Calculating free energies using average force

By Eric Darve AND Andrew Pohorille † ‡

1. Introduction

Many molecular-dynamics computer simulations of chemically and biologically inter-esting systems are devoted to calculating free-energy changes along selected degrees of freedom. In some instances, the full free-energy profile is of interest. For example, non-monotonic changes in the free energy of two small, hydrophobic species in water as a function of their separation, observed in computer simulations, Ludemann (1996), re-flect the changing patterns of hydrophobic hydration and provide important tests of analytical theories of hydrophobic interactions; Pratt (1977). Free-energy maps of small peptide units in vacuum and in water shed light on conformational preferences of the protein backbone; Brooks (1998). The free-energy profiles associated with the transfer of solutes through water-membrane systems yield solute distributions and permeation rates across membranes: see Wilson (1996), Pohorille (1999). In other instances, calculations of free-energy profiles provide a means of estimating the free-energy difference between the end-points which, in turn, yields the relative stabilities of the corresponding states of the system. Determinations of conformational equilibria in flexible molecules and as-sociation constants between molecular species are among important applications of such calculations: see Giraldo (1998), Schaefer (1998).

The free-energy changes along the chosen generalized coordinates can be calculated from molecular simulations by a variety of techniques: see Frenkel (1996), Berne (1997). Most of them require that a sufficient, thermally-representative sample of states of the system is generated at different values of these coordinates. This leads to the interpre-tation of the free-energy changes along the chosen coordinates as the potential of mean force exerted by other coordinates. Only a few methods for calculating this potential can be conveniently, efficiently and generally combined with computer simulations. One such class of methods relies on obtaining the probability density function, $P(\xi_1, \ldots, \xi_p)$, of finding the system at values $\xi_1, \ldots, \xi_p$ of the p selected generalized coordinates. Once this probability density function is estimated with satisfactory accuracy, the potential of mean force, $A(\xi_1, \ldots, \xi_p)$, can be readily calculated as

$$A(\xi_1, \ldots, \xi_p) = -k_B T \log P(\xi_1, \ldots, \xi_p) \tag{1.1}$$

where T is temperature and k_B is the Boltzmann constant.

Another, general, method for calculating the potential of mean force requires calcu-lating the derivatives $\frac{\partial A}{\partial \xi}$ in a series of calculations, in which ξ_i is constrained to fixed values distributed along $[\xi_i^{min}, \xi_i^{max}]$ in the range of interest. Then, the potential of mean force is recovered by numerical integration. The derivative of the free energy is related to the constraint force needed to keep the system at the fixed value of ξ_i. The exact nature of this relationship was a subject of some debate; see Van Gunsteren (1989), Straatsma (1992), Mülders (1996), den Otter (1998), den Otter & Briels (2000). Several initial

† Exobiology Branch, NASA Ames Research Center
‡ Dept. Pharmaceutical Chemistry, University of California, San Francisco

suggestions were found to be valid only under special circumstances; see Van Gunsteren (1989), Straatsma (1992), Mülders (1996). Only recently, the generally valid and practical to use formula was derived for one-dimensional (den Otter (1998), Sprik (1998), Ruiz-Montero (1997)) and multi-dimensional cases (den Otter & Briels 2000). In this paper, this formula is derived in the general context of multi-dimensional reaction coordinates for constrained and unconstrained simulations. All previous derivations were done in the case of constrained simulations only. This formula requires that the constraint force is corrected by geometric factors that depend on $\xi_1, \ldots, \xi_p$ but not on other (usually difficult to define) generalized coordinates. Since the constraint force can be readily calculated in computer simulations, e.g. using the algorithms SHAKE, see Ryckaert (1977), or RATTLE, see Andersen (1983), practical applications of this method are quite feasible.

Compared to the probability-density method, the constraint-force method has several advantages. In particular, it does not require a good guess of the biasing potential to achieve efficient sampling of $\xi_1, \ldots, \xi_p$. Providing such a guess could be a difficult task, especially for qualitatively new problems. Further, data analysis is markedly simpler; no procedure for matching results obtained for overlapping windows is required. However, the constraint-force method also suffers from several disadvantages. It may be inaccurate or inefficient if the potential of mean force is a rapidly-changing function of $\xi_1, \ldots, \xi_p$. In complex cases, involving, for example, the insertion of a peptide into a membrane or the induced fit of an inhibitor into an enzyme, preparation of the system at consecutive, fixed values of the selected degrees of freedom may be difficult, and subsequent equilibration of the system may be slow. In some instances, application of the constraint-force method may lead to quasi-non-ergodic behavior. Finally, information about the dynamic behavior of the system, which may also be of interest in a simulation, is not available in this approach.

