# Hamiltonian and Distance Replica Exchange Method Studies of Met-Enkephalin

## Li Su and Robert I. Cukier*

*Department of Chemistry and the Quantitative Biology Modeling Initiative, Michigan State University, East Lansing, Michigan 48824-1322*

*Received: April 30, 2007; In Final Form: August 1, 2007*

The conformational states of the zwitterionic form of the pentapeptide Met-enkephalin were explored with the use of explicit solvent molecular dynamics (MD). The N and C termini are ionized, as appropriate to polar solvent conditions, and consequently, there is a competition between open forms driven by polar solvation of the ammonium and carboxylate groups and closed forms driven by their salt-bridge formation. Normal MD started from an open state does not sample closed conformations. Sampling was enhanced with a distance replica exchange method (DREM) and with a Hamiltonian replica exchange method (HREM). The potential of mean force (PMF) along an end-to-end distance reaction coordinate was obtained with the DREM. The PMF shows a stable salt-bridge state and the presence of a large region of open states, as hypothesized for conformationally promiscuous small opiate peptides. The HREM systems differ by scaling the peptide−peptide and peptide−solvent electrostatic and Lennard-Jones potentials, with the goal of improving the sampling efficiency with a limited number of systems. A small number of systems were found to be sufficient to sample closed and open states. A principal component analysis (PCA) shows that the HREM-generated fluctuations are dominated by the first two principal modes. The first corresponds to the end-to-end reaction coordinate found in the DREM, and the first mode PMF is similar to the DREM PMF. The second mode describes the presence of two conformations, both of which correspond to the salt-bridge state distance. The conformers differ in the values of neighboring psi and phi dihedral angles, since such psi/phi compensation can still produce the same end-to-end distance. The two-dimensional PMF constructed from the first two PCA modes captures most of the significant backbone conformational space of Met-enkephalin.

## 1. Introduction

Protein molecular dynamics (MD) simulations with explicit solvent run on the current practical time scale of nanoseconds tend to stay around their initial configurations. Even for peptide simulations in explicit solvent started from an extended state, states that are more compact may not be properly sampled. This generic sampling problem, in which barriers large compared with the thermal energy separate stable states, is a major concern in MD and Monte Carlo (MC) simulations. Methods such as multicanonical ensemble,[1,2] simulated tempering,[3,4] and parallel tempering, also referred to as the replica exchange method (REM),[5−11] were designed to address this issue. The original versions of the REM used temperature (TREM), and recently, a Hamiltonian REM (HREM) was introduced.[12] These methods contain two elements: (1) multiple copies of configurations are run independently by MD or MC with different temperatures (TREM) or Hamiltonians (HREM). (2) Two neighbors (different temperatures or Hamiltonians) may be exchanged, according to the Metropolis−Hastings algorithm.[13] In the TREM, sampling is improved at the desired (lowest) temperature by higher-temperature replicas overcoming barriers in the potential energy surface.[14] TREM suffers from the deficiency that the number of system copies needed scales with the square root of the number of degrees of freedom of the system of interest.[12]

To address this difficulty with TREM, Fukunishi et al.[12] introduced the HREM, in which the potential functions in different Hamiltonians differ only in a limited set of the total

number of degrees of freedom required to characterize the system, thereby, in principle, reducing the number of Hamiltonians needed. A specialization of the Hamiltonian REM that is suited to studies in which a reaction coordinate can be identified integrates the umbrella sampling method[13] with the HREM. Namely, different systems are created by adding different umbrella window potential functions to the original system, which serve to restrain the systems to sample desired values along the reaction coordinate. This specialization of HREM, which we will refer to as the distance REM (DREM), was proposed by Sugita and co-workers in the context of a multidimensional temperature and distance replica exchange method and applied to the alanine trimer.[15] We used the DREM to investigate a large conformational change in adenylate kinase that spans its open (ligand-free) and closed (ligand-bound) structures.[16] This DREM application provided greatly enhanced sampling relative to standard umbrella sampling, as measured by improvements in the speed of decay of time correlation functions and the extent of conformational space exploration.[16]

In this work, the conformational space of Met-enkephalin, an opioid pentapeptide with sequence Tyr-Gly-Gly-Phe-Met, is explored with the use of explicit solvent MD simulations and the HREM and DREM. Met-enkephalin has been shown to exhibit great conformational plasticity by experiment[17−22] and computation[2,7,23−31] and provides a good test of methods that can enhance the rate of conformational sampling. Of particular interest is the zwitterionic form (protonated N-terminus and ionized C-terminus), which should predominate in polar media. The competition between a salt-bridged closed form and charge-solvated (terminal peptide charges interacting with solvent

* Corresponding author. Phone: 517-355-9715, ext 263. Fax: 517-353-1793. E-mail: cukier@cem.msu.edu.

Studies of Met-Enkephalin

*J. Phys. Chem. B, Vol. 111, No. 42, 2007* **12311**

dipoles) open forms may lead to a reasonable proportion of closed and open forms. A number of MD-based simulations of Met-enkephalin and the closely related Leu-enkephalin have been carried out. Van der Spoel and Berendsen[25] discussed the backbone and side-chain conformational space of Leu-enkephalin and noted that the zwitterionic form is quite labile and samples folded, closed forms. Aburi and Smith[26] simulated Leu-enkephalin as a function of pH and found that at neutral pH, where they used the zwitterionic form, a roughly equal mixture of folded and unfolded states exists. They found that starting from an open form, the peptide rapidly closes but eventually will reopen on their 10 ns time scale. Nielsen et al.[27] simulated Met- and Leu-enkephalin in zwitterionic forms, found a rapid transition from extended to stable folded conformers, and investigated the details of the side-chain conformations of the folded forms. Shen and Freed[28] compared the results of explicit and implicit solvent simulations and found mostly compact with some extended conformations. Zaman et al.[29] carried out implicit water Langevin dynamics simulations of Met-enkephalin with a variety of force fields to compare their predictions for long (~130 ns) time scales. Karvounis et al.[30] simulated zwitterionic Leu-enkephalin in explicit solvent initiating a number of trajectories from open forms and found a variety of behaviors, including a persistent salt-bridge form. Sanbonmatsu and Garca[31] simulated neutral Met-enkephalin in explicit solvent using the TREM. This study demonstrated the enhanced sampling capability of a REM relative to conventional MD, showing that it could surmount the barriers separating nonhelical from helical conformations in the simulation interval.

The explicit solvent zwitterionic[25−27] studies in which both open and closed conformations were found show that the time scale for these transitions is multinanosecond; thus, in typical simulation times, only a few events can be observed. To obtain a potential of mean force (PMF) and an equilibrium constant between compact and extended conformers, either much longer simulation times must be used or REM or other biasing techniques need to be introduced. The DREM simulations we carry out show that Met-enkephalin has a distinct salt-bridge form that is separated by a low barrier from a broad range of open forms, when the end-to-end distance is used as a PMF reaction coordinate, with a roughly equal mixture of closed and open form conformers. The HREM method supports this conclusion and provides a more complete picture of the conformational states sampled by Met-enkephalin by revealing another significant reaction coordinate; namely, a correlated compensating transition of succeeding psi and phi dihedrals that permits the existence of two distinct conformers for the salt-bridge form. The HREM is very effective in sampling the large conformation space of Met-enkephalin at a reasonable computational cost.

In Section 2, the DREM and HREM are introduced, and details of their applications to the simulation are provided. A new approach to separating a PMF into its energetic and entropic contributions is developed. Section 3 presents our results, and conclusions are given in Section 4.

## 2. Methodology

**Replica Exchange Methods.** The temperature REM constructs independent copies of a system that differ by their temperature. The REM concept can be generalized to a Hamiltonian REM[12] in which the systems differ by their Hamiltonian (in practice, in their potential energy function). As a matter of terminology, we shall refer to these different Hamiltonians as *systems* (versus replicas), since replica connotes

a copy of an item. For HREM and DREM, it is more natural to use this terminology. The term replicas will be reserved for the configurations that are present on any particular MD step. In CUKMODY, our MD program, a given configuration (replica) is maintained on a particular compute node, and the systems (with different potential functions) move onto and out of that node.

It is worth formulating HREM and DREM separately due to some technical issues in their implementations. For HREM, $H_i(X,P) = T(P) + V_i(X)$ where $T(P)$ is the kinetic energy and $V_i(X)$ is the potential function for the $i$th system with phase space coordinates $X,P$. Between exchange attempts, normal MD is run for each system $i$ characterized by $H_i$. When system interchanges are to be attempted, detailed balance

$$\alpha(\mathbf{X},\mathbf{X}' \rightarrow \mathbf{X}',\mathbf{X})P_i(\mathbf{X})P_j(\mathbf{X}') = \alpha(\mathbf{X}',\mathbf{X} \rightarrow \mathbf{X},\mathbf{X}')P_i(\mathbf{X}')P_j(\mathbf{X}) \quad (2.1)$$

must be enforced. Here, $\alpha(\mathbf{X},\mathbf{X}' \rightarrow \mathbf{X}',\mathbf{X})$ is the acceptance probability (transition probability) that configuration $\mathbf{X}$ in the $i$th system and $\mathbf{X}'$ in the $j$th system before exchange results in configuration $\mathbf{X}'$ in the $i$th system and $\mathbf{X}$ in the $j$th system after exchange, and $P_i(\mathbf{X})$ is the Boltzmann distribution at temperature $T = 1/k_B\beta$ for the $i$th system. The Metropolis rule for exchange between two systems,

$$\alpha(\mathbf{X},\mathbf{X}' \rightarrow \mathbf{X}',\mathbf{X}) = \min(1,e^{-\Delta(\mathbf{X},\mathbf{X}'\rightarrow\mathbf{X}',\mathbf{X})}) \quad (2.2a)$$

where

$$\Delta(\mathbf{X},\mathbf{X}' \rightarrow \mathbf{X}',\mathbf{X}) = \beta[(V_i(\mathbf{X}') - V_j(\mathbf{X}')) + (V_j(\mathbf{X}) - V_i(\mathbf{X}))] \quad (2.2b)$$

guarantees that Boltzmann equilibrium in the extended ensemble of the product of all the systems' ensembles will result for a sufficiently long trajectory.[12] If the potential functions differ by a restricted set of degrees of freedom, only those will contribute to eq 2.2b. In our HREM implementation, the potential energy is parametrized as

$$V_i(\mathbf{X}) = \lambda_i V_{PP}(x_P,x_P) + \sqrt{\lambda_i}\, V_{PS}(x_P,x_S) + V_{SS}(x_S,x_S) \quad (2.3)$$

where the terms in eq 2.3 denote peptide−peptide, peptide−solvent, and solvent−solvent interactions, respectively, and $\lambda_i$ is a scaling factor for the Lennard-Jones and electrostatic nonbonded interactions. In explicit solvent simulations, the number of degrees of freedom is dominated by the solvent; thus, the indicated scaling is much reduced relative to the TREM where the global scaling $\beta V_i(\mathbf{X}) = \beta\lambda_i V(\mathbf{X}) \equiv \beta_i V(\mathbf{X})$ is used.

