in two-dimensions using Eq. (12). However, it can be expected that the cumulative free energy resulting from the free energy differences along any closed path is not identically zero due to statistical fluctuations in the sampling; e.g.,  $\sum^{\text{closed path}} \Delta F_{i,j} \neq 0$ . Thus, the calculated free energy differences  $\Delta F_{i,j}$  may not be consistent with a unique set of constant  $F_i$ . In the present application, a least-squares procedure was used to determine a unique set of constants  $F_i$  from all the relative free energy differences  $\Delta F_{i,j}$  between the adjacent windows. Once the  $F_i$  are determined, the PMF's from the individual windows,  $\mathcal{W}_i$ , can be generated with the proper offset constants and no supplementary matching is required.

Different approaches have been proposed to combine the available data avoiding the discarding of any information. To construct the PMF over the whole range of interest, Shen and McCammon used a weighted sum over the individual unbiased distribution functions extracted the  $N_w$  windows [5],

$$\langle \rho(\xi) \rangle = \sum_{i=1}^{N_w} [ \langle \rho(\xi) \rangle ]_{(i)}^{\text{unbiased}}$$

$$\times \left[ \frac{n_i \langle \rho(\xi) \rangle_{(i)}}{\sum_{j=1}^{N_w} n_j \langle \rho(\xi) \rangle_{(j)}} \right], \quad (14)$$

with a straightforward extension to cases of higher dimension. In the following, the weighted distribution function procedure is referred to as W-DF. The  $\xi$ -dependent weighted average expressed by Eq. (14) is not unique. For example, Woolf and Roux used a weighted sum over the individual unbiased PMF's  $\mathcal{W}_i(\xi)$  [9,15],

$$\mathcal{W}(\xi) = \sum_{i=1}^{N_w} \mathcal{W}_i(\xi) \left[ \frac{n_i \langle \rho(\xi) \rangle_{(i)}}{\sum_{j=1}^{N_w} n_j \langle \rho(\xi) \rangle_{(j)}} \right], \quad (15)$$

with a similar expression for the two-dimensional case. In the following, the weighted PMF procedure is referred to as W-PMF. Both W-DF or W-PMF can be used to recombine the results from FEP.

In fact, the W-DF and W-PMF  $\xi$ -dependent weighted sum are very similar to the WHAM equation (6). The three expressions, Eqs. (6), (14) and (15), serve to recombine all available data into a smooth function valid over the whole range of  $\xi$ . Eq. (6) gives more weight to the  $i$ th estimate where the Boltzmann factor of the  $i$ th window potential is large whereas Eqs. (15) and (14) give more weight to the estimate of the  $i$ th window where the biased distribution  $\langle \rho(\xi) \rangle_{(i)}$  is statistically more important. This similarity with the WHAM formulation suggests that Eqs. (15) (or Eq. (14)) and (9) may be solved self-consistently providing a different estimate of the free energy constants  $F_i$ . To examine the validity of the weighted sums equations (15) and (14) the self-consistent W-PMF and W-DF approaches were also tested.

### 3. Illustration of the methods

The alanine dipeptide,  $\text{CH}_3\text{-CO-Ala-NH-CH}_3$ , was chosen for the purpose of illustrating the different methods. This system is convenient because it represents a prototypical model of the protein backbone and has been studied extensively by molecular dynamics. In the present work, one and two-dimensional PMF's characterizing the  $\text{C}_{1-7}$  intramolecular  $\text{N-H} \cdots \text{O}=\text{C}$  hydrogen bond of the alanine dipeptide were considered. The PMF's were calculated in vacuum (one-

dimensional) and in the presence of one isolated water molecule (two-dimensional) using the FEP and the self-consistent WHAM, W-PMF and W-DF methods. All the calculations were done including all hydrogen atoms with the CHARMM PARM22 potential function [16] for the peptides and TIP3P [17] for the water.

To first test the methods in a one-dimensional case, the PMF  $\mathcal{W}(r)$  along the N-H...O=C distance  $r$  of the alanine dipeptide was calculated in vacuum. The potential energy of the system was biased with a harmonic potential,  $\frac{1}{2}K(r - r_i)^2$ , centered on successive values of  $r_i$ , where  $K$  is the harmonic force constant. The conformational sampling was performed with Langevin dynamics; trajectories of 100 ps were generated for five windows, centered on 1.5, 2.5, ..., 5.5 Å with a harmonic force constant  $K$  of 5 kcal/mol/Å<sup>2</sup>. A time step of 0.002 ps was used for all simulations and a friction constant of 25 ps<sup>-1</sup> was applied to all non-hydrogen atoms. Each window was equilibrated during 10 ps starting from the last configuration of the previous window. The self-consistent set of equations (WHAM, W-PMF and W-DF) were iterated until changes in the free energy constants  $F_i$  were less than 0.001 kcal/mol. The free energy differences in the FEP method were calculated by averaging the backward and forward perturbations and the resulting  $\mathcal{W}_i(r)$  were combined together using Eq. (15). To provide a reference for comparison with the PMF calculated with the umbrella sampling method, a single unbiased simulation of 2 ns was generated. In the remainder of the paper, the PMF calculated from the 2 ns simulation is referred to as exact.