In this paper, we propose an alternative and equally general approach to calculating the potential of mean force, which combines several desired features of both methods. As in the constraint-force method, the potential of mean force is obtained by integrating its derivative. This derivative, however, is calculated from unconstrained rather than constrained simulations. The centerpiece of our method is a new, general formula that connects $\partial A/\partial \xi_i$ with the instantaneous force acting on ξ_i. This force acts along the gradient of ξ_i such that if subtracted from the equations of motion the acceleration of ξ_i is zero. This instantaneous force can be also related to the forces of constraint in a constrained simulation. Then, the forces of constraint are applied to maintain ξ_i at a constant value, and the force acting on ξ_i is exactly equal and opposite to these forces of constraint.

The formula that relates $\partial A/\partial \xi_i$ to the instantaneous force acting on ξ_i is different in unconstrained simulations and constrained simulations. However, as will be shown below, it converges to the den Otter-Briels formula at the appropriate limits. The value of the new formula is not only in providing another route to calculating the potential of mean force but also in clarifying the relationship between the thermodynamic force and the force of constraint. By doing so it forms the theoretical basis for highly-efficient methods to calculate the potential of mean force and to investigate rare events (Darve (2001)).

In the next section we derive the formula for $\partial A/\partial \xi_i$. This is done in two steps. First, the expression for $\partial A/\partial \xi_i$ in unconstrained simulations of a Hamiltonian system is obtained. Then, this expression is generalized so that it applies when the system is only approximately Hamiltonian, as is the case in adiabatic approximation. Then we consider two numerical examples – rotation around the C-C bond of 1,2-dichloroethane immersed

in water, and transfer of fluoromethane across the water-hexane interface. These examples involve only a single reaction coordinate. Applications to multidimensional cases will be considered separately. We close the paper with a comparison of the new method with its alternatives.

2. Theory

2.1. *Generalized coordinates*

We assume that we have a set of M particles and we denote by N the total number of degrees of freedom of our system ($N = 3M$). We further assume that exists a Hamiltonian, H, for this system:

$$H(x_1, \ldots, x_N, p_1, \ldots, p_N) = \frac{1}{2} \sum_i \frac{p_i^2}{m_i} + \Phi(x_1, \ldots, x_N)$$

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial p_i}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial x_i}$$

where $(x_1, \ldots, x_n)$ are Cartesian coordinates, $(p_1, \ldots, p_n)$ are the conjugated momenta, Φ is the potential and t is time.

We suppose that a set of $N - p$ functions $(q_1, \ldots, q_{N-p})$ can be defined such that $(\xi_1, \ldots, \xi_p, q_1, \ldots, q_{N-p})$ forms a complete set of generalized coordinates. We will often denote by x the vector $(x_1, \ldots, x_N)$, and similarly for ξ, q, p_ξ and p_q.

The derivative with respect to ξ_i is defined as the derivative computed with ξ_j, $j \neq i$ and q_k, $k = 1, \ldots, N - p$ constant. Using the definition of A, Eq.(1.1), we can write:

$$\frac{\partial A}{\partial \xi_i} = -k_B T \, \frac{\frac{\partial P}{\partial \xi_i}}{P} \tag{2.1}$$

The probability density P for a canonical ensemble can be written as a function of the Hamiltonian H of the system:

$$P(\xi_1^*, \ldots, \xi_p^*) = \frac{1}{\mathcal{N}} \int dx_1 \ldots dx_N dp_1 \ldots dp_N \, \delta(\xi_1 - \xi_1^*) \ldots \delta(\xi_p - \xi_p^*) \exp(-\frac{H}{k_B T}) \tag{2.2}$$

where $\mathcal{N}$ is a normalization factor.

We introduce additional notation to express the Hamiltonian H as a function of the generalized coordinates.

The Jacobian, J, of the transformation from Cartesian to generalized coordinates is denoted by

$$J \stackrel{def}{=} \begin{pmatrix} J_\xi \\ J_q \end{pmatrix} \tag{2.3}$$

where J_ξ are the first p lines and J_q are the remaining lines. We define matrix Z as:

$$Z \stackrel{def}{=} J M^{-1} J^t$$

where J^t is the transpose of matrix J and M is the mass matrix:

$$M = \begin{pmatrix} m_1 & 0 & \ldots & 0 \\ 0 & m_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & m_N \end{pmatrix}$$