If the desire is to enhance the sampling at points along an identified reaction coordinate, then systems that differ by a window potential to maintain a desired distance can be introduced to give a distance REM (DREM). A window potential $W_i(\mathbf{r})$ ($i = 1, 2, ..., N_w$) is added such that $H_i(X,P) = H(X,P) + W_i(\mathbf{r})$, where $\mathbf{r} = f(\mathbf{X})$ denotes a reaction coordinate dependent on $\mathbf{X}$. For a one-dimensional distance reaction coordinate, $r$, the $i$th window function is chosen as a harmonic restraint $W_i(r) = (k_i/2)(r - r_0^i)^2$ with $k_i$ and $r_0^i$ the force constant and equilibrium distance, respectively, and the acceptance probability is

$$\Delta(\mathbf{X},\mathbf{X}' \rightarrow \mathbf{X}',\mathbf{X}) = \beta[(W_i(r') + W_j(r)) - (W_i(r) + W_j(r'))] = \beta k(r' - r)(r_0^j - r_0^i) \quad (2.4)$$

with the last equality specialized to equal force constant values.

Between attempted exchanges, conventional MD simulations are performed for the different systems. The exchanges may be thought of either as configuration exchanges or temperature (TREM) exchanges, window (DREM) exchanges, or scale-of-interaction (HREM) exchanges. Computationally, it is efficient that only temperature, window, or scale-of-interaction needs to be exchanged, versus configuration. Exchanges are attempted only between neighboring windows, because for the method to be effective, the overlap between the windows' probability distributions needs to be adequate. In contrast with the TREM (unless ensemble averages at higher temperatures than the "normal" one are of interest) and HREM, in DREM, the information from all the windows is used directly because it will ultimately provide the potential of mean force along the reaction coordinate. As in any window method, the DREM biases the true Hamiltonian. To unbias the corresponding probability distribution and to combine the trajectory data obtained from the different windows, we use the weighted histogram analysis method (WHAM).[32,33] For the HREM, PMFs are obtained by simply constructing histograms of the relevant coordinates from the $\lambda_i = 1$ trajectory.

The potential of mean force PMF($\mathbf{r}$), which is related to the unbiased probability distribution $p^u(\mathbf{r})$ according to PMF($\mathbf{r}$) = $-k_B T \ln(p^u(\mathbf{r})/p^u(\mathbf{r}_{ref}))$, with $\mathbf{r}_{ref}$ some reference value, may be viewed as an $\mathbf{r}$-dependent free energy, $\Delta A(\mathbf{r})$, since it corresponds to the reversible work of changing $\mathbf{r}$ from $\mathbf{r}_{ref}$ to $\mathbf{r}$. It is of interest to decompose $\Delta A(\mathbf{r})$ into energetic and entropic contributions according to $\Delta A(\mathbf{r}) = \Delta E(\mathbf{r}) - T\Delta S(\mathbf{r})$. However, as has been discussed in various ways,[34-39] this decomposition introduces a statistical problem that is not present in the evaluation of $\Delta A(\mathbf{r})$. A straightforward way to expose the difficulty is to consider the thermodynamic integration[13] expressions for some general free energy, $\Delta A$, and energy, $\Delta E$, change,[37]

$$\Delta A = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda = \int_0^1 \langle H_1 - H_0 \rangle_\lambda d\lambda \quad (2.5a)$$

$$\Delta E = \int_0^1 \frac{\partial \langle H(\lambda) \rangle_\lambda}{\partial \lambda} d\lambda = \int_0^1 \frac{\partial \langle H_0 + \lambda(H_1 - H_0) \rangle_\lambda}{\partial \lambda} d\lambda \quad (2.5b)$$

where the second equalities follow from writing $H(\lambda) = H_0 + \lambda(H_1 - H_0)$, and the brackets denote trajectory averages in the $\lambda$ ensemble. In eq 2.5a, when applied to the DREM type Hamiltonian, only the difference of the window potentials $W_i$($\mathbf{r}$) ($i = 0,1$) contributes, and from a statistical perspective, the fluctuations arising from these few degrees of freedom will be small. In eq 2.5b, the required averages are of the total Hamiltonian, which includes all the peptide and solvent degrees of freedom, leading to large fluctuations that require long trajectories. How this problem is manifest in the context of the WHAM and how to accomplish the decomposition $\Delta A(\mathbf{r}) = \Delta E(\mathbf{r}) - T\Delta S(\mathbf{r})$ is discussed in Appendix A. There, we show that $p^{(u)}(E|r)$, the unbiased (by the window potentials) conditional probability of total energy values $E$, given a value of the reaction coordinate $r$, is required to obtain $\Delta E(\mathbf{r})$. In this regard, it is useful to think of the DREM perspective of a fixed system (window potential) having its configuration updated in an exchange move. The condition for bringing in a new configuration is based on the window energy differences in eq 2.4. The total energy, $E(\mathbf{r})$, fluctuates without the constraint of this Metropolis rule. In addition to this REM source of fluctuation, the usual $E \sim \sqrt{N}$, where $N$ is the total number of degrees of freedom source, is also present. Therefore, the quality of $\Delta E(\mathbf{r})$

and $\Delta S(\mathbf{r})$ will be considerably less for a given simulation length than for $\Delta A(\mathbf{r})$.

**Molecular Dynamics Simulations.** The CUKMODY protein molecular dynamics code, which uses the GROMOS96[40] force field, was modified to incorporate the HREM based on the DREM[16] code. The systems are run independently on different nodes of a Linux cluster computer, and when exchanges are attempted, information is passed using the message passing interface technique implemented as MPICH. SHAKE[41] is used to constrain bond distances, enabling a 2 fs time step, and temperature is globally controlled with a Berendsen thermostat[42] with relaxation time of 0.2 ps.

For the evaluation of the electrostatic and the attractive part of the Lennard-Jones energies and forces, the PME method[43] was applied with a direct-space cutoff of 8.52 Å, an Ewald coefficient of 0.32 for DREM and 0.45 for HREM, and a $30 \times 30 \times 30$ reciprocal space grid. Applying a scale factor to the potential in the context of an Ewald method, in which the evaluation of energy and force in the reciprocal space is based on a structure factor (a sum over atoms), necessitates assignment of the scale factor to atoms (versus atom−atom interactions). For example, the scaling in eq 2.3 is based on assigning $\sqrt{\lambda_i}$ (0 $< \lambda_i \le 1$) to the protein atoms and not scaling the solvent atoms. The electrostatic and Lennard-Jones interactions are uniformly scaled; therefore, as $\lambda_i$ decreases, softer Lennard-Jones and reduced electrostatic interactions are obtained.

All simulations were carried out in a box with 30.0 Å sides, having 864 waters initially. The starting Met-enkephalin configuration was obtained from an NMR ensemble (pdb 1PLW).[20] In this configuration, the end-to-end distance (nitrogen of the N-terminus to carboxylate carbon of the C-terminus) is 10.5 Å; this distance in the ensemble of 80 lowest energy structures is ~10−11 Å. The peptide was immersed in the water box, and 51 overlapping waters were removed. Before the DREM, the systems were run for 1 ns with the force constants $k = 4.0$ kcal/mol and equilibrium distances 3.0−11.0 Å in 0.5 Å increments, and 8.0−16.0 Å in 0.5 Å increments for the two 17-system simulations. Exchanges were attempted every 100 steps for the DREM. For an odd number, $2n + 1$, of systems, exchange attempts were alternated among 1∘2, ..., $2n − 1∘2n$ and 2∘3, ..., $2n∘2n + 1$. The relatively small force constant used permitted good overlap of windows for the 0.5 Å window equilibrium distance increments and has the advantage of enhancing the sampling of the orthogonal degrees of freedom. For the HREM, five systems were used with the scale factors set to $\lambda_1 = 1$, $\lambda_2 = 0.925$, $\lambda_3 = 0.85$, $\lambda_4 = 0.775$, and $\lambda_5 = 0.7$. They are started from the same initial configuration, and the first 3 ns are considered as the equilibration time. Exchanges were attempted every 40 steps.