To illustrate the methods in the case of a two-dimensional free energy surface, the PMF of the C<sub>1-7</sub> hydrogen bond,  $\mathcal{W}(r, d)$ , was calculated in the presence of a single water molecule at a distance  $d$  from the carbonyl oxygen of the C=O group ( $d$  is the water hydrogen – carbonyl oxygen distance). The two-dimensional free energy surface was computed using a protocol similar to that described above. The potential energy of the system was biased with two harmonic potentials, centered on successive values of the distances  $r$  and  $d$ , i.e.,  $w_i(r, d) = \frac{1}{2}K_r(r - r_i)^2 + \frac{1}{2}K_d(d - d_i)^2$ , where  $K_r$  and  $K_d$  are the harmonic force constant of the window potential. After equilibration for 10 ps, Langevin trajectories of 100 ps were generated for an eight by eight grid of

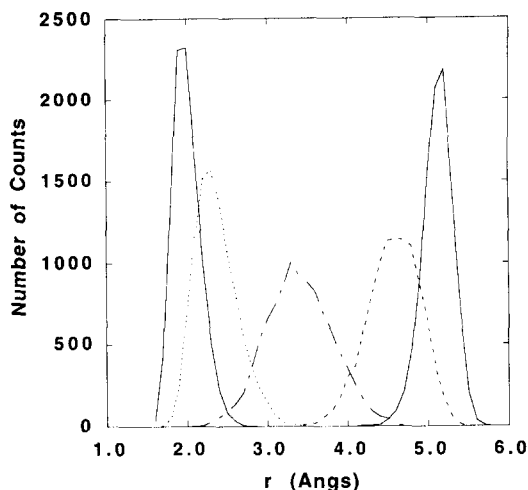


Fig. 1. Histograms generated for each window for calculation of the PMF of the intramolecular hydrogen bond characterizing the C<sub>1-7</sub> conformation. The histograms were calculated with a bin width of 0.1 Å. A total number of 10000 data was used from the time-series. The average position from each window (1 to 5) was 2.05403, 2.41345, 3.44748, 4.62168 and 5.16744 Å, respectively, and the rms fluctuation was 0.17047, 0.26313, 0.41399, 0.32528 and 0.19046 Å.

windows along  $r$  and  $d$ , centered on 1.0, 1.5, 2.0, ..., 5.5 Å with force constants of 10 kcal/mol/Å<sup>2</sup>. The self-consistent set of equations (WHAM, W-PMF and W-DF) were iterated until changes in the free energy constants  $F_i$  were less than 0.001 kcal/mol. No stable self-consistent solution could be found with W-PMF. The free energy differences in the FEP method were calculated by averaging the backward and forward perturbations and the resulting  $\mathcal{W}_i(r)$  were combined together using a two-dimensional extension of Eq. (15).

#### 4. Results and discussion

The histograms of the five windows are shown in Fig 1; all the calculated PMF's along the N-H...O=C intramolecular distance  $r$  are shown in Fig 2. The three PMF's calculated self-consistently (WHAM, W-PMF and W-DF) agree well with the PMF extracted from the 2 ns unbiased simulation. The results from FEP differ most, in particular at the local minimum near 5.05 Å. The largest deviations with the PMF extracted from the unbiased simulation are in the order of 0.10