Matrix Z can be written as:

$$Z = \begin{pmatrix} Z_\xi & Z_{\xi q} \\ Z_{q\xi} & Z_q \end{pmatrix}$$

where Z_ξ is a $p \times p$ matrix, $Z_{\xi q}$ a $p \times (N-p)$ matrix, and Z_q a $(N-p) \times (N-p)$ matrix.
  The inverse of Z is denoted by A:

$$A = \begin{pmatrix} A_\xi & A_{\xi q} \\ A_{q\xi} & A_q \end{pmatrix}$$

  Using generalized coordinates, the Hamiltonian of the system takes the form:

$$H(\xi, q, p_\xi, p_q) = \frac{1}{2} p_\xi^t Z_\xi p_\xi + \frac{1}{2} p_q^t Z_q p_q + p_\xi^t Z_{\xi q} p_q + \Phi(\xi, q) \tag{2.4}$$

where p_ξ^t and p_q^t are the transposes of vectors p_ξ and p_q.
  Inserting the expression for P from Eq. (2.2) into (2.1), we obtain:

$$\frac{\partial A}{\partial \xi_i} = \frac{\int dq\, dp_q\, dp_\xi\, \frac{\partial H}{\partial \xi_i} \exp(-\frac{H}{k_B T})}{\int dq\, dp_q\, dp_\xi\, \exp(-\frac{H}{k_B T})} \tag{2.5}$$

with a change of variables from Cartesian coordinates to generalized coordinates. For all functions F, we define the statistical average of F at fixed $\xi^* = (\xi_1^*, \ldots, \xi_p^*)$ as:

$$\langle F \rangle_{\xi^*} = \frac{\int dx_1 \ldots dx_N\, dp_1 \ldots dp_N\, \delta(\xi_1 - \xi_1^*) \ldots \delta(\xi_p - \xi_p^*)\, \exp(-\frac{H}{k_B T})\, F(x_1, \ldots, x_N)}{\int dx_1 \ldots dx_N\, dp_1 \ldots dp_N\, \delta(\xi_1 - \xi_1^*) \ldots \delta(\xi_p - \xi_p^*)\, \exp(-\frac{H}{k_B T})}$$

$$= \frac{\int dq\, dp_q\, dp_\xi\, F(x_1, \ldots, x_N)}{\int dq\, dp_q\, dp_\xi\, \exp(-\frac{H}{k_B T})}$$

where in the last equation $\xi = \xi^*$. With this notation, we can rewrite Eq. (2.5) as:

$$\frac{\partial A}{\partial \xi_i} = \left\langle \frac{\partial H}{\partial \xi_i} \right\rangle_\xi \tag{2.6}$$

After differentiating both sides of Eq. (2.4) and integrating over p_ξ and p_q, we obtain a new expression for Eq. (2.6)

$$\nabla_\xi A = \left\langle \nabla_\xi \Phi + k_B T\, \nabla_\xi \log |J| \right\rangle_\xi \tag{2.7}$$

  The derivative of the free energy can be seen as resulting from two contributions: the mechanical forces acting along ξ and the variations of the volume element associated with the generalized coordinates. This formula has been previously derived in many papers, e.g. Ruiz-Montero (1997) and den Otter & Briels (2000), and is also given by Frenkel and Smit (1996).

2.2. *Thermodynamic force*

In this and the following Sections, we use the fact that, for a given ξ^*, it is possible to choose a basis q such that:

$$Z_{q\xi}(\xi^*, q) = 0, \quad \forall q. \tag{2.8}$$

This choice of q leads to a simplified derivation of our analytical results.

Eq. (2.6) explicitly depends on the choice of all generalized coordinates, including q. As this is not practical from a computational point of view, we now modify this equation to obtain an expression independent of the choice of q. This is done by analytically integrating as many terms as possible in Eq. (2.6).

We start by simplifying the notation:

$$x'_i \overset{def}{=} \sqrt{m_i}\, x_i \qquad\qquad \nabla'_i \overset{def}{=} \frac{1}{\sqrt{m_i}} \frac{\partial}{\partial x_i} \tag{2.9}$$

$$p'_{x_i} \overset{def}{=} \frac{p_{x_i}}{\sqrt{m_i}} \tag{2.10}$$

The symbol $\cdot$ denotes a dot product or a matrix-vector product.

We start from the equation for the time evolution of p_{ξ_i}:

$$\frac{dp_{\xi_i}}{dt} = -\frac{\partial H}{\partial \xi_i} \tag{2.11}$$

The momentum vector p_ξ is defined as the derivative of the Lagrangian with respect to $\dot{\xi}$:

$$p_{\xi_i} \overset{def}{=} \sum_j [A_\xi]_{ij} \frac{d\xi_j}{dt} + \sum_k [A_{\xi q}]_{ik} \frac{dq_k}{dt} \tag{2.12}$$