**Principal Component Analysis.** Principal component analysis[44-49] (PCA) diagonalizes the covariance matrix with elements $(\boldsymbol{\sigma})_{ij} = \langle \delta\alpha_i \delta\alpha_j \rangle$ of the atom fluctuations $\delta\alpha_i = \alpha_i - \langle \alpha_i \rangle$ from their trajectory-averaged $\langle \alpha_i(t) \rangle = \int_0^T \alpha_i(t) \, dt/T$ values, where the $\alpha_i = \{x_i, y_i, z_i\}$ denotes the Cartesian components of the position of the $i$th atom. PCA decomposes the configuration point $\mathbf{X}(t) = (x_1(t), y_1(t), .., z_N(t))^T$ as

$$\mathbf{X}(t) = \sum_{i=1}^{3N} [\mathbf{X}(t) \cdot \mathbf{m}_i] \mathbf{m}_i = \sum_{i=1}^{3N} p_i(t) \mathbf{m}_i \quad (2.6)$$

where $\mathbf{m}_i$ is the (orthonormal) eigenvectors of the covariance matrix $\boldsymbol{\sigma}$, the corresponding eigenvalues are denoted as $\lambda_i^2$, and $p_i(t)$ is the mode displacements. The eigenvalues are related to the mean square fluctuation (MSF) of the atoms over the

Studies of Met-Enkephalin

*J. Phys. Chem. B, Vol. 111, No. 42, 2007* **12313**

trajectory as MSF $= \sum_i \lambda_i^2$. In the rotated Cartesian coordinate basis defined by $\mathbf{m}_i$ ($i = 1, 2, ..., 3N$), the largest eigenvalue captures the largest fraction of the MSF; the second largest, the next largest fraction of the MSF; etc. In favorable cases, a small set of modes capture most of the fluctuation, and they may describe collective motion. The PCA analysis can be used as a severe test of simulation convergence because the larger eigenvalues correspond to the slower motions of the peptide. Several convergence tests have been proposed.[50,51] Amadei and co-workers[50] introduced a root-mean-square inner product (RMSIP) measure,

$$RMSIP = \left[ \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{n} \mathbf{m}_k(t) \cdot \mathbf{m}_i(t') \right]^{1/2} \quad (2.7)$$

that evaluates the overlap of a subset of $n$ modes obtained from different time intervals of the total trajectory. Here, we take $t$ and $t'$ to be two disjoint time intervals of the trajectory and use the resulting RMSIPs to monitor the stability of these modes.

The PCA is carried out by using ANALYZER,[52] a program written for the purpose of analyzing trajectory data by a wide variety of methods.

**Fluctuation Measures.** The conventional RMSF$j$ (root-mean-square fluctuation for atom $j$) measure of protein fluctuations (that can be compared to B-factors)[53] is defined as

$$RMSF_j = \sqrt{\left( \int_0^T dt(\mathbf{r}_j(t) - \bar{\mathbf{r}}_j) \cdot (\mathbf{r}_j(t) - \bar{\mathbf{r}}_j) \right)/T} \quad (2.8)$$

where $\bar{\mathbf{r}}_j = \sqrt{\int_0^T dt \mathbf{r}_j(t)/T}$ is the trajectory average position vector for atom $j$.

A time-dependent version that is summed over all the atoms, RMSF_$t(s)$, can also be defined:

$$RMSF\_t(t) = \sqrt{\left( \sum_{j=1}^{N_{at}} (\mathbf{r}_j(t) - \bar{\mathbf{r}}_j) \cdot (\mathbf{r}_j(t) - \bar{\mathbf{r}}_j) \right)/N_{at}} \quad (2.9)$$

### 3. Results

**DREM Simulations.** Two DREM simulations of 6 ns each were run to cover the end-to-end distance (Tyr1 terminal nitrogen to Met5 terminal carboxylate carbon) intervals 3−11 and 8−16 Å. To see if the DREM method provides an advantage relative to a conventional window method, in which each distance-restrained system is simulated independently (no exchanges), a no-DREM simulation was run for the 3−11 Å range. Figure 1a compares the no-DREM and DREM results over the time intervals 2−4 and 4−6 ns. The 0−2 ns interval was considered as an equilibration period, since even for the DREM simulation, the results were not converged. Clearly, the DREM simulation provides better convergence than no-DREM in the more difficult, larger coordinate interval. The close-in distance result (∼3−4 Å) for the no-DREM is also not yet converged because it looks much like the unequilibrated DREM 0−2 ns result. Thus, as has been concluded in both simpler[15] and more complex protein systems,[16] the replica exchanges are accelerating the exploration of configuration space relative to no exchanges.

The PMF for the two distance ranges is shown in Figure 1b. The neighbor window average acceptance ratios ranged from ∼30 to 45%, with the smallest value corresponding to the 6.0−6.5 Å systems (reflecting the peak in the PMF) for the 3−11 Å simulation and from ∼35 to 45% for the 8−16 Å simulation. The itineration over time of (1) different configurations visiting
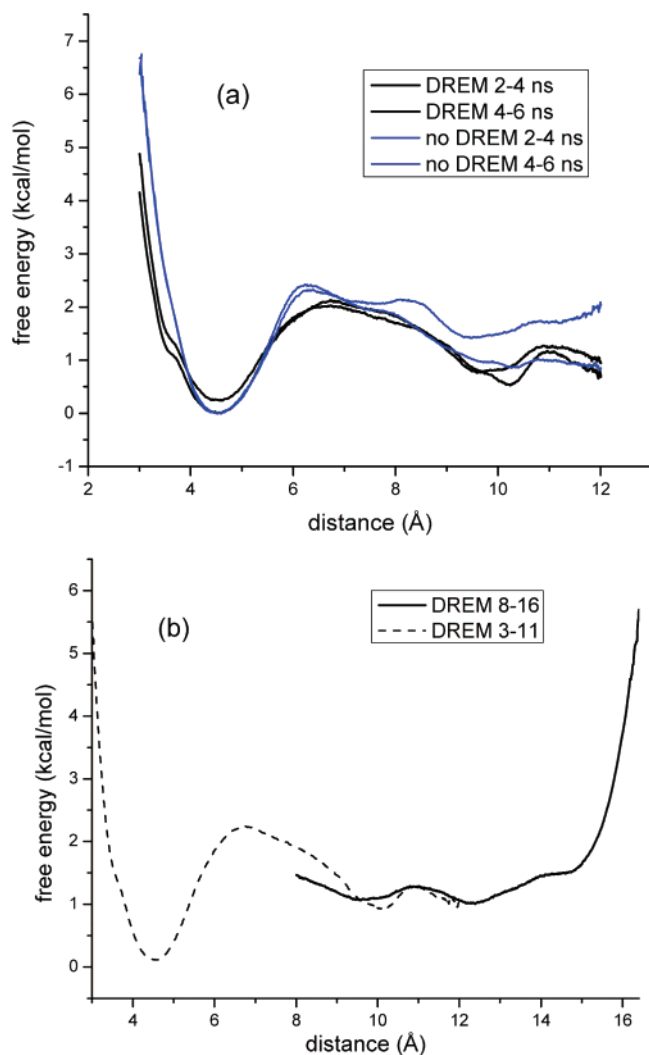


**Figure 1.** (a) The PMF for the end-to-end distance generated in the 3−11 Å DREM and no-DREM simulations for the time intervals 2−4 and 4−6 ns. The lack of convergence of the no-DREM simulation at larger distances is evident. (b) The PMF for the end-to-end distance generated by the DREM simulations for the 3−11 Å and the 8−16 Å ranges.

a fixed system (a given window equilibrium distance) and (2) different systems migrating in and out of a particular configuration showed that the procedure provided good mixing in the extended ensemble. For each simulation, the PMFs constructed also over the 8−16 Å interval are not converged for the 0−2 ns interval; thus, Figure 1b presents the result for the 2−6 ns interval. Examining the 8−16 Å PMFs on 1 ns (2 ns) intervals provides an error estimate of ±0.25 (±0.125) kcal/mol. A similar result was found for the 3−11 Å simulation, with the error concentrated in the 8−11 Å range. The DREM simulations' accuracy may degrade at the lower and upper limits of the restraint interval. Thus, the lack of agreement in the 8−11 Å range between the simulations may reflect this end-point issue as well as statistical fluctuations.

**DREM PMF.** The closed, salt bridge structure around 4.5 Å is the most stable, well-defined state. The minimum around 4.5 Å for the end-to-end distance corresponds approximately to the conventional geometric definition of a salt bridge. In protein interiors, there is some contention as to whether salt bridges are stabilizing.[54] For a small peptide, the issue is somewhat different, relying more on the difference between the free energy of formation of one salt bridge surrounded mainly

**12314** *J. Phys. Chem. B, Vol. 111, No. 42, 2007*

Su and Cukier

**TABLE 1: The Acceptance Ratio for the Time Span of the HREM Simulation**[a]

| potential index | $\lambda_a$ | $\lambda_b$ | acceptance ratio |
|---|---|---|---|
| $0 \leftrightarrow 1$ | 1.0 | 0.925 | 0.424 |
| $1 \leftrightarrow 2$ | 0.925 | 0.85 | 0.451 |
| $2 \leftrightarrow 3$ | 0.85 | 0.775 | 0.457 |
| $3 \leftrightarrow 4$ | 0.775 | 0.7 | 0.431 |

[a] Exchanges were attempted every 40 fs.

by solvent versus the charge-dipole free energy of solvation of the two termini in more extended configurations. The PMF shows that salt bridge conformers are strongly stabilized and localized when defined by the chosen reaction coordinate. The salt bridge is separated from a broad, open region in the interval $10-14$ Å by a 2 kcal/mol barrier around 7 Å. This is a small barrier on the scale of thermal energy ($k_B T = 0.6$ kcal/mol for $T = 300$ K). Nevertheless, without the use of a reaction coordinate window method to sample these close-in configurations, such states are rarely sampled, as noted in the Introduction. Suggestions for why it is difficult to sample closed conformers starting from open forms will be presented in the HREM discussion. The broad region of flat PMF is bounded at the large distance by rapidly rising values, indicating that the peptide is being stretched too much. Even though there is no well-defined open state, a kind of equilibrium constant can still be defined by using the relative populations to the left and right of the barrier at 7 Å. This equilibrium constant slightly favors (about 55%) the salt-bridge versus open states, indicating that in a high dielectric solvent, both closed and open forms of zwitterionic Met-enkephalin are well represented.

**HREM Replica Exchange Diagnostics.** The HREM has the virtue of not requiring specification of a reaction coordinate or its dimensionality. The issue is to limit the number of systems required while still providing robust sampling. In explicit solvent simulations, not scaling the solvent−solvent interactions should provide a reduction in the number of systems relative to a temperature REM that scales all the degrees of freedom. As in all REM versions, the choice and optimization of the acceptance probability of attempted exchanges in the HREM is a central issue. There should be an optimal acceptance probability, because for low exchange probability, the rate of movement through configuration space is small, whereas for high exchange probability, the movement through configuration space is slow. Table 1 lists the acceptance ratios for the HREM simulation; they are all around 0.44. Predescu and co-workers[13] analyzed the optimization of the TREM acceptance ratio for a multidimensional oscillator system and found that 0.3874 is the optimal acceptance ratio, with the efficiency falling off slowly around this value. The uniformity of the acceptance probability values that we find indicate that the choice of the number and spacing (the specific $\lambda_i$ values) are appropriate.