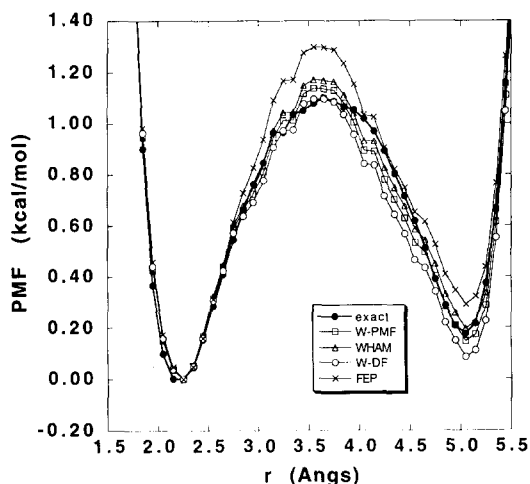


Fig. 2. PMF of the intramolecular hydrogen bond characterizing the  $C_{1-7}$  conformation. ( $r$  corresponds to the  $N-H \cdots O=C$  distance). The PMF's calculated from the unbiased 2 ns simulation (exact), free energy perturbation (FEP), and the self-consistent weighted histogram analysis method (WHAM), the weighted PMF method (W-PMF), and the weighted distribution function (W-DF) are shown. The FEP windows were recombined using Eq. (15).

kcal/mol with WHAM and 0.06 kcal/mol with W-PMF in the region of the central barrier. In this region a larger error may be expected because the number of counts in the histogram bins of the biased simulations is smaller; the relative error in the histogram bins is in the order of  $1/\sqrt{N}$ , where  $N$  is the number of counts [10]. The average number of counts is around 2500 for  $r$  around 2 and 5 Å, whereas it is only 600 around 3.5 Å due to the influence of the central barrier. For comparison, the number of counts in those bins was 1200 for the 2 ns simulation. The convergence in the barrier region could be improved by using an approximate guess to the PMF as an overall biasing potential in addition to the successive harmonic window potentials, i.e.,  $w_i(r) = \frac{1}{2}K(r - r_i)^2 - \mathcal{W}^{\text{guess}}(r)$ , such that the resulting free energy profile of the system is more uniform [18].

The free energy constants  $F_i$  used in Eq. (4) to unbias the individual  $\mathcal{W}_i$  are given in Table 1. Generally, the values calculated with FEP using Eq. (12) and those obtained with WHAM and W-PMF using Eq. (9) are in good accord with the free energy constants extracted from the 2 ns unbiased simulation. In the W-PMF method, the best estimate of the PMF is constructed from a linear combination of the individ-

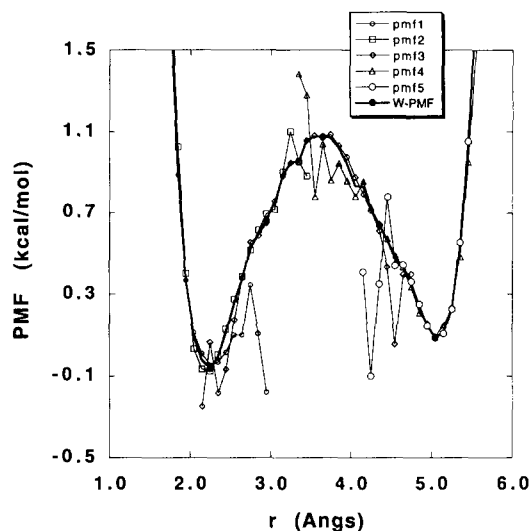


Fig. 3. Individual  $\mathcal{W}_i$  from the five windows shown with the final W-PMF estimate.

Table 1  
Comparison of the free energy constant  $F_i$

	Free energy constants (kcal/mol)				
	Exact	WHAM	W-PMF	W-DF	FEP <sup>a</sup>
$F_1$	0.000	0.000	0.000	0.000	0.000
$F_2$	-1.020	-1.027	-1.041	-1.059	-1.095
$F_3$	-0.321	-0.316	-0.358	-0.410	-0.229
$F_4$	-0.715	-0.745	-0.802	-0.874	-0.716
$F_5$	-0.437	-0.467	-0.525	-0.597	-0.403

<sup>a</sup>Calculated from an average of the forward and backward free energy perturbation with Eq. (12).

ual  $\mathcal{W}_i$  using Eq. (15) whereas those quantities enter as  $\exp(-\mathcal{W}_i/k_B T)$  in Eq. (6) in WHAM and in W-DF. The unbiased  $\mathcal{W}_i$  from each individual window are shown in Fig 3. The individual  $\mathcal{W}_i$  are in good agreement with the final unbiased estimated PMF in the center of the individual windows, although they deviate significantly at the boundaries of the window. In the W-PMF and W-DF methods, the individual estimates are combined with a weighting proportional to the occurrence in the corresponding histograms using Eqs. (14) and (15), respectively. The histograms of the five windows, shown in Fig. 1, are centered on the successive value of the  $H \cdots O$  distance where the individual estimates are accurate.