We can differentiate both sides of Eq. (2.12) with respect to t and use Eq. (2.11) to obtain an expression for $\frac{\partial H}{\partial \xi_i}$. As the right-hand side of Eq. (2.12) is the sum of two products, its derivative contains four terms:

$$\begin{aligned}
\frac{\partial H}{\partial \xi_i} = -\frac{dp_{\xi_i}}{dt} = &-\sum_j \frac{d[A_\xi]_{ij}}{dt} \frac{d\xi_j}{dt} - \sum_j [A_\xi]_{ij} \frac{d^2\xi_j}{dt^2} \\
&- \sum_k \frac{d[A_{\xi q}]_{ik}}{dt} \frac{dq_k}{dt} - \sum_k [A_{\xi q}]_{ik} \frac{d^2 q_k}{dt^2}
\end{aligned} \tag{2.13}$$

Because q satisfies Eq. (2.8), the last term in Eq. (2.13) is equal to zero. Eq. (2.13) can be further transformed using the chain rule of derivation to obtain:

$$\frac{\partial H}{\partial \xi_i} = -\sum_j [Z_\xi^{-1}]_{ij} \frac{d^2\xi_j}{dt^2} + \sum_{jk} [Z_\xi^{-1}]_{ij} \, [\frac{\partial Z_\xi}{\partial x'} \cdot p'_x]_{jk}\, p_{\xi_k} - \sum_k \frac{d[A_{\xi q}]_{ik}}{dt} \frac{dq_k}{dt} \tag{2.14}$$

By expressing p'_x in terms of p_ξ and p_q, one can prove that the second term of Eq. (2.14) is equal to

$$\begin{aligned}
\sum_{jk} [Z_\xi^{-1}]_{ij}\, [\frac{\partial Z_\xi}{\partial x'} \cdot p'_x]_{jk}\, p_{\xi_k} = &\sum_{jklr} [Z_\xi^{-1}]_{ij} \frac{\partial [Z_\xi]_{jk}}{\partial x'_l} [J'_\xi]_{rl}\, p_{\xi_r} p_{\xi_k} \\
&+ \sum_{jklr} [Z_\xi^{-1}]_{ij} \frac{\partial [Z_\xi]_{jk}}{\partial x'_l} [J'_\xi]_{r+p,l}\, p_{q_r} p_{\xi_k}
\end{aligned} \tag{2.15}$$

where J' is defined as J but with x replaced by x'.

In the last equation, we have split the right-hand side into odd and even functions of p_ξ and p_q . We want to compute

$$\int dp_q dp_\xi \exp(-\frac{H}{k_B T})\frac{\partial H}{\partial \xi_i}$$

Because we chose a basis q such that Eq. (2.8) is true, the function $\exp(-\frac{H}{k_B T})$ is even in p_ξ and p_q. Therefore in Eq. (2.15), all odd terms in p_ξ and p_q cancel whereas even terms contribute. An analytical integration over p_{ξ_k} leads to:

$$\Big\langle \sum_{jk} [Z_\xi^{-1}]_{ij} \, [\frac{\partial Z_\xi}{\partial x'} \cdot p'_x]_{jk} \, p_{\xi_k} \Big\rangle_\xi = k_B T \Big\langle \sum_{jkrl} [Z_\xi^{-1}]_{ij} \frac{\partial [Z_\xi]_{jk}}{\partial x'_l} [Z_\xi^{-1}]_{kr} \frac{\partial \xi_r}{\partial x'_l} \Big\rangle_\xi \qquad (2.16)$$

Using the fact that vectors $\frac{1}{m_s}\frac{\partial q_r}{\partial x_s}$ satisfy Eq. (2.8) and are therefore orthogonal to $\nabla \xi_1,..., \nabla \xi_p$, one can prove that the third term on the right-hand side of Eq. (2.14) does not contribute to $\langle \nabla_\xi H \rangle_\xi$. In matrix notation, inserting Eq. (2.16) in Eq. (2.14) gives:

$$\langle \nabla_\xi H \rangle_\xi = k_B T \, \langle \sum_l \frac{1}{m_l} Z_\xi^{-1} \cdot \partial_l Z_\xi \cdot Z_\xi^{-1} \cdot \nabla \xi \rangle_\xi \; - \; \langle Z_\xi^{-1} \frac{d^2 \xi}{dt^2} \rangle_\xi \qquad (2.17)$$

where we denote $\partial_l Z_\xi = \frac{\partial Z_\xi}{\partial x_l}$.