To examine whether all the configurations (replicas) can visit a particular system (with a particular $\lambda_i$ value) and whether given configurations are visited by all the systems, we display these time trajectories in Figure 2. From the plots, it is clear that all the configurations (replicas) can be visited by all the systems (a−e), and conversely, all the systems can be visited by all the configurations (replicas) (f−j). In plot a, the points are vertically connected, and that the plane is covered demonstrates the desired itineration. In all the other plots, we just insert a point to indicate occupancy. On this scale, all the plots look similar and uniform in time, which supports the desired feature that the systems undergo a random walk in the whole exchange range. Examined at higher resolution (a shorter time interval), plot a, for example,
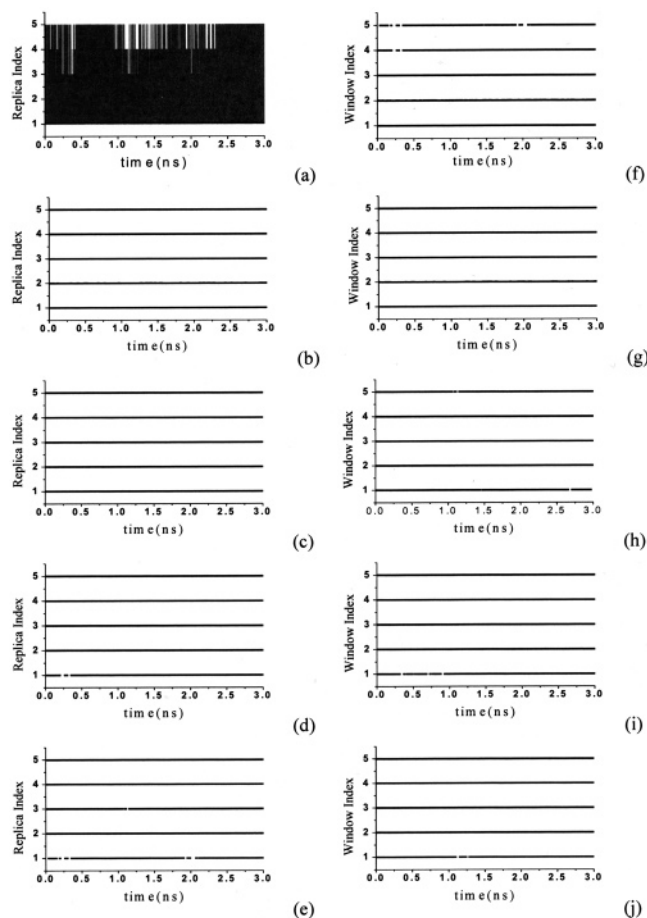


**Figure 2.** (a−e) Migration of systems into and out of a given configuration. (f−j) Migration of configurations (replicas) into and out of a given system ($\lambda_i$ scale value). The figures from top to bottom correspond to, for a−e, configurations 1−5 and, for f−j, systems 1−5. In a, the points are vertically connected, and the coverage demonstrates the desired itineration. Note that in view of the number of data points that are plotted, it appears as if, at a particular time, several replicas occupy the same window or several systems visit the same replica; this does not happen.

does show that replica 0 is slightly favored by lower-numbered systems, as actually can be inferred from the white space in the plot.

**HREM Trajectory PCA.** Principal component analysis is a useful tool for the systematic investigation of the conformational sampling of a peptide,[47,48] since it can reduce the high dimensional configuration space to an essential subspace that contains most of the significant, large-scale motions. Only window 1 ($\lambda_1 = 1$) corresponds to the normal force field, and all our following analysis for HREM is based on window 1 for this reason. The first 3 ns is considered as an equilibration period, and the PCA is constructed for the backbone atoms. Over the trajectory, the first mode represents around 40% of the total backbone variance (the total MSF); the first 2 modes, around 55%; and the first 10 modes, around 90%; showing that the HREM simulation of met-Enkephalin does provide a division into essential and remaining subspaces. An issue with PCA is that the slow relaxation of the large fluctuations may prevent the fast convergence of the $(\sigma)_{ij}$ covariance matrix,[55] which results in slow convergence of the essential space spanned by those large fluctuation modes. So before going through any detailed examination of the trajectories using PCA, we first test the convergence in the essential subspace with the RMSIP[50] method (eq 2.5), which can examine convergence by comparing

Studies of Met-Enkephalin

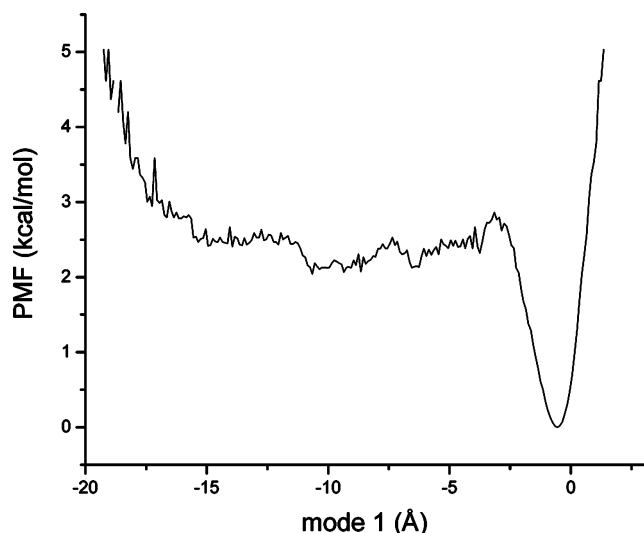*J. Phys. Chem. B, Vol. 111, No. 42, 2007* **12315**



**Figure 3.** The PMF corresponding to the PCA mode 1 displacement for all $\lambda = 1$ system snapshots.

**TABLE 2: The RMSIPs (for HREM Window 1) for the Three 3 ns Intervals (3−6, 6−9, and 9−12 Ns) Using the PCA First Mode, First 2 Modes and First 10 Modes**

|  | 3−6 vs 6−9 ns | 3−6 vs 9−12 ns | 6−9 vs 9−12ns |
|---|---|---|---|
| first mode | 0.972 | 0.991 | 0.980 |
| first 2 modes | 0.918 | 0.959 | 0.909 |
| first 10 modes | 0.926 | 0.988 | 0.914 |

the overlap of modes constructed by using different time intervals of the simulation. We compare RMSIPs among three time intervals, 3−6, 6−9, and 9−12 ns, for the window 1 trajectory ($\lambda_1 = 1$) for the first mode, the sum of the first 2 modes (these 2 modes will be used for the following analysis), and the sum of the first 10 modes. The RMSIP results, which are a direct measure of the projection of the basis of one subspace onto the other subspace, are listed in Table 2. All the RMSIP values are close to their limiting values of 1.0. The mode 1 result indicates that the three PCA first mode vectors constructed from the three time intervals are essentially the same. That suggests there might be some reaction coordinate correlated with the PCA first (slowest) mode and that 3 ns intervals are sufficiently long to capture this motion. The RMSIP values for the first 2 and 10 modes show a similar convergence. Thus, we may be confident that the HREM simulation over 3 ns intervals are sufficient to capture the fluctuations of the important PCA modes of Met-enkephalin.

**HREM 1D PCA PMF.** That the first mode captures ∼40% of the total MSF suggests its use as a reaction coordinate, and it should be more objective in character than that based on, for example, an end-to-end distance, as in the DREM simulation. The first mode displacement trajectory, $p_1(t)$, defined in eq 2.6 exhibits numerous transitions between its extreme values, which shows that a broad range of distances is being sampled repeatedly during the simulation time. Thus, it is legitimate to construct a potential of mean force. The PMF, constructed by making a histogram of $p_1(t)$ for the 3−6 ns time interval is displayed in Figure 3. Since we have already shown good PCA convergence, for the following discussion, unless specified, we will focus on the 3−6 ns interval trajectories for convenience. (Note that in PCA, the eigenvectors are determined only up to a sign, and consequently, there is no meaning to the direction of increasing mode coordinate.) There is a well-defined well around −0.5 Å, and a broad plateau region from 15 to −5 Å. The difference between the lowest points of the two wells is

2.0 kcal/mol, and the barrier between them is 2.9 kcal/mol. The barrier is not high, which suggests that switching between these two states should not be difficult. Although the deep well is 2 kcal/mol lower than the broad well, it is much narrower and suggests that there is an energy/entropy compensation tradeoff between the broad and sharp wells. These features support the observations of the coexistence of different conformations of Met-enkephalin in water and of a lack of distinguishable secondary structure.

The type of atom displacements that correspond to the first PCA mode can be inferred by calculating the correlation coefficient of the end-to-end distance for the first mode (obtained from eq 2.6 with $i = 1$) and for the true trajectory. The correlation coefficient is 0.998, which suggests that mode 1 is reporting on the end-to-end distance fluctuations. The PCA has thus succeeded in singling out the principal collective motion that spans the open to closed conformations.

**HREM 2D PCA PMF.** The second PCA mode has a significant contribution (15%) to the overall motion and motivates construction of the two-dimensional PMF displayed in Figure 4. There are two wells around (−0.5 Å, 2.3 Å) and (−0.5 Å, 7.0 Å) that correspond to the localized well in the one-dimensional PMF plot in Figure 2. Representative α-carbon (CA) pictures of the dominant backbone conformations are also shown. Configurations in the two wells have end-to-end distances corresponding to salt-bridged conformers, versus the more extended states of all the other displayed configurations. Figure 5 shows representative CA wire frame backbone structures with the Gly-2 and Gly-3 atoms explicit for the two wells. The differences in the backbone structures, with the parallel and antiparallel carbonyls, come from the differences of the Gly-2 Ψ and Gly-3 Φ dihedral angles. The first row of Figure 6 shows the Ramachandran plots for Gly2, Gly3, and Phe4, with snapshots picked from the PCA first mode within the range −0.6 to −0.4 Å and, in the second row, for all snapshots for comparison purposes. Although the configurations corresponding to the two deep wells in the 2D PMF have large differences in their Gly-2 Ψ and Gly-3 Φ dihedral angles, the main patterns of their backbones are similar, illustrating the known[56] feature that the Ψ($i$) and Φ($i + 1$) values of residues $i$ and $i + 1$ can compensate and still lead to overall similar structures. This is the only local mechanism in peptides (or proteins) that can lead to a structure with essentially the same overall conformation. The Ramachandran plots in the first row of Figure 6 are quite similar to those generated by van der Spoel and Berendsen[24] in their study of zwitterionic Leu-enkephalin solvated by water. Their simulations were started from a salt bridge and an open form that rapidly closes and remains mostly closed, so they mainly sample salt bridge conformers. They, too, note a Gly2 Ψ and Gly3 Φ compensation for these salt-bridged conformers.