The two-dimensional free energy surfaces were

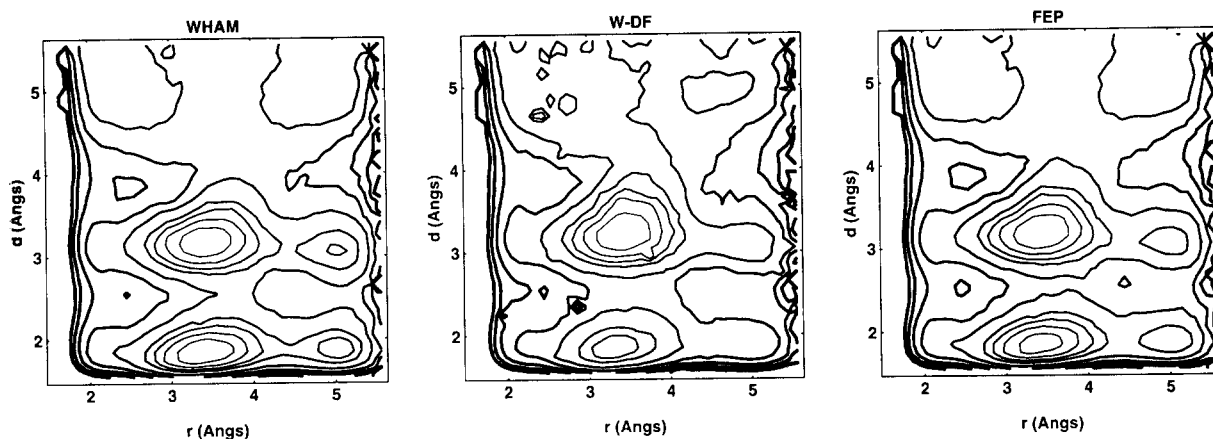


Fig. 4. PMF's of the intramolecular hydrogen bond characterizing the  $C_{1-7}$  conformation in the presence of a water molecule ( $r$  corresponds to the N-H...O=C distance and  $d$  to the water-H...O=C distance). The results obtained with the free energy perturbation (FEP) and from the self-consistent weighted histogram analysis method (WHAM) and weighted distribution function (W-DF) are shown. The contours are traced at the levels 0.25, 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, 3.0, and 3.5 kcal/mol with an increasingly thick line. More information about the six local minima of the free energy surfaces is given in Table 2.

Table 2  
Minima of the two-dimensional free energy surface

Position ( $\text{\AA}$ )		Free energy (kcal/mol)		
$r$	$d$	WHAM	W-DF	FEP
3.35	1.85	0.000	0.000	0.000
3.35	3.15	0.042	–	-0.084
3.35	3.25	–	-0.564	–
2.15	5.45	1.051	–	0.887
2.45	4.65	–	1.008	–
5.05	1.85	0.639	0.635	0.677
5.05	3.05	0.729	–	0.730
4.95	3.05	–	0.635	–
4.85	4.95	1.120	–	0.998
4.55	4.95	–	0.820	–

calculated from the 64 biased simulations using the WHAM, W-DF and FEP approaches. Unexpectedly, it was not possible to reach self-consistency in the case of the W-PMF method for the two-dimensional free energy surface. The reason for this failure is not clear. Nevertheless, it indicates that the approach has severe limitations. The free energy surface calculated with WHAM, W-DF and FEP are shown in Fig. 4; information about the minima of the free energy surfaces is given in Table 2. Whereas the WHAM and FEP free energy surfaces look qualitatively similar, the free energy surface obtained with the W-DF approach

appears to be significantly different. In the case of WHAM and FEP, the location of the six local minima is identical. The relative free energy is slightly different with the absolute minimum found at a different location; e.g., the absolute minimum of one surface is a relative minimum on the other surface, although the free energy difference between the two minima is very small (less than 0.1 kcal/mol). In the case of self-consistent W-DF, the relative energy and even the position of the six minima is appreciably different.

The simple examples considered here suggest that WHAM is the most reliable approach to unbiased and recombine all available data from umbrella sampling calculations. WHAM has many advantages over the more traditional approaches used to unbiased and recombine the umbrella sampling calculations. A particularly important advantage is that WHAM can be easily extended to any number of coordinates. A multidimensional free energy surface may be useful even if the function of interest is a PMF along a single coordinate. For example, sampling problems along a coordinate  $\xi_2$  can be present during the calculation of a PMF along the coordinate  $\xi_1$  [15]. In such a situation, the computation of a two-dimensional free energy surface,  $\mathcal{W}(\xi_1, \xi_2)$ , may be an effective approach to obtain  $\mathcal{W}(\xi_1)$ , which itself can be obtained from a direct numerical integration,

$$e^{-[\mathcal{W}(\xi_1) - \mathcal{W}(\xi_1^*)]/k_B T} = \frac{\int d\xi_2 \langle \rho(\xi_1, \xi_2) \rangle}{\int d\xi_2 \langle \rho(\xi_1^*, \xi_2) \rangle}. \quad (16)$$