If we denote by λ the vector of RATTLE Lagrange multipliers – Andersen (1983) – they are by definition such that:

$$Z_\xi \lambda \stackrel{def}{=} -\frac{d^2 \xi}{dt^2}. \qquad (2.18)$$

We now summarize what we have obtained so far. We started our derivation from Eq. (2.6), which relates the derivative of A with respect to ξ_i to the average of $\frac{\partial H}{\partial \xi_i}$. We observed that this expression is not very useful, because it depends on a particular choice of generalized coordinates. We transformed this expression by analytically integrating some terms and obtained Eq. (2.17). This new expression is much more useful than the initial one, Eq. (2.6), as it can be computed numerically without any explicit reference to a particular choice of generalized coordinates. Finally, by inserting Eq. (2.18) in the last term of Eq. (2.17), we obtain:

$$\nabla_\xi A = \Big\langle \lambda + k_B T \sum_l \frac{1}{m_l} Z_\xi^{-1} \cdot \partial_l Z_\xi \cdot Z_\xi^{-1} \cdot \partial_l \xi \Big\rangle_\xi \stackrel{def}{=} \Big\langle F_\xi^{(1)} \Big\rangle_\xi \qquad (2.19)$$

Eq. (2.19) has a similar interpretation to Eq. (2.7) although the terms are different. The first term λ is related to the force acting along ξ, which is the opposite of the constraint force. The second term $\sum_l \frac{1}{m_l}\partial_l Z_\xi^{-1} \cdot \partial_l \xi$ is a correction term which accounts for the variation of an infinitesimal volume element in generalized coordinates.

2.3. *Decoupled degrees of freedom*

It is often desirable to consider a situation where ξ is decoupled from the other degrees of freedom. By decoupling we mean that $\frac{d^2 \xi}{dt^2}$ is not a function of the coordinates q, but instead is governed by some other equation of motion. In the previous paper, Darve (2001), we derived the formula for $\frac{\partial A}{\partial \xi}$ that applies to a single reaction coordinate. In this paper, this formula is generalized to a multi-dimensional case.

One example of decoupling is a constrained simulation in which ξ is constant. In this

case $\dot{\xi} = 0$ and $\frac{d^2\xi}{dt^2} = 0$. We will see that using Eq. (2.19) we recover the result of den Otter and Briels (2000). Our derivation can thus be seen as a generalization of their result. Another choice, which was previously discussed in Darve (2001), is a diffusion equation such that the motion of ξ is random and approximately adiabatic. The choice of a Langevin equation is a convenient one because adiabatic approximation can be achieved simply by varying the diffusion constant.

Deriving the relation for $\nabla_\xi A$ in the decoupled case requires modifying the probability density of p_ξ. Previously this density was given by:

$$f_\xi = \exp(-\frac{\frac{1}{2}p_\xi^t Z_\xi p_\xi}{k_B T}) \tag{2.20}$$

For a constraint simulation, f_ξ becomes a Dirac delta function at the location of the constraint whereas for the other decoupled case, f_ξ is a constant function. Since the equation for the decoupled case can be used for an arbitrary f_ξ, it can be seen as a generalization of Eq. (2.19). Thus, one can implement the equation for the decoupled case (Eq. (2.21)) and use it in all situations.

After calculating analytically the integral over p_ξ in Eq. (2.19) with f_ξ given by Eq. (2.20), we obtain the correction for the decoupled case:

$$\frac{\partial A}{\partial \xi_i} = \frac{\left\langle \frac{1}{|Z_\xi|^{1/2}} \left(\lambda + \sum_j [Z_\xi^{-1}]_{ij} \left((\frac{d\xi}{dt})^t \cdot \tilde{\mathcal{H}}_j \cdot (\frac{d\xi}{dt}) + \frac{k_B T}{2} \nabla' \xi_j \cdot \nabla' \log |Z_\xi| \right) \right) \right\rangle_\xi}{\left\langle \frac{1}{|Z_\xi|^{1/2}} \right\rangle_\xi} \tag{2.21}$$

In this equation, we have used the notation $\tilde{\mathcal{H}}_j$ for:

$$\tilde{\mathcal{H}}_j = Z_\xi^{-1} \, J_\xi' \, \mathcal{H}_j \, (J_\xi')^t \, Z_\xi^{-1} \tag{2.22}$$

Note that $\tilde{\mathcal{H}}_j$ is a function only of the first and second derivatives of ξ with respect to Cartesian coordinates, and thus can be easily computed numerically.

2.4. *Constrained simulation*

In the particular case of a constraint simulation, $\dot{\xi} = 0$, which leads to:

$$\frac{\partial A}{\partial \xi_i} = \frac{\left\langle \frac{1}{|Z_\xi|^{1/2}} \left(\lambda + \frac{k_B T}{2} \sum_j [Z_\xi^{-1}]_{ij} (\nabla' \xi_j \cdot \nabla' \log |Z_\xi|) \right) \right\rangle_\xi}{\left\langle \frac{1}{|Z_\xi|^{1/2}} \right\rangle_\xi} \tag{2.23}$$

This is the formula obtained by den Otter and Briels. Note that this formula is applicable to the case of several degrees of freedom. Several authors derived a similar equation for a single reaction coordinate (see den Otter (1998), Sprik (1998)). Note that in Eq. (2.23), Z_ξ is a matrix, $|Z_\xi|$ denotes its determinant and λ is a vector. This contrasts with the case of a single reaction coordinate in den Otter (1998), Sprik (1998).