The finding that the Gly2 and Gly3 are concentrated in two distinct regions of Ψ and Φ suggests that these residues may be participating in hydrogen bonding. For the "parallel" carbonyl arrangement of Figure 5, there can be Gly3 N−H to Met5 O=C hydrogen bonding/salt-bridge interactions. For the "antiparallel" arrangement, there can be Gly2 C=O Met5 N−H as well as Gly2 N−H Met5 O=C hydrogen bonding/salt-bridge interactions. Selected snapshots of the parallel and antiparallel conformers do show these interaction patterns that most likely aid in stabilizing the direct terminal salt-bridge interaction. However, hydrogen bond analysis of the trajectories shows that none of these additional interactions persists. The relevant distances are constantly fluctuating and are mainly beyond the
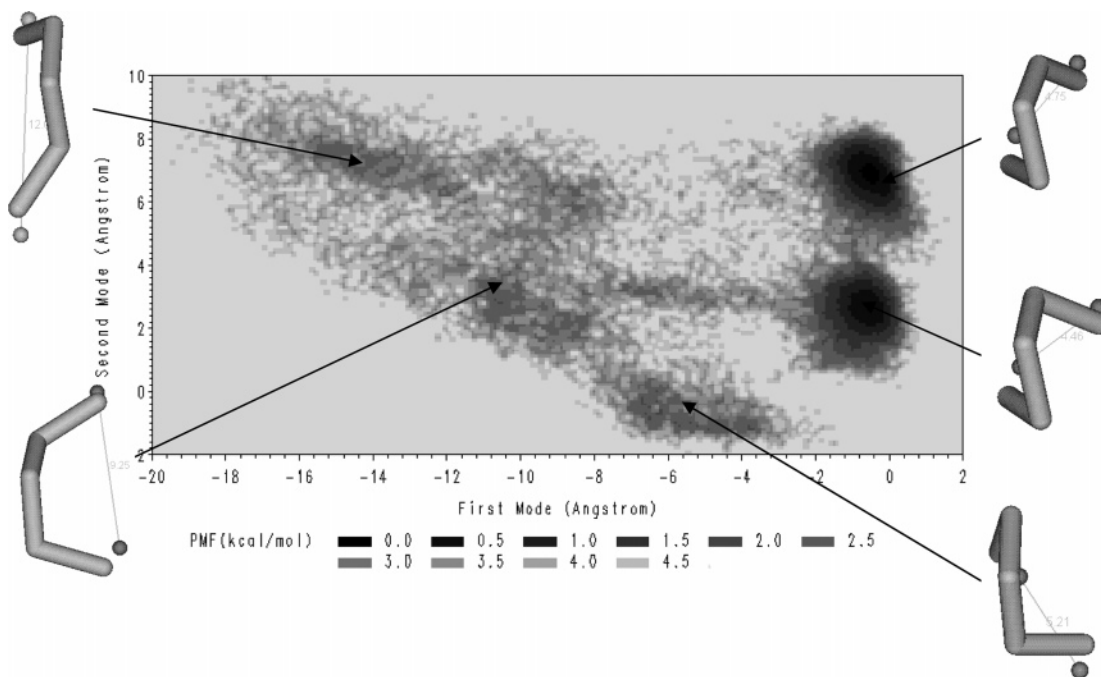
**Figure 4.** The 2D PMF for the first and second PCA modes for all $\lambda = 1$ system snapshots. Backbone CA stick plots of the configurations in the dense places are shown with the distance between Tyr1 backbone nitrogen and Met5 carboxyl carbon shown.
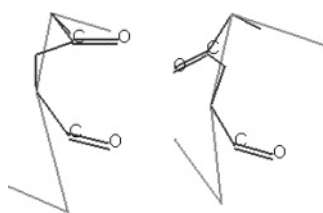


**Figure 5.** CA wire plots for the two salt bridge conformers shown in Figure 4, with the Gly-2 and Gly-3 backbone atoms shown explicitly that illustrate the $\Psi(2)$ and $\Phi(3)$ dihedral angle compensation mechanism.

distance that would permit such a conclusion. This result is based on both the HREM simulation and the DREM simulation in which the data for the end-to-end distance of 4.5 Å is used for the analysis. If it were persistent, the antiparallel arrangement would be similar to a 2−5 $\beta$ turn. NMR experiments on Leu-enkephalin in aqueous solution have not found evidence for a $\beta$ turn.[57]

In the remaining region of the 2D PMF plot in Figure 4, there is a less concentrated but much larger area of relatively low free energy than in the area around the two deep wells. The barrier between the deep and broad areas is only around 3 kcal/mol, and the difference between them is roughly within 2−3.5 kcal/mol, both of which are not very large on a thermal scale, but significantly, the pathway between two areas is very narrow. This observation suggests that although Met-enkephalin samples both regions, one favored by energy and the other favored by entropy, the rate of transition between them may be slower than might be inferred from the small free energy barrier. Starting from an extended configuration, there are many configurations that correspond to the extended states area to first sample, and then a restricted region in configuration space must be found that corresponds to the small transition area in Figure 4 to enter the salt-bridge region. This feature may explain the discrepancy between the experimental results[17−22] that Met-enkephalin or Leu-enkephalin shows great flexibility and a lack of definite conformations in water and the difficulty of going from the extended to closed conformations found through conventional

MD simulations.[2,7,23−31] Indeed, for the zwitterionic Met-enkephalin simulated here, an 18 ns normal MD simulation with the same starting configuration and conditions as for the HREM window 1 ($\lambda = 1$) is trapped in extended states and seldom samples configurations corresponding to the deep wells.

**HREM PMF Tyr1 Terminal Nitrogen to Met5 Terminal Carboxylate Carbon.** The zwitterionic form of Met-enkephalin should be capable of salt bridge formation between the terminal Tyr1 amide and Met5 carboxyl groups. As already noted, there is a strong correlation between the end-to-end distance fluctuations for the complete trajectory and that found from the first PCA mode. In Figure 7a, we display the 1D PMF along this end-to-end distance calculated from the trajectory of the HREM window 1 for three time intervals, 3−6, 6−9, and 9−12 ns, along with the total 18 ns (3−21 ns) interval. The agreement among the three PMFs is good with the exception of the region 9−14 Å that has a discrepancy of about 1 kcal/mol. The greater fluctuation in this region is most likely due to the extensive configuration space sampling for these less constrained regions. For comparison, Figure 7b displays a "pseudo PMF" generated from the 18 ns normal MD simulation that, as we noted in the previous section, does not adequately sample close-in conformations. There is a flat, broad well from 9 to 15 Å, similar to the HREM or DREM simulations, but the salt bridge region is not sampled reliably.

Comparison of the DREM PMF in Figure 1b with the corresponding HREM-generated PMF shows that they are in good agreement. There is some difference in the salt bridge peak. The DREM peak is shifted down ∼0.25 kcal/mol relative to the HREM. In both, there is evidence for a slight double well structure around 10 and 12 Å. Thus, the HREM simulation, in which there is no explicit restraint potential, does a remarkably good job of mirroring the results from the DREM simulation. In principle, a DREM simulation should provide a more accurate PMF along a chosen reaction coordinate than a HREM simulation with respect to the particular reaction coordinate. Furthermore, the total time over all processors in our simulation, including the equilibration times, was 204 ns for the DREM and 105 ns for the HREM simulations. However,
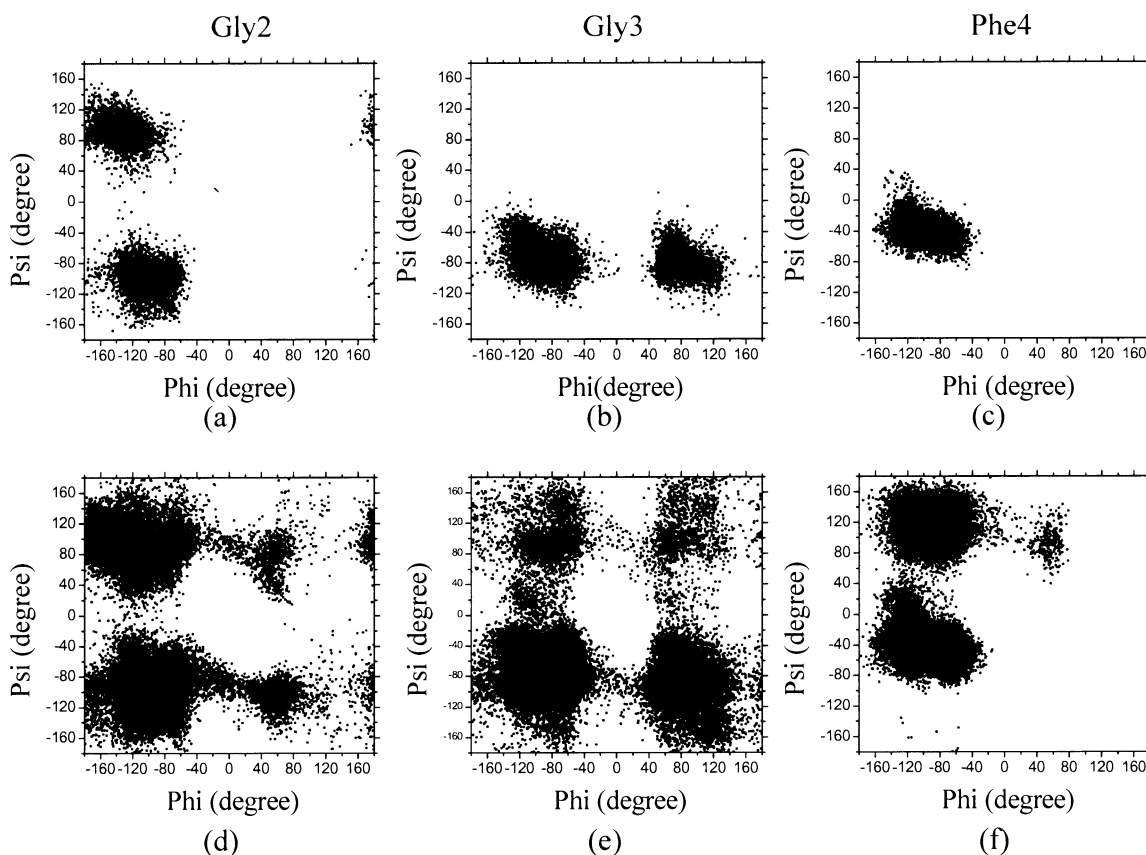
Studies of Met-Enkephalin

*J. Phys. Chem. B, Vol. 111, No. 42, 2007* **12317**



**Figure 6.** Ramachandran plots for Gly2, Gly3, and Phe4 of Met-enkephalin: (a−c) snapshots that are in the PCA first mode deep well (−0.6 to −0.4 Å); (d−f) all snapshots.

it is not so straightforward to compare a simulation that uses restraints with a given strength (the DREM) with a simulation that uses only a scaled potential energy.