Furthermore, the convergence properties of umbrella sampling calculations may be exploited more effectively using WHAM. Generating short umbrella sampling simulations for a large number of narrow windows is computationally more advantageous than generating longer simulations with a smaller number of wider windows [19]. This observation can be demonstrated using a crude argument. Assuming that the dynamics of the umbrella sampling coordinate is governed by a simple damping constant  $\gamma$ , the sampling of the window histogram (in one dimension) should take place on a time scale of  $\tau_w \sim \gamma/K$ , where  $K$  is the force constant of the harmonic window potential. If  $N_w$  simulations are used to cover the whole range  $L$ , the force constant  $K$  of the umbrella sampling potential must be chosen to insure a proper overlap between the adjacent windows, i.e., each window should cover a range of  $\Delta L = L/N_w$  and the value of  $K$  should be on the order of  $k_B T / \Delta L^2$  based on the magnitude of the rms fluctuations. It follows that the total simulation time  $T_{\text{tot}}$  needed to generate the  $N_w$  windows varies as  $\sim L^2/N_w$ . Thus, it is more advantageous to run short umbrella sampling simulations for a large number of narrow windows (the simulation time required to prepare and equilibrate the windows, which is also important, is ignored in this simple analysis) [19]. Traditionally, the uncertainty in matching adjacent windows introduced a cumulative error and there was no clear advantage to generating a large number of windows. With WHAM it should be possible to take advantage of this convergence property.

## 5. Summary

Automatic schemes (WHAM, W-PMF, W-DF and FEP) to unbiased and recombine the windows in umbrella sampling calculation were described and implemented. Tests with the alanine dipeptide system indicated that the approaches generate similar results in

the case of a one-dimensional PMF. Although the four different approaches can, in principle, be extended to the case of multidimensional free energy surfaces, tests with a PMF depending on two coordinate showed that WHAM is the most reliable approach.

## References

- [1] J.G. Kirkwood, *Statistical mechanics of fluid mixtures*, *J. Chem. Phys.* 3 (1935) 300.
- [2] D. Chandler, *J. Chem. Phys.* 68 (1978) 2959–2970.
- [3] G.M. Torrie and J.P. Valleau, *Chem. Phys. Lett.* 28 (1974) 578–581.
- [4] J.P. Valleau and G.M. Torrie, *A guide for Monte Carlo for statistical mechanics*, in: *Statistical Mechanics, Part A*, B.J. Berne, ed. (Plenum Press, New York, 1977) pp. 169–194.
- [5] J. Shen and J.A. McCammon, *Chem. Phys.* 158 (1991) 191–198.
- [6] M. Mezei, P.K. Mehrotra and D.L. Beveridge, *J. Am. Chem. Soc.* 107 (1985) 2239–2245.
- [7] C. Haydock, J.C. Sharp and F.G. Prendergast, *Biophys. J.* 57 (1990) 1269–1279.
- [8] R.W. Zwanzig, *J. Chem. Phys.* 22 (1954) 1420–1426.
- [9] T.B. Woolf and B. Roux, *J. Am. Chem. Soc.* 116 (1994) 5916–5926.
- [10] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman and J.M. Rosenberg, *J. Comp. Chem.* 13 (1992) 1011–1021.
- [11] E.M. Boczek and C.L. Brooks III, *J. Phys. Chem.* 97 (1993) 4509–4513.
- [12] A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.* 63 (1989) 1195–1198.
- [13] C.M. Bennet, *J. Comp. Chem.* 22 (1976) 245–268.
- [14] C.L. Brooks III and L. Nilsson, *J. Am. Chem. Soc.* 115 (1993) 11034–11035.
- [15] T. Woolf, S. Crouzy and B. Roux, *Biophys. J.* 67 (1994) 1370–1386.
- [16] B.R. Brooks, R.E. Brucoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.* 4 (1983) 187–217.
- [17] W.L. Jorgensen, R.W. Impey, J. Chandrasekhar, J.D. Madura and M.L. Klein, *J. Chem. Phys.* 79 (1983) 926–935.
- [18] T.C. Beutler and W.F. van Gunsteren, *J. Chem. Phys.* 100 (1994) 1492–1497.
- [19] S. van Duijneveldt and D. Frenkel, *J. Chem. Phys.* 96 (1992) 4655–4668.