3. Numerical Results

To examine the performance of the method based on Eq. (2.21), we studied two test cases. One example involved calculating the potential of mean force for the rotation of the C-C bond in 1,2-dichloroethane (DCE) dissolved in water. In the second example, the potential of mean force for the transfer of fluoromethane (FMet) across the water-hexane interface was obtained.

The first system consisted of a DCE molecule surrounded by 343 water molecules, all placed in a cubic box whose edge length was 21.73 Å. This yielded a water density approximately equal to 1 g/cm^3. The second system contained one FMet molecule and a lamella of 486 water molecules in contact with a lamella of 83 hexane molecules. This system was enclosed in a box, whose x, y-dimensions were 24×24 Å^2 and the z-dimension, perpendicular to the water-hexane interface, was equal to 150 Å. Thus, the system contained one liquid-liquid interface and two liquid-vapor interfaces. The same geometry was used in a series of previous studies on the transfer of different solutes across the water-hexane interface; see Darve (2001). In both cases, periodic boundary conditions were applied in the three spatial directions.

For DCE in water, the potential of mean force was calculated along ξ, defined as the Cl-C-C-Cl torsional angle. For the transfer of FMet across the water-hexane interface, ξ was defined as the z component of the distance between the centers of mass of the solute and the hexane lamella (since both cases involved only one-dimensional potentials of mean force, we drop the subscript i following ξ). For each system, three sets of calculations were performed. They yielded $A(\xi)$, using the probability-density method and the methods of the constraint force from unconstrained and constrained simulations.

To obtain $A(\xi)$ from the probability density method, a series of simulations was performed. For DCE, we used a single window and a biasing potential obtained previously (Darve (2001)). The trajectory was 2 ns long. For FMet, ξ was constrained by a harmonic potential in five overlapping windows. No biasing potential was applied. For each window, a molecular-dynamics trajectory 2.4 ns long was obtained. From this trajectory the probability density, $P(\xi)$, was calculated. The probability density in the full range of ξ was constructed by matching $P(\xi)$ in the overlapping regions of consecutive windows. $A(\xi)$ was calculated from the complete $P(\xi)$ using Eq. (1.1).

Calculations of $\frac{\partial A}{\partial \xi}$ from unconstrained simulations were very similar. For DCE, we used a biasing potential and one window. For FMet we did not use a biasing potential and divided the full range of ξ into five windows. In these simulations, however, there was no need for windows to overlap. The molecular-dynamics trajectory in each window was 1.5 ns long. In each molecular dynamics step, the force of constraint was calculated using RATTLE. Since no biasing force was applied the average force in each bin along ξ was simply the arithmetic average of the instantaneous forces.

$\frac{\partial A}{\partial \xi}$ was obtained from constrained simulations by generating a series of trajectories, in which ξ was fixed at several values uniformly spanning the full range of interest. For DCE, simulations were carried out at 37 values of ξ in the range between 0 and 180 deg. This corresponds to 5 deg separation between two values of ξ. For FMet, ξ was fixed at 102 values between -10.1 Å and 10.1 Å (0.2 Å separation between two values). The constraints on ξ were enforced using RATTLE. The average thermodynamic force was obtained by correcting the calculated constraint force according to Eq. (2.23). Once calculations of $\frac{\partial A}{\partial \xi}$ were completed for all discrete values of ξ, $A(\xi)$ was obtained by numerical integration.

The potentials of mean force for rotation of DCE in water and transfer of FMet across the water-hexane interface, obtained from all three methods, are shown in Figs. 1 and 2, respectively. For DCE, gauche and trans conformations were found to have nearly the same free energy, and were separated by a barrier 4.2 kcal/mol high. The gauche conformation corresponds to a torsion angle of about $+/-$ 60 deg (i.e. Cl is not in the plane defined by C-C-Cl) and the trans conformation corresponds to a torsion angle of 180 deg (i.e. this is the conformation where the two Cl are on the opposite sides of the C-

C bond). These results are in close agreement with the results obtained previously using the same potential functions (Benjamin (1993), Pohorille (1993)). For FMet, the free energy between dissolving this molecule in water and in hexane was found to be 0.6 kcal mol^{-1}. An appreciable minimum in the potential of mean force, approximately 1.4 kcal mol^{-1} deep, was observed near the interface. A very similar profile of $A(\xi)$ was obtained using the particle-insertion method of Pohorille, Chipot, New and Wilson (1996).