**DREM Entropy Energy Separation.** The separation of the PMF according to $\Delta A(\mathbf{r}) = \Delta E(\mathbf{r}) - T\Delta S(\mathbf{r})$ was evaluated using the scheme introduced in the Methodology Section. Figure 8 shows the resulting energies for the two DREM simulation ranges for several time intervals. Note that the energies are given for the 2−4 and 2−6 ns ranges to indicate the scale of fluctuation in their determination. The open range data, after the barrier at 7 Å, exhibits the greatest fluctuations, as does the PMF, although the PMF fluctuations are much smaller for the same time intervals. In spite of these inaccuracies in the determination of $\Delta E(r)$ and $\Delta S(r)$, some conclusions are safe to make. Excluding the ends of the ranges, there seems to be a minimal dependence of entropy and energy on the reaction coordinate. In the barrier region, if there is an entropic increase, according to Figure 8a, it is quite modest. As shown in another fashion below, the entropy associated with the peptide is mainly due to side-chain atom fluctuations, and these are not very different along the reaction coordinate.

The close in data in Figure 8a (note that it is plotted starting at 3.5 Å versus in Figure 7, which starts at 3.0 Å) would suggest that the entropy is increasing as the end-to-end distance $r$ decreases. One possible source for this increase is that as the charged ends are forced together, the effective dipole moment of the zwitterion decreases, and this leads to a decrease in solvent ordering. However, examination of the window data shows that the distances ∼3.0 Å are not well-sampled. This is not a problem for $\Delta A(r)$, but for the separation into its components, it is a severe issue. An attempt was made to examine the close in distances by running a DREM simulation

with 17 windows that span 3.0−5.0 Å with uniform separation of 0.125 Å between neighboring windows using much stiffer (64 kcal/mol) force constants. Now the energy is more in accord with the free energy of Figure 8a, but the sampling is insufficient to obtain a reliable $\Delta A(r)$. Since presumably the energy increases with decreasing distance in this region mainly due to van der Waals repulsion of the terminal groups, the free energy should be dominated by this energy contribution.

The 8−16 Å data in Figure 8b show that the energy fluctuates on essentially the same scale for the 2−4 and 2−6 ns data intervals, indicating that a substantially longer time simulation would be required for its accurate determination in this distance interval. Again, the problem of insufficient sampling, now around 16.0 Å, arises. Another 17-window DREM simulation over the interval 14.5−16.5 Å with 64 kcal/mol force constants was run. The free energy profile over this range is essentially identical to the results in Figure 8b. The energy evaluated over different time intervals is quite well converged and is increasing with increasing reaction coordinate distance in qualitative agreement with the free energy. Thus, there is not a large entropy decrease when Met-enkephalin is stretched.

In view of the difficulties with the above approach, another method that focuses on the entropic contribution from the peptide degrees of freedom for different reaction coordinate distances is worth investigating. In addition, it is instructive to consider the nature of the configurational space exploration of Met-enkephalin on the basis of the time course of the peptide atom fluctuations for different restrained distances. Of course, separating state functions into components cannot be done rigorously unless the components correspond to independent sets of degrees of freedom,[58] but that may be reasonable for the peptide and solvent degrees of freedom. The RMSF_$t(t)$
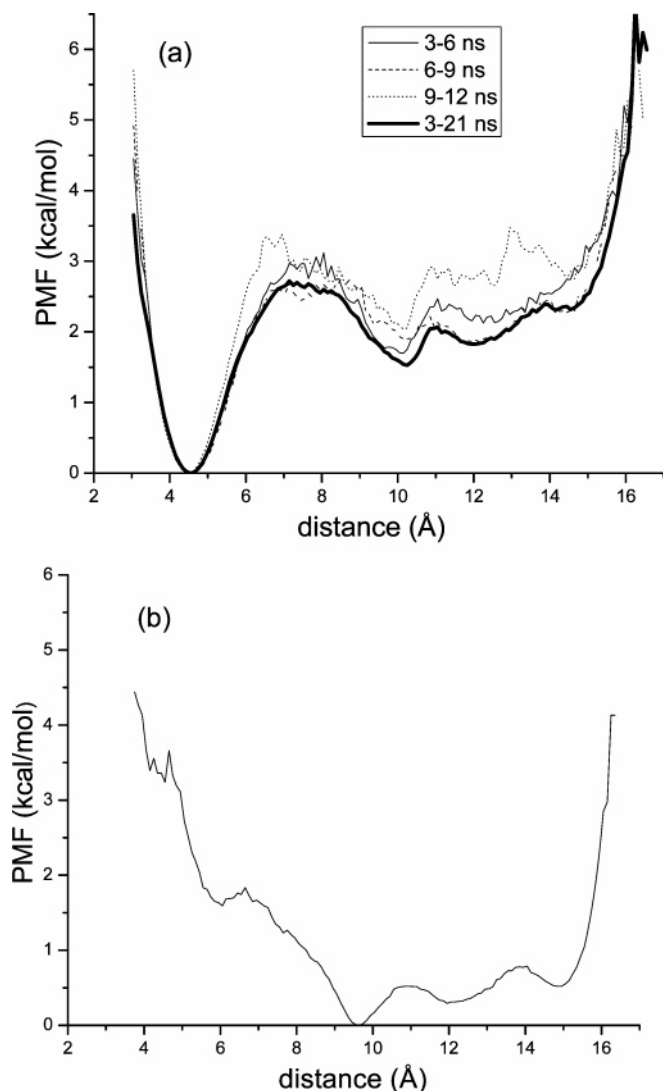
**Figure 7.** (a) The PMF for the end-to-end distance generated by the HREM simulation for all systems with $\lambda = 1$, averaged over several 3 ns intervals and the total averaged over 18 ns (bold line). (b) The "pseudo PMF" for the end-to-end distance generated by an 18 ns normal ($\lambda = 1$) MD simulation.



**Figure 8.** The free energy and energy from the DREM simulations. (a) The 3.5−11 Å range. (b) The 8−16 Å range. The time ranges for generating the energies are given, indicating the greater fluctuation inherent to the energy determination that is most evident for the open form of Met-enkephalin.

defined in eq 2.9 is a measure of the time course of the fluctuations from the average configuration for all the atoms (main and side chain) of the peptide. Figure 9 displays this time dependence; there are large fluctuations that indicate the extensive sampling of configuration space. The 4.5 Å data is somewhat more constrained than for 12 Å, suggesting that conformations corresponding to the salt bridge structure fluctuate less than those of the open form. The large fluctuations in RMSF_$t(t)$ demonstrate that long averaging times will be required in obtaining quantities such as $\Delta E(r)$ and $\Delta S(r)$. Another view of the peptide fluctuations is shown in Figure 10, which displays the RMSF$_j$'s (cf. eq 2.8), the conventional time-averaged root-mean-square fluctuation for each atom, $j$. The backbone atom RMSF$_j$'s, indicated in the figure with squares, are much smaller than those of the side-chain atoms. The backbone RMSF$_j$'s for the 4.5 Å distance are somewhat smaller than found in typical proteins, whereas for the 12.0 Å distance they are comparable to the values obtained from B-factors in protein crystallography. The open form atom fluctuations are larger than those of the salt-bridge structure for most backbone and all side-chain atoms. It should be clear from Figure 10 that the side-chain atoms dominate the entropy. The
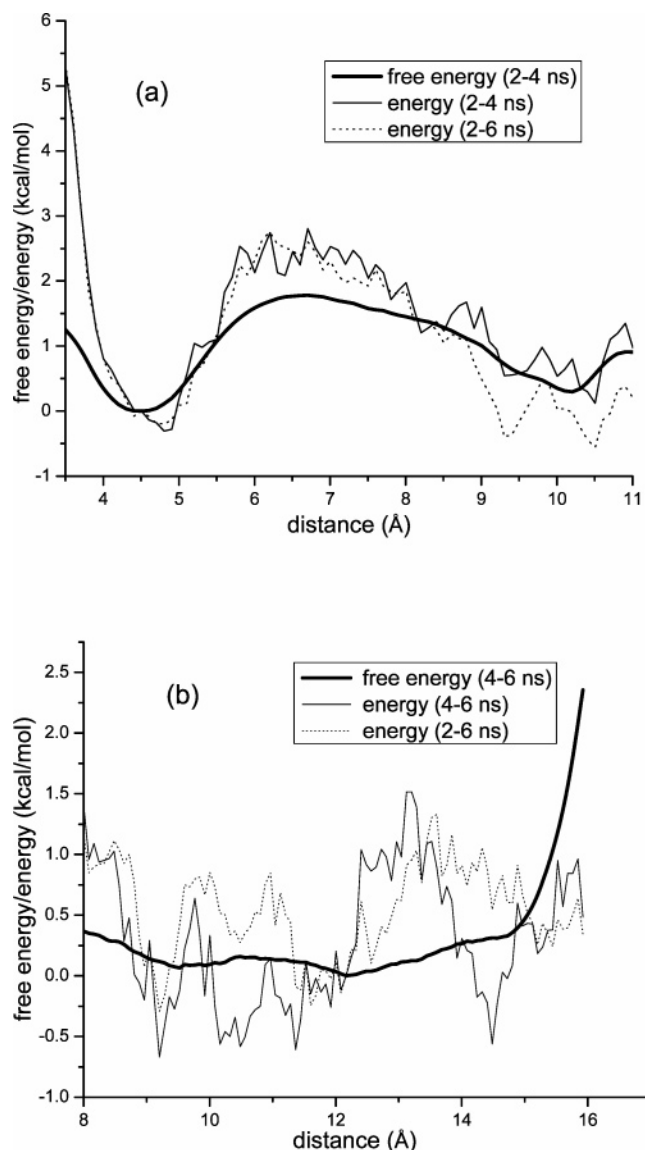
additional 14.5−16.5 Å DREM simulation shows that the backbone RMSF$_j$ values for 16.0 Å are even smaller than those found for 4.5 Å, indicating that the peptide is stretched, whereas the side-chain RMSF$_j$ values are comparable.