4. Discussion

In both numerical examples presented in the previous section, the method based on calculating the probability density along ξ and both methods relying on calculating $\frac{\partial A}{\partial \xi}$ yield the potentials of mean force that are identical to within statistical error. This confirms applicability of Eq. (2.19) to the calculation of the potential of mean force in unconstrained simulations.

The method based on Eq. (2.19) has an important advantage over the probability density method. No post-processing of the data obtained from different windows, such as WHAM (Kumar (1995)), is needed. The average force in a given bin along ξ is simply the arithmetic average of instantaneous forces recorded in this bin in all windows (if different biasing forces were used in different windows they have to be subtracted before the average is calculated). In fact, no overlapping between consecutive windows is needed if a sufficiently good estimate of the average force is obtained from one window.

The new approach does not suffer from the disadvantages of the method based on calculating the force in constrained simulations. These disadvantages were discussed in the Introduction. In addition, calculating the forces of constraints becomes less demanding. In constrained simulations, analytical formulas for calculating forces of constraints cannot be used. Instead, iterative procedures with very low tolerance, sometimes requiring double-precision arithmetic, have to be applied. This is needed to prevent drift of the constraint from the preset value due to the accumulation of numerical errors. This problem, however, does not exist in unconstrained simulations. Accuracy in calculating the forces of constraints does not influence motion of the system. This calculation is just a measurement performed on the system and should be done sufficiently accurately that numerical errors associated with this measurement have only negligible contribution to the statistical error of the average force. This is not a very stringent requirement.

REFERENCES

ANDERSEN, H. C. 1983 RATTLE: a 'velocity' version of the SHAKE algorithm for molecular dynamics calculations. *J. Comput. Phys.* **52**, 24-34.

BENJAMIN, I. & POHORILLE, A. 1993 Isomerization reaction dynamics and equilibrium at the liquid-vapor interface of water — a molecular dynamics study. *J. Chem. Phys.* **98**, 236-242.

BERNE, B. J. & STRAUB, J. E. 1997 Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opin. Struct. Biol.* **7**, 181-189.

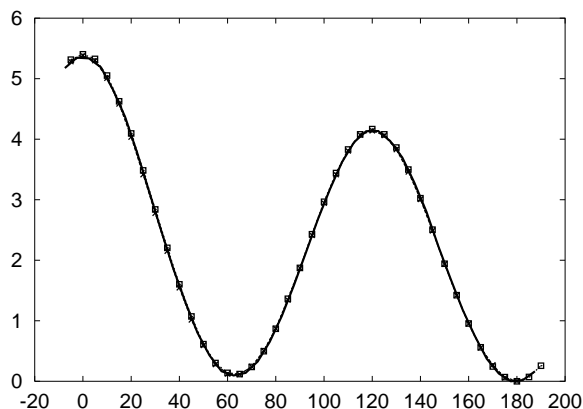BROOKS, C. L., III 1998 Simulations of protein folding and unfolding. *Curr. Opin. Struct. Biol.* **8**, 222-226.

FIGURE 1. The free energy of rotating DCE around the C-C bond computed using the probability density method (———) and the methods of the constraint force from unconstrained (—□—) and constrained simulations (—×—). On the x-axis is the value of the Cl-C-C-Cl torsional angle (in deg). On the y-axis is the free energy (in kcal mol^{-1}).
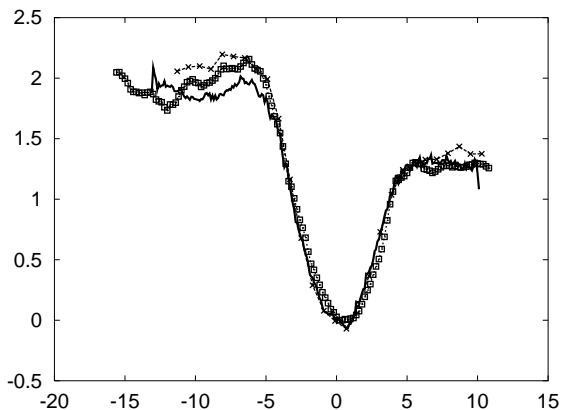

FIGURE 2. The free energy of transferring FMet across the water-hexane interface computed using the probability density method (———) and the methods of the constraint force from unconstrained (—□—) and constrained simulations (—×—). On the x-axis is the value of of the reaction coordinate ξ (in Å). On the y-axis is the free energy (in kcal mol^{-1}).

CHERN, S. S., CHEN, W. H. & LAM, K. S. 1999 *Lectures on differential geometry.* World Scientific.