An estimate of the entropy difference of the peptide between two restrained distances arising from differences in configurational sampling may be obtained from this simulation data as follows. If $Z[t] \equiv \text{RMSF}\_t(t)$ is assumed to be generated by a stochastic process with a Gaussian probability density, then

$$p(Z)\, dZ = \frac{1}{2\pi C^{1/2}} \exp\left(-\frac{1}{2} C^{-1}(\delta Z)^2\right) dZ \qquad (3.1)$$

with $C = \langle(\delta Z)^2\rangle$ the distribution's width, and $\delta Z = Z - \langle Z\rangle$. That $Z[t]$ might have a Gaussian distribution could be suggested from its form as a sum of a large number of variables and an appeal to the central limit theorem.[59] (Actually, the distribution is formed from those for the three Cartesian dimensions, each of which is assumed to be Gaussian, and is therefore skewed
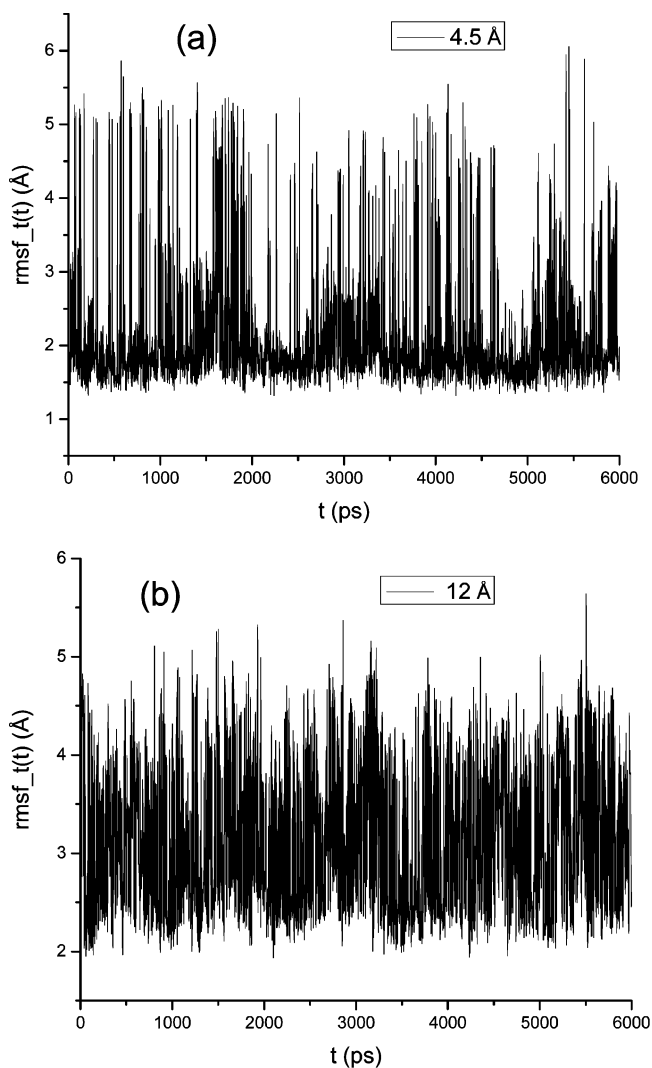
Studies of Met-Enkephalin

*J. Phys. Chem. B, Vol. 111, No. 42, 2007* **12319**



**Figure 9.** All atom RMSF_$t(t)$ defined in eq 2.9 versus time (a) for DREM distance 4.5 Å and (b) for DREM distance 12.0 Å.
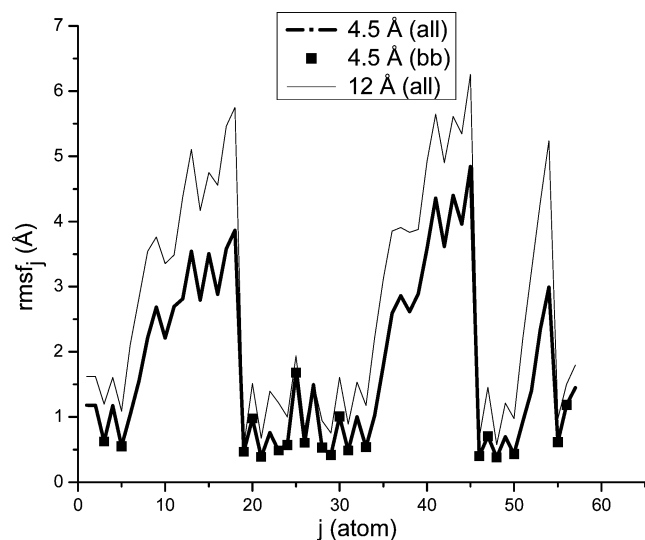


**Figure 10.** The RMSF$_j$'s for each atom $j$ for the DREM distances 4.5 and 12.0 Å, with the backbone atom RMSF$_j$'s for 4.5 Å indicated with squares.

from Gaussian, but that does not affect results based on the relative widths of distributions for two restraint distances.) The corresponding entropy is given by $S = -k_B \int p(Z) \ln p(Z) \, dZ$, and this leads to

$$\Delta S = S_2 - S_1 = k_B \ln[C_2/C_1] = k_B \ln[\text{MSF}_2/\text{MSF}_1] \tag{3.2}$$

The last equality, which follows from the definition of $C$ and eqs 2.8 and 2.9, introduces the total mean square fluctuation $\text{MSF} = \sum_{j=1}^{N_{at}} \text{RMSF}_j^2$ and shows that it provides a measure of the entropy difference between the two restrained distance subensembles. Applying these considerations to the data in Figure 10 provides a value of $T\Delta S = 0.45$ kcal/mol. Thus, the entropic effect from conformational sampling at the indicated distances is quite small. The conclusion of a small entropic effect for distances excluding the ends of the range, based on the scheme in Appendix A, agrees with this analysis These considerations rely on the Gaussian nature of the components of $\delta Z = Z - \langle Z \rangle$. The data in Figure 9 show that these fluctuations are not well described as a skewed Gaussian, most likely due to the correlations among the atom positions that contribute to $Z$. Nevertheless, since the entropy for the various restrained distances is dominated by the side-chain fluctuations and these are not greatly affected by the restraint distance, the peptide entropy over the sampled reaction coordinate range is not changing significantly.

## 4. Conclusions

Two approaches, DREM and HREM, to improving the MD sampling of zwitterionic Met-enkephalin in explicit solvent were investigated. Using the DREM with the one-dimensional reaction coordinate of the end-to-end distance provides a potential of mean force that exhibits a well-localized salt-bridge state separated by a small barrier from a broad region. An advantage of a reaction coordinate in the DREM context is that, as evident from eq 2.4, statistical fluctuations are minimized, since the exchanges depend only on differences in energy that are specific to the one restrained degree of freedom. The difficulties found in separating the PMF into its entropic and energetic contributions in part reflect the loss of this statistical advantage. For a reaction coordinate simulation, free energy convergence depends on only that degree of freedom, whereas the energy convergence will depend on all the system degrees of freedom. Therefore, energy and entropy convergence must be much slower.

The HREM simulation succeeded in sampling configurations that correspond to all accessible distances along the DREM reaction coordinate without a restraint to force the system to the desired distances. It did so quite efficiently in the sense that only five windows were required to generate results that, as monitored by the PCA modes, show good convergence properties in the simulation time. In addition, and essential for properly exploring the configuration space of Met-enkephalin, the mode 1 trajectory repeatedly sampled configurations that correspond to all the relevant end-to-end distances, and this permitted construction of free energy quantities, such as a potential of mean force. The PCA was effective in singling out this dominant motion in the first mode.

The HREM simulation trajectory, analyzed with the PCA decomposition, also revealed the presence of another significant PMF dimension that is mainly composed of a correlated local change of the Gly2 $\Psi$ and Gly3 $\Phi$ dihedral angles. That two very different regions of dihedral space for these residues are stably occupied, even though the overall structure, as monitored by the end-to-end distance, corresponds to a salt-bridge conformer, results from this local $\Psi(i)$ and $\Phi(i + 1)$ compensation.

An advantage of the HREM is that it does not require a commitment to a particular reaction coordinate, or a particular dimensionality of reaction coordinate. The objectivity of the

HREM in this regard is an argument in its favor relative to the DREM in which a reaction coordinate must be specified. Of course, the HREM has the drawback that it could fail to sample adequately regions of configuration space that would reveal some crucial phenomenon. The TREM is objective because all degrees of freedom are thermally excited. However, for a system with many degrees of freedom, the chain of replicas required may become impractically large. The HREM can be viewed as an attempt to excite only important degrees of freedom but that requires a decision as to which are the important degrees of freedom. For explicit solvent simulations, not having to excite the solvent degrees of freedom is a great advantage since their number is an order of magnitude larger than the number of peptide (and protein) degrees of freedom in typical MD simulations.

Even though Met-enkephalin is small when compared with proteins, the dispersion of its fluctuations as indicated by Figure 9 and their scale (dominated by the side chains), as indicated by Figure 10, are both large. These features stand in contrast with a typical folded protein. Thus, properly sampling the Met-enkephalin configuration space is more an issue of having to explore a large space with small potential energy barriers than of having to overcome a large barrier along a reaction coordinate for some transformation, as in a folded, stable protein. The result is a free energy profile that does not exhibit a large barrier yet requires a great deal of simulation time and special MD methods, such as the DREM and HREM.

## Appendix A

An expression suited for obtaining the free energy decomposition into energy and entropy contributions $\Delta A(\mathbf{r}) = \Delta E(\mathbf{r}) - T\Delta S(\mathbf{r})$ by the weighted histogram analysis method[32,33] can be developed as follows. For simplicity, we use a scalar coordinate $r$ for the derivation. Introduce an $r$-dependent partition function,

$$Q(r) = \int d\mathbf{\Gamma}\, \delta(r - f(\mathbf{X}))e^{-\beta H}(\mathbf{\Gamma}) \tag{A.1}$$

where $\mathbf{\Gamma} = P, X$ denotes the phase space. Then,

$$-\beta \Delta A(r) = \ln(Q(r)/Q(r_{\text{ref}})) = \ln(p^{(u)}(r)/p^{(u)}(r_{\text{ref}})) \tag{A.2}$$

where $p^{(u)}(r) = Q(r)/Q$ is the unbiased probability density corresponding to $H(X,P)$ and $Q$ is the canonical ensemble partition function. The desired average energy,

$$E_c(r) = -\frac{\partial}{\partial \beta} \ln Q(r) = \int d\mathbf{\Gamma}\, H(\mathbf{\Gamma})\delta(r - f(\mathbf{X}))e^{-\beta H}(\mathbf{\Gamma})/Q(r) \tag{A.3}$$

can be expressed using eq A.3 as

$$E_c(r) = \int d\mathbf{\Gamma}\, dE\, E\delta(E - H(\mathbf{\Gamma}))\delta(r - f(\mathbf{X}))e^{-\beta H}\mathbf{\Gamma}/Q(r) =$$
$$\int dE\, E p^{(u)}(E,r)/p^{(u)}(r) = \int dE\, E p^{(u)}(E|r) \tag{A.4}$$

with the definition of the unbiased joint probability density,

$$p^{(u)}(E,r) = \int d\mathbf{\Gamma}\, \delta(E - H(\mathbf{\Gamma}))\delta(r - f(\mathbf{X}))e^{-\beta H(\mathbf{\Gamma})}/Q \tag{A.5}$$

Equations A.3–A.5 make it clear that the decomposition of a free energy along a reaction coordinate into energy and entropy is related to $p^{(u)}(E|r)$, the conditional probability of energy values $E$ given a value of the reaction coordinate $r$, and therefore

requires information about all the degrees of freedom, peptide, and solvent. This shows, in the context of a reaction coordinate free energy method, that it is much more demanding statistically to obtain $\Delta E(\mathbf{r})$ and $\Delta S(\mathbf{r})$ versus $\Delta A(\mathbf{r})$. (If the derivation is carried out using the configuration space and configurational partition function instead of the phase space, the energy values will be those coming from the total potential energy.)

WHAM combines the data from multiple windows of different bias window potentials by writing the true (unbiased by the window potentials) estimated probability density as a linear combination of the window biased probability densities, $p_i^{(b)}(r)$

$$p^{(u)}(r) = \sum_{i=1}^{N} \frac{n_i p_i^{(b)}(r)}{G(r)} \tag{A.6}$$

with

$$G(r) = \sum_{i=1}^{N} n_i e^{-\beta(W_i(r) - f_i)} \tag{A.7}$$

Here, $n_i$ is the number of data points in the $i$th window, and $f_i$ is the free energy of the system with added potential $W_i(r)$. The free energy values implicit in eqs A.6 and A.7 are obtained by an iterative scheme, and their values are then used to obtain the unbiased probability distribution. The extension of WHAM to obtain the joint probability density $p^{(u)}(E,r)$ defined in eq A.5 is straightforward because the window potential $W_i(r)$ is not changed. Therefore,

$$p^{(u)}(E,r) = \sum_{i=1}^{N} \frac{n_i p_i^{(b)}(E,r)}{G(r)} \tag{A.8}$$

and the desired conditional probability average is obtained from eq A.4.

## References and Notes

(1) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249.
(2) Hansmann, U. H. E.; Okamoto, Y.; Eisenmenger, F. *Chem. Phys. Lett.* **1996**, *259*, 321.
(3) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776.
(4) Lyubartsev, A.; Laaksonen, A. *Applied Paralled Computing* **1998**, *1541*, 296.
(5) Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1986**, *57*, 2607.
(6) Geyer, C. J. Markov Chain Monte Carlo Maximum Likelihood. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*; Keramidas, E. M., Ed.; Interface Foundation: Fairfax Station, 1991.
(7) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
(8) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140.
(9) Wang, J. S.; Swendsen, R. H. *Prog. Theor. Phys.* Supplement **2005**, 317.
(10) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604.
(11) Lei, H.; Duan, Y. *Curr. Opin. Struct. Biol.* **2007**, *17*, 187.
(12) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058.
(13) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic: San Diego, 1996.
(14) Predescu, C.; Predescu, M.; Ciobanu, C. V. *J. Phys. Chem. B* **2005**, *109*, 4189.
(15) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042.
(16) Lou, H. F.; Cukier, R. I. *J. Phys. Chem. B* **2006**, *110*, 24121.

Studies of Met-Enkephalin

*J. Phys. Chem. B, Vol. 111, No. 42, 2007* **12321**

(17) D'Alagni, M.; Delfini, M.; DiNola, A.; Eisenberg, M.; Paci, M.; Roda, L. G.; Veglia, G. *Eur. J. Biochem.* **1996**, *240*, 540.

(18) Graham, W. H.; Carter, E. S.; Hicks, R. P. *Biopolymers* **1992**, *32*, 1755.

(19) Higashijima, T.; Kobayashi, J.; Nagai, U.; Miyazawa, T. *Eur. J. Biochem.* **1979**, *97*, 43.

(20) Marcotte, I.; Separovic, F.; Auger, M.; Gagne, S. M. *Biophys. J.* **2004**, *86*, 1587.

(21) Surewicz, W. K.; Mantsch, H. H. *Biochem. Biophys. Res. Commun.* **1988**, *150*, 245.

(22) Takeuchi, H.; Ohtsuka, Y.; Harada, I. *J. Am. Chem. Soc.* **1992**, *114*, 5321.

(23) Hansmann, U. H. E. *Eur. Phys. J. B* **1999**, *12*, 607.

(24) Hansmann, U. H. E.; Okamoto, Y. *J. Comput. Chem.* **1997**, *18*, 920.

(25) van der Spoel, D.; Berendsen, H. J. C. *Biophys. J.* **1997**, *72*, 2032.

(26) Aburi, M.; Smith, P. E. *Biopolymers* **2002**, *64*, 177.

(27) Nielsen, B. G.; Jensen, M. O.; Bohr, H. G. *Biopolymers* **2003**, *71*, 577.

(28) Shen, M. Y.; Freed, K. F. *Biophys. J.* **2002**, *82*, 1791.

(29) Zaman, M. H.; Shen, M. Y.; Berry, R. S.; Freed, K. F. *J. Phys. Chem. B* **2003**, *107*, 1685.

(30) Karvounis, G.; Nerukh, D.; Glen, R. C. *J. Chem. Phys.* **2004**, *121*, 4925.

(31) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 225.

(32) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.

(33) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40.

(34) Fleischman, S. H.; Brooks, C. L. *J. Chem. Phys.* **1987**, *87*, 3029.

(35) Yu, H. A.; Karplus, M. *J. Chem. Phys.* **1988**, *89*, 2366.

(36) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. *J. Chem. Phys.* **2003**, *119*, 3559.

(37) Peter, C.; Oostenbrink, C.; van Dorp, A.; van Gunsteren, W. F. *J. Chem. Phys.* **2004**, *120*, 2652.

(38) Qian, H.; Hopfield, J. J. *J. Chem. Phys.* **1996**, *105*, 9292.

(39) Meirovitch, H. *Curr. Opin. Struct. Biol.* **2007**, *17*, 181.

(40) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P. *Biomolecular simulation: the GROMOS96 manual and user guide*; Vdf hochschulverlag AG an der ETH: Zürich, 1996.

(41) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(42) Berendsen, H. H. C.; Postma, J. P. M.; Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(43) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

(44) Cox, T. F.; Cox, M. A. A. *Multidimensional scaling*, 2nd ed.; Chapman & Hall: Boca Raton, 2001.

(45) García, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696.

(46) García, A. E.; Blumenfeld, R.; Hummer, G.; Krumhansl, J. A. *Physica D* **1997**, *107*, 225.

(47) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412.

(48) Romo, T. D.; Clarage, J. B.; Sorensen, D. C.; Phillips, G. N. *Proteins: Struct. Funct. Genet.* **1995**, *22*, 311.

(49) Teodoro, M. L.; Phillips, G. N.; Kavraki, L. E. *J. Comput. Biol.* **2003**, *10*, 617.

(50) Amadei, A.; Ceruso, M. A.; DiNola, A. *Proteins: Struct. Funct. Genet.* **1999**, *36*, 419.

(51) Hess, B. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2002**, *65*, 031910.

(52) Lou, H. *Analyzer*, 1.0 ed.; East Lansing, 2005.

(53) McCammon, A.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Cambridge University Press: Cambridge, 1987.

(54) Kumar, S.; Nussinov, R. *J. Mol. Biol.* **1999**, *293*, 1241.

(55) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. *J. Phys. Chem.* **1996**, *100*, 2567.

(56) Korn, A. P.; Rose, D. R. *Protein Eng.* **1994**, *7*, 961.

(57) Sakarellos, C.; Gerothanassis, I. P.; Birlirakis, N.; Karayannis, T.; Sakarellosdaitsiotis, M. *Biopolymers* **1989**, *28*, 15.

(58) Smith, P. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1994**, *98*, 13735.

(59) Gnedenko, B. V. *Theory of Probability*; Chelsea Publishers: New York, 1962.