DARVE, E., WILSON, M. A. & POHORILLE, A. 2001 Calculating free energies using scaled-force molecular dynamics algorithm. *Mol. Sim.*, in press.

DEN OTTER, W. K. 2000 Thermodynamic integration of the free energy along a reaction coordinate in Cartesian coordinates. *J. Chem. Phys.* **112**, 7283-7292.

DEN OTTER, W. K. & BRIELS, W. J. 2000 Free energy from molecular dynamics with multiple constraints. *Mol. Phys.* **98**, 773-781.

DEN OTTER, W. K. & BRIELS, W. J. 1998 The calculation of free-energy differences by constrained molecular-dynamics simulation. *J. Chem. Phys.* **109**, 4139-4146.

FRENKEL, D. & SMIT, B. 1996 *Understanding molecular simulation.* Academic Press.

GIRALDO, J., WODAK, S. J. & VAN BELLE, D. 1998 Conformational analysis of GPA and GPAP in aqueous solution by molecular dynamics and statistical methods. *J. Mol. Biol.* **283**, 863-882.

JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W. & KLEIN, M. L. 1983 Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935.

JORGENSEN, W. L., MADURA, J. D. & SWENSON, C. J. 1984 Optimized potential energy functions for liquid hydrocarbons. *J. Am. Chem. Soc.* **106**, 6638-6646.

KUMAR, S., ROSENBERG, J. M., BOUZIDA, D., SWENDSEN, R. H. & KOLLMAN, P. A. 1995 Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **16**, 1339-1350.

LUDEMANN, S., SCHREIBER, H., ABSEHER, R. & STEINHAUSER, O. 1996 The influence of temperature on pairwise hydrophobic interactions of methane-like particles: a molecular-dynamics study of free-energy. *J. Chem. Phys.* **104**, 286-295.

MARTYNA, G. J., KLEIN, M. L. & TUCKERMAN, M. 1992 Nose-Hoover chains — the canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635-2643.

MÜLDERS, A., KRÜGER, P., SWEGAT, W. & SCHLITTER, J. 1996 Free energy as the potential of mean constraint force. *J. Chem. Phys.* **104**, 4869-4870.

POHORILLE, A. & WILSON, M. A. 1993 Isomerization reactions at aqueous interfaces. *Reaction Dynamics in Clusters and Condensed Phases — The Jerusalem Symposia on Quantum Chemistry and Biochemistry* (Jortner, J., Levine, R. D. & Pullman, B., eds.), Kluwer, Dordrecht, **26**, 207.

POHORILLE, A., CHIPOT, C., NEW, M. H. & WILSON, M. A. 1996 Molecular modeling of protocellular functions.
Pacific Symposium on Biocomputing '96 (Hunter, L. & Klein, T. E., eds.), World Scientific, Singapore, 550-569.

POHORILLE, A. & WILSON, M. A. 1996 Excess chemical potential of small solutes across water-membrane and water-hexane interfaces. *J. Chem. Phys.* **104**, 3760-3773.

POHORILLE, A., WILSON, M. A., CHIPOT, C., NEW, M. H. & SCHWEIGHOFER, K. S. 1999 Interactions of small molecules and peptides with membranes. Lesczynski, J. (Ed.), *Computational Molecular Biology*, Theoretical and Computational Chemistry, Elsevier, Amsterdam, 485-526.

PRATT, L. R. & CHANDLER, D. 1977 Theory of hydrophobic effect. *J. Chem. Phys.* **67**, 3683-3704.

RUIZ-MONTERO, M. J., FRENKEL, D. & BREY, J. J. 1997 Efficient schemes to compute diffusive barrier crossing rates. *Mol. Phys.* **90**, 925-941.

RYCKAERT, J., CICCOTTI, G. & BERENDSEN, H. J. C. 1977 Numerical integration of the Cartesian equations of motion for a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327-341.

SCHAEFER, M., BARTELS, C. & KARPLUS, M. 1998 Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* **284**, 835-848.

SPRIK, M. & CICCOTTI, G. 1998 Free energy from constrained molecular dynamics. *J. Chem. Phys.* **109**, 7737-7744.

STRAATSMA, T. P., ZACHARIAS, M. & MCCAMMON, J. A. 1992 Holonomic constraint contributions to free-energy differences from thermodynamic integration molecular-dynamics simulations *Chem. Phys. Lett.* **196**, 297-302.

Van Gunsteren, W. F. 1989 Methods for calculation of free energies and binding constants: Successes and problems, In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, Van Gunsteren, W. F. & Weiner, P. K. (Eds.), ESCOM, 27-59.

Wilson, M. A. & Pohorille, A. 1996 Mechanism of unassisted ion transport across membrane. *J. Am. Chem. Soc.* **118**, 6580-6587.